Chapter 2. Measurement

As far as their sources of data are concerned, economists are in a better position than most social scientists. Many goods are priced and transacted in markets. This produces a **wealth of data** that can be used to **motivate theories**, **test their predictions**, and to **guide policy**. Our goal is to develop theories that are consistent with this data. The data suggests which factors we need to include in our theory and which factors do not seem to be relevant. We use these theories to make **qualitative and quantitative forecasts** and to evaluate the **effects of different policies**. With data we can **test competing theories**, comparing what actually happens to what different theories predict it should happen.

For example, surveys of consumer prices, conducted on a monthly basis, provide information on the evolution of prices through time. Combining this information with observations on other economic variables allows the development of theories that explain why prices change over time.

The goal of **National Income accounting** is to provide a systematic method for aggregating the output of diverse sectors producing different goods and services into a **single measure of overall economic activity**. Like accounting for a firm, it allows analyzing the state of an economy at **a given point in time**, the **changes over time**, and **differences across countries**.

We will pay special attention to the distinction between **real** (expressed in terms of goods) and **nominal** (expressed in terms of money) variables. For instance, the (hourly) **nominal wage** is the number of dollars received for one hour of work, while the (hourly) **real wage** is the amount of goods and services that can be purchased with one hour of work. To the extent that money is mostly valued by the command over goods and services it provides (either today or at some point in the future), we will focus on real variables. Nonetheless, we will use nominal variables as an step in the contruction of real variables.

We will discuss many macroeconomic variables, some of which are **stocks** and some are **flows**. Flows are measured as an **amount per unit of time**, in contrast stocks are measured as the **amount at a given point in time**. Consider a bathtub and open the faucet. After 2 minutes there is a stock of water in the bathtub of 120 liters and each second a flow of 1 additional liter of water adds to that stock. As you can see stocks and flows are clearly related. A stock is often an accumulation of flows over time. For instance, your monthly **saving is a flow** and the **balance of your savings account** is the associated **stock**.

Finally, it is important to keep in mind that national accounting is based on a more or less reasonable **set of conventions**, that slowly adapt to a changing environment.

1 Gross Domestic Product

We will denote gross domestic product as Y. **Gross Domestic Product** (GDP) is the **market value** of the **final** goods and services **produced** in a **country** over a **period of time**. GDP is the **basic measure of economic activity**. GDP is a **flow variable** since it is measured per unit of time.[1]

[1] Diane Coyle's book "GDP: a brief but affectionate history" provides an overview of the history of measurement of economic activity. The first attempts to measure national wealth and production date back to the XVII when William Petty used them to evaluate the affordability of the Second Anglo-Dutch war. Nonetheless measures of aggregate economic activity did not become comprehensive until the Great depression of 1929 when governments felt the need to accurately assess the costs of the recession. For Canada, measures of aggregate output have been published on a regular basis since 1947 (one can find GDP measures for earlier times, but these have been

There are three approaches to measuring GDP: the production (**value-added**) approach, the **income** approach, and the **expenditure** approach.

The **production (or value-added)** measure of GDP adds the market value of the final goods and services produced in the economy or, equivalently, the **value added of all the productive units in the economy**.

The **income measure** of GDP counts all the **income generated in the production process in the economy.**

The **expenditure measure** of GDP adds the value of **total purchases** of final goods and services in the economy.

All three approaches will give the **same measure of GDP**.

$$Production = Income = Expenditure = Y$$

Let's illustrate it with an example. Consider a simplified economy consisting of two firms. Firm 1 produces steel using workers as the only input. It sells the steel for \$100 to a second firm that produces cars. The steel firm pays its workers \$80, pays \$10 dollars of taxes to the government and keeps the rest as profits. Firm 2, the car producer, buys \$100 worth of steel and hires workers for \$50 to produce cars. The market value of the cars produced is \$210, \$200 are sold to families and the remaining \$10 remains unsold and are kept in the form of inventories. The car producer pays \$40 worth of taxes.

The government hires \$50 worth of workers to monitor the traffic (police).

Finally, the consumers work for the firms and for the government earning total wages of \$180. Since the consumers are also the ultimate owners of the firms they receive an after-tax profit of \$30. The following table summarizes the information from our simple example.

Firms	**Steel co.**	**Car co.**
Revenues	**100**	**210**
Sales	100	200
Change inventories	0	10
Expenses	**80**	**150**
Wages	80	50
Other inputs	0	100
Taxes	**10**	**40**
Profits	**10**	**20**

Government	
Tax revenues	50
Wages (Police Services)	50

Consumers	
Wage Income	180
Profits distributed by firms	30

Figure 1: Calculating GDP: A simple economy.

Let's use this simple economy to **illustrate the three approaches** to calculating GDP.

constructed retrospectively using a quasi-archeological approach).

1.1 The production (value-added) approach to GDP

Gross Domestic Product (GDP) is the **market value of the final goods and services** produced in a country over a period of time.

In the value-added (or production) approach, GDP is calculated as the sum of **value added to goods and services in production across all productive units in the economy**. We would like to add the value the value of all goods and services produced in the economy and then subtract the value of the intermediate goods used in the production of other goods to **avoid double-counting**, i.e. in the previous example the steel company produces an input, steel, that is used by the car company. The market value of the cars obviously includes the value of the steel used to produce them, so if we simply added the values of steel and cars we will be overestimating the true amount of (new) goods produced.

To compute GDP following the value-added approach in our example, we proceed as follows. The economy produces \$360 worth of goods and services (steel, cars and police services), of which \$100 are intermediate goods so our value-added measure of GDP would be \$360-\$100=\$260.[2] [3]

We can look at this example in a different way. Suppose that both firms merge so that the sale of steel took place inside the new firm and is no longer recorded.

Firms	**~~Steel co.~~**	**~~Car co.~~**	**Merged co.**
Revenues	**~~100~~**	**~~210~~**	**210**
Sales	~~100~~	~~200~~	200
Change inventories	~~0~~	~~10~~	10
Expenses	**~~80~~**	**~~150~~**	**130**
Wages	~~80~~	~~50~~	130
Other inputs	~~0~~	~~100~~	0
Taxes	**~~10~~**	**~~40~~**	**50**
Profits	**~~10~~**	**~~20~~**	**30**

Government	
Tax revenues	50
Wages (Police Services)	50

Consumers	
Wage Income	180
Profits distributed by firms	30

Figure 2: Calculating GDP: Changes in the productive structure I.

All we would see would be one firm producing \$210 worth of cars, paying workers \$130, \$50 worth of taxes and making a profit of \$30. Our measure of GDP remains unchanged, i.e. GDP is invariant to the distribution of firms or the productive structure of the economy.

Note 1: Intermediate goods and services are excluded to avoid double counting (a tire bought by BMW to incorporate to a car is an intermediate good but a tire bought by a family to replace a flat one in their car is a final good).

[2] Alternatively, we could add the value added in each productive unit in the economy. For instance, in our example the \$210 worth of cars produced by the car company include \$100 worth of steel, an intermediate input, and therefore the value added of this company is \$210-\$100 = \$110.

[3] In the case where taxes are zero and the government borrows to finance the police services, value added and therefore GDP from the production approach remains unchanged.

Note 2: Goods that are produced and not sold are considered as purchased by the firm that produces them, i.e. as an increase in its inventories. In a sense **everything that is produced is recorded as sold** (either to customers or to the own firm). As a result of this convention, **total final production** (value-added approach) always coincides with **total purchases of final goods and services** (expenditure approach).

Note 3: **Only new (current-year) production** of goods and services is counted in GDP. Imagine next year, year t+1, everyone produces the same, but the Car company sells not only all year t+1 production, but also the $10 inventory of cars that were produced in year t. This transaction will not be part of year t+1 GDP, since it was already counted in year t when the cars were actually produced.

The summary of year t+1 production/transaction is given by

Firms	**Steel co.**	**Car co.**
Revenues	**100**	**210**
Sales	100	220
Change inventories	0	-10
Expenses	**80**	**150**
Wages	80	50
Other inputs	0	100
Taxes	**10**	**40**
Profits	**10**	**20**

Government	
Tax revenues	50
Wages (Police Services)	50

Consumers	
Wage Income	180
Profits distributed by firms	30

Figure 3: Calculating GDP: Sales vs. inventories.

The same logic applies applies to second-hand goods.

Note 4: Since **government services** (police) are not (usually) sold in markets and therefore they have no market price the standard practice is to value them at the **cost of** inputs of **production**.

Finally, notice that the **flow of labor** (for instance the $80 of wages of the steel company) is **not treated as an intermediate good**. If, in our example, one of the employees in the car company who was a lawyer quits her job and starts an independent firm that provides legal advice to the car company.

Then, the market value of this legal advice would be considered as an intermediate input, but GDP would remain unchanged (since the increase in value added from the new firm, the small legal company, is compensated by a decrease in value added of the car firm).

1.2 The income-approach to GDP

A second way of computing **GDP is from the income side**, i.e. how is the revenue obtained through production divided among the factors of production. Or alternatively how much is paid to workers and how much is paid to the owners of capital for their contribution to production (the

Firms	Steel co.	Car co.	Legal co.
Revenues	**100**	**210**	**10**
Sales	100	200	10
Change inventories	0	10	0
Expenses	**80**	**150**	**10**
Wages	80	~~50~~ 40	10
Other inputs	0	~~100~~ 110	0
Taxes	**10**	**40**	**0**
Profits	**10**	**20**	**0**

Government	
Tax revenues	50
Wages (Police Services)	50

Consumers	
Wage Income	180
Profits distributed by firms	30

Figure 4: Calculating GDP: Changes in the productive structure II.

government will collect some of this income, taxes, from workers and firms). **Every dollar that is produced is a dollar of income earned by someone**.

In our example total **wages**, denoted by W, amount to \$180, **profits before taxes** and **interest payments** (and including depreciation, i.e. gross profits), denoted by Π, are \$80. The income measure of GDP is again \$260.[4]

$$Y = W + \Pi$$

Note 5: Rents of owner-occupied housing are included as (non-labor) **income** (they are also included in production, you can think of them as housing services, and expenditure, as the purchase of these housing services). When somebody rents an apartment, the rent he pays to the landlord becomes part of GDP. When the landlord rents a house to a tenant, the landlord is selling a service just like a haircut. Over 60% of Canadians own their houses and in a sense, they are their own landlords. Since counting renter-occupied housing as GDP and ignoring owner-occupied housing seems inconsistent, Statistics Canada calculates the amount that owner-occupied houses would rent for if they were rented out and imputes this amount as income for the owner.[5]

Note 6: Transfer payments are excluded from GDP. Social Security payments, unemployment insurance payments, subsidies to producers, interest payments on government debt... These transfers are simply reallocations of income (rather than new income) from certain agents to other agents. This is what happens with the (corporate) taxes paid by the firms in our example.

[4] In the case where taxes are zero and the government borrows to finance the police services, profits before taxes remain unchanged and therefore GDP from the income approach remains unchanged.

[5] In the national accounts, owner occupiers are deemed to be unincorporated businesses producing housing services (which they consume). The imputation is made so that the treatment of owner-occupied housing in GDP is comparable to that of tenant-occupied housing, which is valued by rent paid. This practice keeps GDP invariant as to whether a house is occupied by its owner or rented.

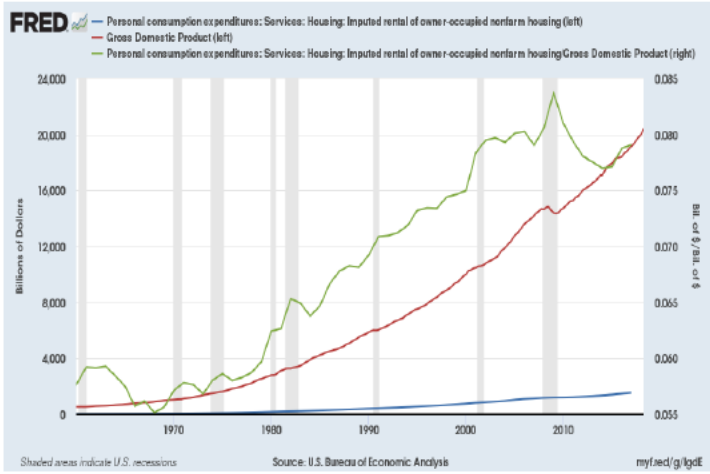

Figure 5: Housing services and GDP.

1.3 The expenditure approach to GDP

A third way of calculating GDP is from the **expenditure side**, i.e. **who buys** the **final goods and services** that are produced in the economy in a given year. This is probably the **most interesting calculation of GDP for modeling purposes**.

We can consider 4 broad categories of agents populating an economy:

1. **Households** or families (ultimate owners of the factors of production)
2. **Firms**
3. **Government**
4. The **rest of the world**

Any economic agent belongs to one, and only one, of these 4 categories.

Final goods purchased by **households** are (in general) classified as **consumption** (denoted by C), by **firms as investment** (denoted by I), by the **government as government purchases** (denoted by G). Investment goods (as opposed to intermediate goods) are those (final) goods purchased by firms that will be used in the production process for several years, such as machinery, buildings or structures. We will often refer to these goods that are used for the production of other goods but are not used up in the stages of production (as opposed to intermediate goods) as **capital**. Goods produced abroad and sold in Canada are **imports** (denoted by M) and goods produced in Canada and sold abroad are **exports** (denoted by X).

Then GDP from the **expenditure side**, can be written as,

$$Y \equiv C + I + G + (X - M) = C + I + G + NX$$

where NX stands for **net exports**.

To see this notice that any good purchased by the households has to be produced domestically, superscript d, or abroad, superscript m for imported. In line with this we can write consumption as $C = C^d + C^m$, investment as $I = I^d + I^m$, and government expenditure as $G = G^d + G^m$. As a result since all imports are sold to any of the three domestic categories total imports will be, $M = C^m + I^m + G^m$. At the same time a certain amount of domestic output, Y, will be exported, X, and therefore domestic production is,

$$Y = C^d + I^d + G^d + X.$$

Then

$$Y = C - C^m + I - I^m + G - G^m + X = C + I + G + X - (C^m + I^m + G^m)$$
$$Y = C + I + G + (X - M)$$

In our previous example the value of consumption is $200 worth of cars sold to families, investment is the $10 worth of cars that the second firms holds in the form of inventories and government expenditure is just the $50 worth of police services purchased by the government. Since there is no foreign sector exports equals imports equals zero. So GDP is again $260.[6]

Note 7: Purchases of **new constructed houses** are classified **investment**. A house (as a machine) will provide services for several years. In this sense, although houses are (usually) purchased by households, these purchases are considered as investment. In line with note 5 the rents of owner-occupied housing are included as consumption.

Note 8: Investment includes depreciation, i.e. the replacement of the capital goods that break down in the process of production is included as part of investment. Imagine that the delivery of steel (the steel co. is responsible for this) requires the use of 3 cars. On average these cars last for 3 years, so one car breaks down every year. Depreciation is the value of those capital goods (the broken car) that are lost (wear and tear) in the process of production.[7]

To illustrate this, assume a car is worth $10, when the old car breaks down the steel company buys the car that in the previous example remained in the inventory of the car firm.

Firms	**Steel co.**	**Car co.**
Revenues	**100**	**210**
Sales	100	~~200~~ 210
Change inventories	0	~~10~~ 0
Expenses	**80**	**150**
Wages	80	50
Other inputs	0	100
Taxes	**10**	**40**
Profits	**10**	**20**

Government	
Tax revenues	50
Wages	50

Consumers	
Wage Income	180
Profits distributed by firms	30

Figure 6: Calculating GDP: Investment and depreciation.

The calculation of GDP would remain unchanged, since the car bought by the steel company is not treated as an intermediate good but rather as a capital good, since it will contribute to production in the future.

Gross measures include depreciation while **net measures** exclude it. Since depreciation is reflected in final good prices, in the value of expenditure, and in the payments to the factors of

[6] In the case where taxes are zero and the government borrows to finance the police services, government expenditure remains unchanged and so does GDP from the expenditure approach.

[7] Houses, like any other capital goods, also depreciate.

production since profits are recorded before depreciation (income), it is incorporated when GDP is measured using the expenditure, production, and income approaches. Nonetheless, it is worth noticing that the **true value of new production available for** private or public **use** is **best captured** by **net** measures (i.e. the part of investment that is used to replace depreciated capital goods it is not really an increase in the amount of available goods and services).

Note 9: This note is related to note 2. Since **increases in inventories** are treated as **value added in the production approach** they are treated as **investment in the expenditure approach** and as revenues and therefore **profits in the income approach**.

The following two figures take a deeper look at the components of GDP from the expenditure perspective,

TABLE 2.9 Gross Domestic Product for Canada for 2004

Component of GDP	*$ Billions*	*% of GDP**
GDP	1290.2	100
Consumption	721.2	55.8
Durables	95.1	7.4
Semi-durables	61.5	4.8
Nondurables	177.6	13.8
Services	389.1	30.2
Investment	233.6	18.1
Nonresidential	142.3	11.0
Residential	83.6	6.5
Change in inventories	7.8	0.6
Government expenditures	281.6	21.8
Government consumption	248.5	19.3
Government investment	33.1	2.6
Net exports	54.3	4.2
Exports	492.6	38.2
Imports	−438.3	−34.0

Source: Statistics Canada, CANSIM database, Table 380-0002.
*Percentages do not add up to 100 because of rounding.

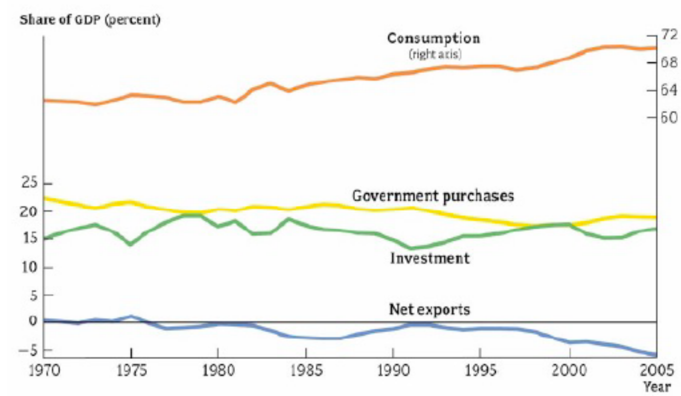

Figure 7: Canadian (left) and US (right) GDP.

The following diagram, the **circular flow of economic activity** for a closed economy, summarizes the flows of spending, inputs, and goods and services that make up GDP.

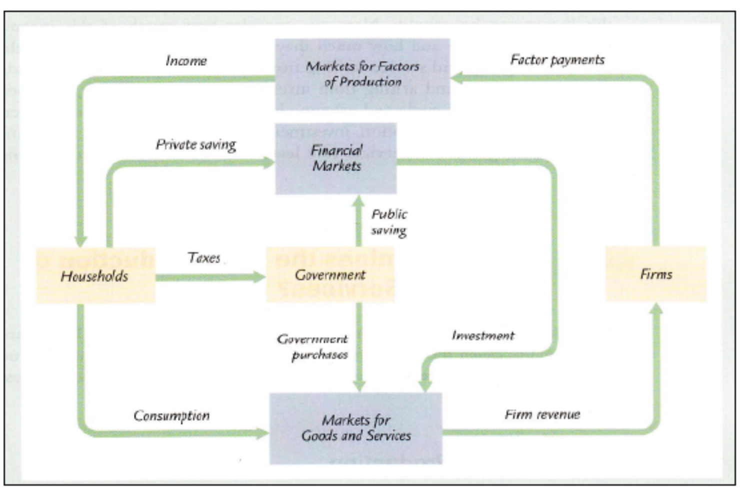

Figure 8: The circular flow of economic activity.

Finally, a related measure of economic activity is **Gross National Product** (GNP). GNP measures the value of output produced by domestic (Canadian) factors of production, whether or not the production takes place inside Canadian borders. For example, if an oil well in Venezuela is owned by Canadian residents, then the profits of this oil well will be included in Canadian GNP, but not in Canadian GDP.

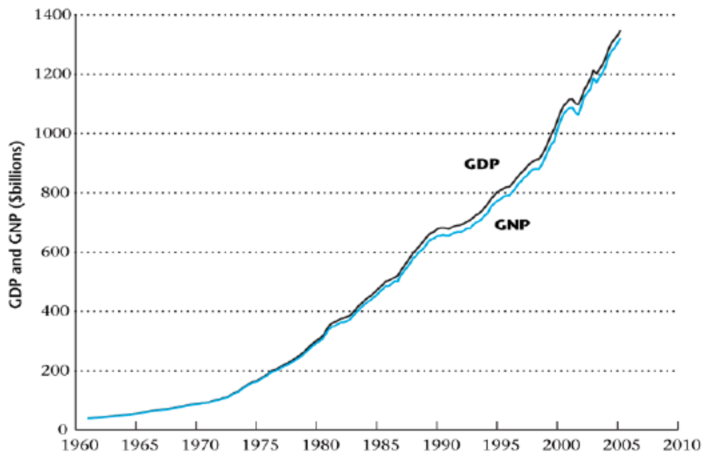

Figure 9: GDP vs GNP.

1.4 Comparing production across time: Nominal vs. Real GDP

Gross Domestic Product (GDP) is the **market value** of the final goods and services produced in a country over a period of time. To calculate GDP different goods are aggregated using their market prices, i.e. what people pay for these goods and services, which, in a sense, is (the best) a measure of the value they place on those goods and services. Comparing production through time is difficult since **both prices and quantities change** across periods.

Let's illustrate it with a simple example. Imagine an economy in which the only goods produced are apples and oranges. In year 1, 50 apples and 100 oranges are produced and the price of apples and oranges are $1 and $0.8 respectively. In year 2, 80 apples and 120 oranges are produced and the price of apples and oranges are $1.25 and $1.6 respectively.

TABLE 2.10 Data for Real GDP Example

	Apples	Oranges
Quantity in year 1	50	100
Price in year 1	$1.00	$0.80
Quantity in year 2	80	120
Price in year 2	$1.25	$1.60

Figure 10: Calculating Real GDP: An example.

Calculating nominal GDP it is straightforward here as there are **only final goods**. Multiplying prices by quantities each year and adding them up we reach,

$$NGDP_1 = Q_1^o * P_1^o + Q_1^a * P_1^a = \$130$$
$$NGDP_2 = Q_2^o * P_2^o + Q_2^a * P_2^a = \$292$$

Now we can calculate the percentage increase in nominal GDP between the two periods as (292-130)/130=1.25=125%, i.e. nominal GDP more than doubles.

Nonetheless this growth rate is **not very informative** since what **we are interested in** are the **changes in quantity** (production) from one year to the other, since it includes not only changes in the quantities produced but also changes in the prices of those goods. The solution is to value production in different years using constant prices, i.e. to calculate **real GDP**. For instance if we take year 1 as the **base year**, real GDP would be,

$$NGDP_1 = RGDP_1 = Q_1^o * P_1^o + Q_1^a * P_1^a = \$130$$
$$RGDP_2 = Q_2^o * P_1^o + Q_2^a * P_1^a = 80 * 1 + 120 * 0.8 = \$176$$

And the growth rate of real GDP is (176-130)/130=.354=35.4%. This growth rate is purged from changes in prices capturing the **increase in the quantity of goods** produced we are interested on. Any **change in real GDP** reflects a **change in actual production** (i.e. a change in the quantity of goods produced), since by construction we are **holding prices constant**.[8]

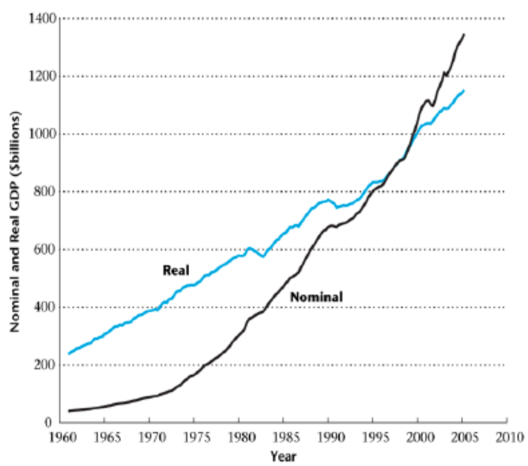

Figure 11: Nominal and Real GDP. Base year 1997.

Since the relevant measure of production is real GDP in general when we talk about GDP we **mean "real"** rather than "nominal".

There is a final issue when measuring production that is related to **changes in quality** of the goods produced. The quality of apples and oranges does not change much from one year to the next or over a decade. This is not true for many other goods included in GDP such as cars or computers. If the average computer produced this year is 10% "better" than the average computer produced last year and we are interested in the change in production between those two years, we

[8]Finally, notice that the choice of base year has an impact on our measure of real GDP and in its change over time and therefore if we use year 2 as our base year we will get slightly different results. In practice a geometric average of the prices of both years is used. The book discusses this approach.

should count each computer produced this year as the equivalent of 1.1 computers of the previous year. For instance, if both years we produce 100 computers the **quality-adjusted quantity** of computers produced this year is 100*1.1=110, while the quality-adjusted quantity of computers produced last yer is 100.

The next **question** is **how to measure "better"**. The approach used by statisticians to adjust for these quality improvements is to look at the market for computers and determine how this market values computers with different characteristics in a given year. Suppose that today, the price of a computer with a speed of 4 GHz is 10% higher than the price of an otherwise identical computer with 3 GHz. Since people are willing to pay 10% more for the additional GHz, one can say that an increase of 1 GHz increases the quality of a computer by 10%, or equivalently that a computer with 4 GHz is equivalent to 1.1 computers with 3 GHz. This approach is known as **hedonic pricing**. Some calculations suggest that for a given price, the quality of computers has increased, on average, by 18% per year since 1981. Along these lines a typical computer thirty years later, in 2011, provides roughly 143 times ((1.18)^30) more computing power than the typical personal computer in 1981. In this sense one computer in 2011 is equivalent to 143 1981-computers.

1.5 What do we want to measure with GDP? What is in it and shouldn't be there? What is missing?

Our goal when computing GDP is to provide a **measure** of the overall level of economic activity, but when we use it to make comparisons in a single country through time or across countries at a given time, we presume that it also captures something about the quality of life (**standard of living**) in those different societies or at different points in time.

GDP only includes activities that are **transacted in markets** and as a result a fraction of economic activity is not accounted for (household production, informal sector, illegal activities).

GDP does not capture the **depletion of natural resources** that takes place with the process of production. In the same way that net output subtracts from current production the "depletion" of capital goods, depreciation. An adequate measure of production, particularly in resource abundant countries, should be adjusted for the depletion of non-renewable natural resources.

GDP does not capture **pollution, or traffic congestion, or crime**, that are important for living standards. "Bads" are not included, only goods.

GDP does not capture the effects of **income inequality, health or literacy** in well-being beyond their impact on current production.

Although it seems that most people (and economists) use GDP per capita as a measure of how well a society satisfies the needs of its citizens this measure should be taken with **caution** given the previous limitations. As a result, over the last three decades or so, there has been a growing body of empirical evidence attempting to measure **subjective well-being** or **happiness**, its determinants and its connection with per capita GDP. Some examples of these measures are the Human Development Index constructed by the the UN or the Index of Social Well-Being from the OECD . These more comprehensive measures of well-being are **finding slowly their way** into serious **policy discussions** (read Frank (2008), listed in the syllabus).

Finally, figure (12) depicts a measure of life satisfaction against income per capita –on average countries with higher income per capita enjoy higher levels of life satisfaction and, in most cases, within a country life satisfaction also increases in income.

Figure 12: Life satisfaction and per capita income.

2 GDP: Where It Comes From and Where It Goes

2.1 What determines the Total Production of Goods and Services: The aggregate production function

An economy's GDP, the total production of final goods and services, depends on (1) the **quantity of inputs**, called factors of production, and (2) the ability to **turn inputs into output**, as represented by the production function.

We will focus on the two most important factors of production: physical **capital**, K, and **labor**, N. Capital, which is the result of a continuous process of investment, includes man-made means of production such as machines, plant, equipment, infrastructure,...[9] Labor is measured

[9] We should think about a broad definition of capital that includes, tools, machinery, buildings, public and private infrastructure... one of the key features of capital is its reproducible nature (**man-made**), capital as opposed to most natural resources could be produced. To think about aggregate production before the industrial revolution will require the introduction of **land** as an additional input, for modern capitalist economies this distinction seems unnecessary. For instance, in Britain, where long time series are available, the value of non-urban land as a percentage of total wealth has decreased from 65% at the beginning of the 18th century to 2% nowadays. For the U.S. in 1995 the share of oil purchases over GDP was 3.5% according to the Energy Information Administration,

as the number of workers. Furthermore, we assume these workers are homogeneous.

Households own both inputs of production. Most households are also workers, they own labor, and some households are the owners of the firms and therefore, indirectly, the owners of the capital owned by firms.

In the economy there is a large number, an aggregate, of firms that produce (and sell) a **composite good** that we call output or **GDP** and denote by Y. We can describe the technology accessible to this aggregate of firms using a production function, which describes the **technological possibilities for converting inputs into output**. The production function is given by,

$$Y = F(K, N) \tag{1}$$

The **factors of production** together with the **production function** determine the quantity of **goods and services produced**. For the moment let's assume the **inputs of production are fixed**, i.e. there are $\overline{N}$ identical workers who operate $\overline{K}$ identical units of capital.[10] Then the quantity of goods produced is given by

$$Y = F\left(\overline{K}, \overline{N}\right) = \overline{Y}$$

which in turn equals the economy's output, i.e. **GDP from the production approach**. Since we assume that the quantities of inputs are fixed, output is also fixed at a level $\overline{Y}$.[11]

The **marginal product of labor,** MPN, is the **extra** amount of **output** produced by an **extra** unit of **labor**, holding the amount of **capital constant**. Using our production function, (1), we can express the marginal product of labor as,

$$MPN = F\left(\overline{K}, \overline{N}+1\right) - F\left(\overline{K}, \overline{N}\right)$$

The first term on the right-hand side is the amount of output produced with $\overline{K}$ units of capital and $\overline{N}+1$ workers; the second term is the amount produced with $\overline{K}$ units of capital and $\overline{N}$ workers. The MPN is the difference between these two amounts of output. **Graphically**, the marginal product of labor is the **slope of the production function** when plotted in the labor space, i.e. holding capital constant. **Mathematically**, the marginal product of labor is the **derivative with respect to labor of the production function**.

Similarly, the **marginal product of capital**, MPK, which is defined accordingly, is given by the extra amount of output produced by an extra unit of capital, holding the amount of labor constant.

Up to this point, aside from the fact that output depends on capital and labor, we have **not placed any additional restrictions** on the production function. Next, on the ground of some **observations about real economies**, we will state some (desirable) **properties of the production function**. These properties we will place some **restrictions on the shape of** (1).

1. **More inputs yield more output**. For instance, if we were to consider a firm that writes computer software (output) using computers (capital) and programmers (labor), the previous

an agency of the U.S. Government. Finally, most firms incorporate into their production process a wide variety of **intermediate goods** produced by other firms. These other firms produce those goods using capital and labor, so those intermediate goods are just combinations of capital and labor. Technically our aggregate production function is a **value added production function** and value added already excludes any intermediate inputs.

[10] In principle production function can accommodate heterogenous capital goods that would be aggregated into the firm's capital stock using market prices.

[11] Intuitively you can think of the production function as being a recipe to make pancakes (output) combining eggs (labor) and flour (capital).

statement requires that the more computers the firm owns or the more programmers it hires the larger the number of lines of code it produces. I believe this requires little discussion.

2. There is some **"ideal"**, although flexible, **mix of inputs**. For instance, continuing with our previous example this ideal mix is, maybe, one computer per programmer. Let's say that a programmer-computer pair produces 20 lines of code per hour. If we add a second programmer to an existing computer-programmer pair output would increase, since more inputs yield more output. The second programmer can revise the code while the first one adds new lines, or can write code while the other goes to the restroom. Nonetheless, the additional lines of code resulting from the additional programmer (holding the number of computers constant) will be less than the ones written when there was a single programmer, for instance let's say that the two programmers-computer triplet produces 30 lines of code per hour. This discussion implies, that the **additional** amount of **output** produced, as we **increase one input**, **decreases** as the quantity of that input increases, holding the **other input constant**.

3. **If we double capital and labor output is doubled**. Continuing with the software company example, if the company purchased twice as many computers and hired twice as many programmers, it would, in principle, be able to produce twice as many lines of code.

Next, using these desirable properties we can place the **following restrictions** on the production function, (1).

"More inputs yield more output" implies that the **production function** is always **increasing**, or that its slope is always positive, or that **marginal products are always positive**, $MPN = F\left(\overline{K}, \overline{N}+1\right) - F\left(\overline{K}, \overline{N}\right) > 0$ and $MPK = F\left(\overline{K}+1, \overline{N}\right) - F\left(\overline{K}, \overline{N}\right) > 0$. These 3 statements are equivalent.

"The "ideal" mix of inputs" implies **diminishing marginal products** which in turn imply that the production function is **concave**.

The following figure depicts **graphically** these two properties focusing on labor (i.e., holding capital constant).

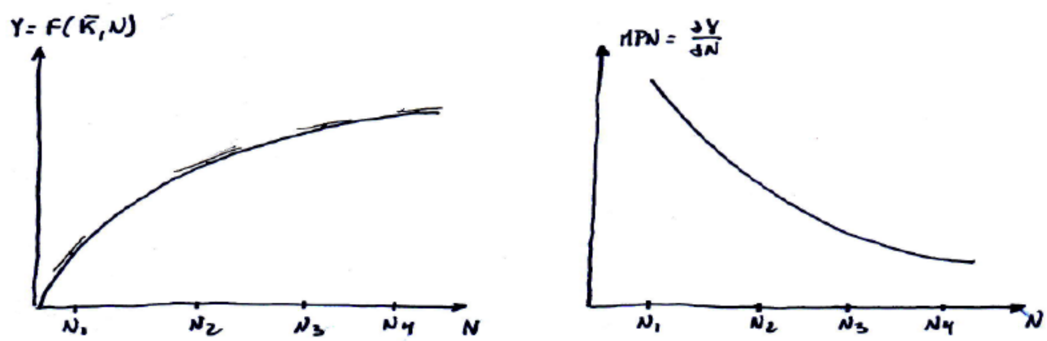

Figure 13: The production function and the marginal product of labor.

Notice that the first desirable property of the production function implies that the **marginal products are always positive**. This second (desirable) property implies that the marginal products are always **decreasing**.

Next, **"if we double the two inputs we double output"** implies that the production function exhibits **constant returns to scale** in capital and labor. Mathematically, the production function is homogeneous of degree 1 in capital and labor. Formally, this property of the production function can be expressed as,

$$F(cK, cN) = cF(K, n) = cY$$

where $c > 0$ is any positive constant, i.e. if we change both inputs by a proportion c, let's say 2, output changes by that same proportion, i.e. it doubles.

Additionally, if the production function is **constant returns to scale** in capital and labor, **we can always express** it as

$$Y = F(K, N) = MPK * K + MPN * N$$

Next, combining the last two properties of the production function –decreasing marginal products and constant returns to scale– we will derive an additional implication, the **complementarity between capital and labor** in production. This complementarity implies that **the more labor** (capital) the **firm uses**, the **more productive becomes its capital** (labor). Going back to the software example, we already know that adding one additional programmer to an existing computer-programmer pair would lead to an increase of 10 lines of code per hour. That is the marginal product of labor moves from 20 lines of code down to 10 lines of code when we move from one programmer to two programmers having only one computer. This is just an illustration of the diminishing marginal product of labor. Now, imagine that the firm equips the new programmer with an additional computer. How many lines of code would this new computer-programmer pair produce? Since all computers and all programmers are identical, they would produce 20 additional lines of code. This is just an illustration of constant returns to scale. What has happened to the **marginal product of labor** when we added the **new computer**? Well, it moved back from 10 to 20 lines of code.

Formally, the previous discussion implies that the **marginal product of labor** (capital) **increases in** the quantity of **capital** (labor), i.e. **capital and labor are complements** in production. Graphically,

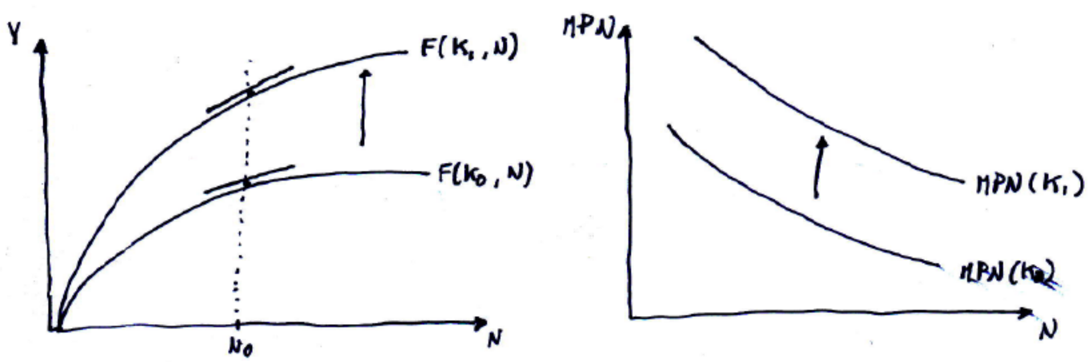

Figure 14: Capital and the marginal product of labor: Complementarity.

Notice that **complementarity** is **not an additional restriction** on the production function, but simply a **consequence of the other restrictions**, particularly of 2. and 3. That is, if the world is one where diminishing marginal products and constant returns to scale are prevalent, then the world is one in which capital and labor are complementary in production.

It is worth noticing that under constant returns to scale the **size of firms does not matter** for **aggregate outcomes**, i.e. it is equivalent in terms of production to have ten small firms or a single firm with ten times the capital and labor of those small firms,

$$\underbrace{F\left(\frac{K}{10},\frac{N}{10}\right)+...+F\left(\frac{K}{10},\frac{N}{10}\right)}_{10times}=\underbrace{\frac{1}{10}F\left(K,N\right)+...+\frac{1}{10}F\left(K,N\right)}_{10times}=F\left(K,N\right).$$

Since the size of individual firms does not matter we will use a **representative firm approach**, i.e. a single firm that uses all the capital and employs all the workers and produces all the output in the economy, summarizing all the choices made by firms in the economy. Nonetheless, this **artificial firm**, which is simply a sum of all the firms in the economy, **is competitive** in the sense that it takes prices (the price of capital, the price of labor and the price of its own product) as given. We will discuss more later about the competitive assumption.

Finally, the production function reflects the **available technology** at a point in time for turning **capital and labor** into **output**. Imagine someone **invents a better way to produce a good** resulting in **more output** for the **same amounts of capital and labor**. This will alter the production function. Thus, technological change alters the production function.

The first panel of the following figure depicts the **production function in a three-dimensional space** (capital, labor, output). The red plane in the second panel delimits a section of the production function **holding one of the inputs constant.**

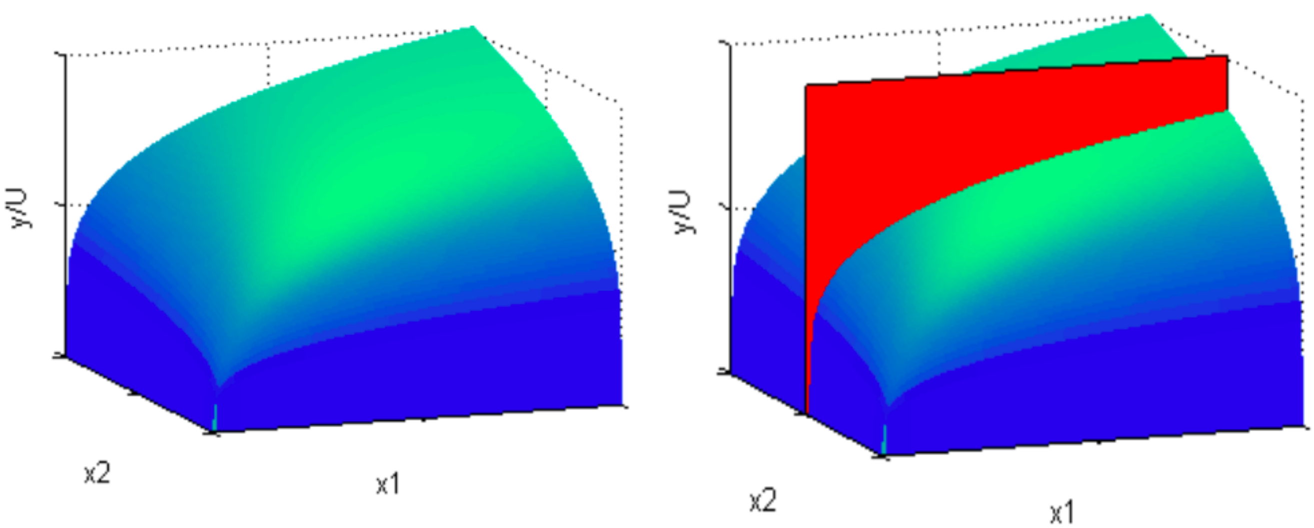

Figure 15: The production function in 3D.

Some **examples** of production functions (that may or may not satisfy the desirable conditions).

Linear production function (maybe the future, with robots and workers being perfectly interchangeable for production purposes): $Y = aK + bN$. In this case there is no ideal mix.

Leontieff production function (taxi services, before self-driving cars): $Y = min\left\{K, N\right\}$. In this case the ideal mix is very strict, exactly the same amounts of capital per unit of labor (or labor per unit of capital) with any excess being useless.

2.2 How is National Income Distributed to the Factors of Production?

So far, we have seen how **GDP** is determined through the **production approach**, i.e. how to model it using the aggregate production function. In this view, GDP is the amount of **output**

that results from transforming, by means of the **production function**, the available amounts of **capital and labor** into goods and services. Since the factors of production and the production function determine the amount of goods and services produced, GDP from the production approach, they also determine **GDP from the income approach**, national income.

Factor prices are the amounts paid to the factors of production –the **wage** the workers earn, i.e. the return per unit of labor, and the return the owners of capital collect, the **return per unit of capital**.

The standard assumption we make about the typical firm is that it is **competitive**. This assumption rests on the observation that most firms are **small relative to the markets** in which they trade, so basically have no influence on market prices –factor prices and the price of the goods and services they produce. Therefore, the competitive firm **takes the prices of its output and its inputs as given**, i.e. beyond its control. As we will see, when markets are competitive these **prices are determined** by the supply and demand for each good and factor, i.e. by the **joint actions of all market participants** (workers, owners of capital, producers, and customers).[12] If the overall demand for a good exceeds its supply, the price increases, if it falls short, the price decreases. In competitive markets **prices ensure** that **quantities supplied and demanded coincide.**

The typical firm rents capital and hires labor that transforms into output. As we did for the aggregate economy, we represent the relationship between inputs and output by means of the

[12] When you studied microeconomics, you learnt how prices are set, by supply and demand. But the supply and demand analysis is used for finding out the price of one good (for instance potatoes), assuming that you already know the price of all the other goods (apples, oranges, labor, machines, and everything else in the economy). What happens when you do not know the price of anything? What if you just have some "capital" and some "labor" around? Will a competitive market create prices that ensure that all capital and labor will be used? To make things concrete, consider a simple farm economy, with 100 workers and 10 identical farms owned by 10 owners. The only good produced and consumed are potatoes. Land (capital) and technology are fixed. Let's assume that production exhibits diminishing returns to labor and the additional output produced by the last worker (the 100th worker) is 5,000 potatoes per year.

Furthermore, assume that all workers are willing to work at any wage. It seems like a **recipe for exploitation**. Let's sum up

1. 100 workers that will work regardless of the wage.
2. 10 farm owners trying to hire workers (identical farms).
3. Diminishing returns to labor with the additional output of the last worker being 5,000 potatoes.

How are prices determined? Will all the workers be employed? Let's start with 10 workers per farm and assume that the owners meet in the pub and agree to pay 3,000 potatoes per worker. But by the time farmer #7 is back in his plot of land, he has figured out that if he hires one additional worker he will pay him 3,000 potatoes and get 5,000 potatoes, i.e. he can make a profit of 2,000 potatoes. So he tries to hire that additional worker. But where can he get it? Since the supply of workers is fixed at 100 he needs to hire it away from another farmer. He hires him away from the other farm by offering him 10 more potatoes per year. Of course all the other farmers have the same thoughts, they will hire from other farms bidding up the price of labor, the wage, 3,010, 3,020, 3,050... this happens **not out of the kindness of the owners** of farms but because they are **greedy**. The owners compete against each other, **acting in their individual self-interest**, and unintentionally **raising the wage** of the workers. The cycle continues until the wage is 5,000. At this point each farmer is content with the workers he has. The benefit of hiring one more worker is equal to the cost of doing so, i.e. the marginal product of labor is equal to the wage.

So we started off with an assumption – fixed supply of labor – that made it look like workers would be ripe for exploitation. But there are two sides to a fixed number of workers: It also means that business owners can only hire additional workers at the expense of other owners, bidding up their price, the wage. The wage depends on a fact of agriculture, of the nature of farming; the wage depends on how many potatoes can an additional worker produce. It does not depend on the kindness of the farmers or on how desperate workers are. This market we have just described is a competitive market.

In a competitive market, prices are not set either by individual producers or consumers but rather by their aggregate actions production function

$$Y = F(K, N).$$

This firm sells its output at a **price** P, hires workers at a **nominal wage** w^N, and compensates the owners of capital at a **nominal return** r^N for each **unit of capital** it rents. In general, we will use the superscript N for nominal variables while we will denote real variables without a superscript, that is w^N is the nominal wage while w is the real wage.

Notice that when we speak of **firms renting capital** we are assuming that **households own the economy's stock of capital**. Under this assumption, households rent out their capital, just as they sell their labor. In reality the ownership of capital is indirect because firms own capital and households own the firms so real-world **firms have two functions**: **owning capital** and **producing output**. Since we are interested in understanding how factors of production, in particular capital, are compensated it is helpful to assume that there exists an **explicit market where capital is rented out**. As a result, we assume that firms only produce output and capital is owned directly by households.

We assume the goal of the typical firm is to **maximize** (nominal economic) **profits**, i.e. the difference between revenues and costs.[13] Nominal economic profits are therefore given by

$$EconomicProfits^N = \underbrace{PF(K, N)}_{Revenues} - \underbrace{w^N N - r^N K}_{Costs} \tag{2}$$

Profits depend on the **product price**, the **factor prices**, and **factor quantities** hired by the firm (which, given the production function, determine its output).

Since the competitive firm takes all **three prices as given** it can only **choose** the amounts of **capital and labor** it hires. Notice that once capital and labor are chosen, output is also implicitly chosen. The competitive firm makes these choices in an attempt to maximize (2). But how does the **typical firm decide** how much labor or capital to hire?

Let us first consider the **quantity of labor the firm employs.** For this purpose, we assume the quantity of capital is fixed at a level $\bar{K}$. The **more labor** the firm employs, the **more output** it produces, and therefore, the **higher its revenues**. At the same time the more labor the firm employs, the **higher the wage bill** it needs to pay, and therefore, the **higher its costs**. Since the firm wants to maximize its profit by choosing the right amount of labor it should consider how the decision to hire an additional worker affects profits, i.e. it needs to compare the change in revenues with the change in costs associated with one additional worker. The **additional revenue** associated with one additional worker is simply $P * MPN$ while the **additional cost** is w^N, as a result the change in profit from hiring an additional worker is given by

$$\Delta EconomicProfits^N = P * MPN - w^N.$$

The firm would want to hire labor up to the point where the **increase in profits associated with an additional worker is driven down to zero**, i.e. up to

$$P * MPN - w^N = 0 \rightarrow MPN = \frac{w^N}{P} \equiv w \tag{3}$$

[13]As we will see with competitive markets any firm that does not maximize profits will be incurring in losses. These losses, which will be repeated period after period, will eventually drive the non-profit maximizing firms out of the market.

where w is the real wage, i.e. the payment to labor measure in units of output rather than dollars.

The firm takes the real wage as given and **hires workers up** to the point where the **marginal product of labor**, the units of output produced by an additional worker, are **equated** to the **real wage**, the units of output required to compensate this worker.[14] The following figure depicts graphically the optimal choice of labor, denoted by N^*. The first panels uses the production function, while the second panel uses the marginal product of labor.

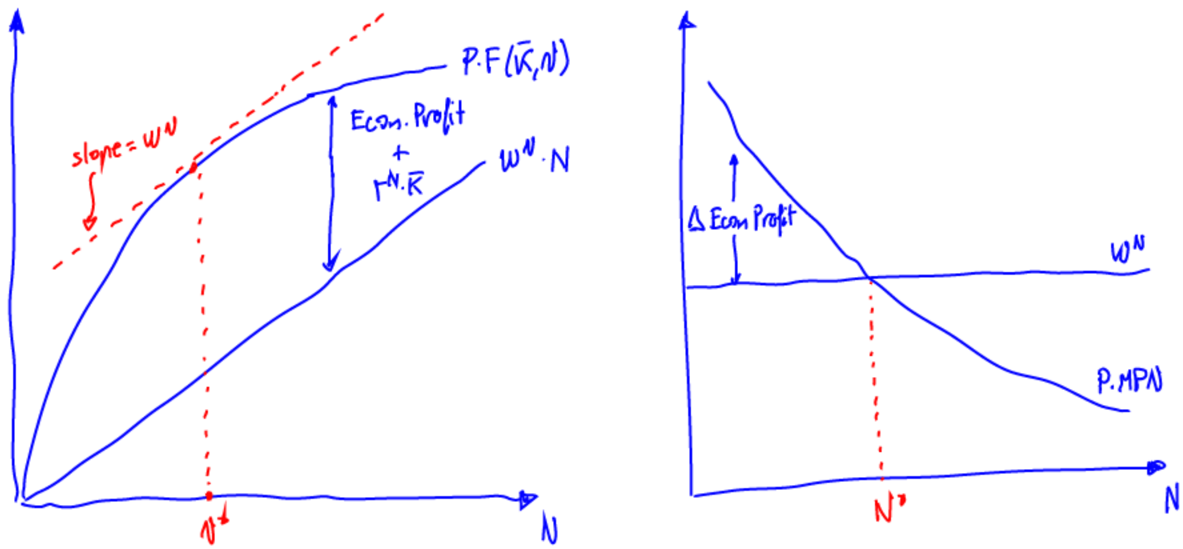

Figure 16: The optimal choice of labor by the firm.

A similar reasoning could be applied to capital reaching the following **renting rule for capital**

$$P * MPK - r^N = 0 \rightarrow MPK = \frac{r^N}{P} \equiv r, \tag{4}$$

which states that the typical firm will **rent capital** up to the point where the **marginal product of capital** is **equated** to the **real return to the owners of capital**, i.e. its nominal compensation expressed in units of output.

We can now explain how **competitive markets distribute income** between **capital and labor**. If **all firms** in the economy hire labor and capital in the same factor markets, all of them **are competitive**, and all of them **maximize profits**, then **what is the return paid** to each factor of production?

Well, conditions (3) and (4) suggest that each factor of production would be paid its marginal product. The **total real wages** paid to labor are therefore $w * N = MPN * N$. The **total real return to** the owners of **capital** for their contribution to production is given by $r * K = MPK * K$.

As we already know, what remains after firms have paid the factors of production their marginal products is the **economic profit**. As a result, economic profit **in real terms** is given by

$$EconomicProfits = F(K, N) - wN - rK,$$

[14]Notice that the fact that the production function exhibits diminishing returns to labor ensures that there is one optimal level of labor.

since we are interested in the **distribution of national income**, we rearrange the previous expression to reach

$$Y = MPN * N + MPK * K + EconomicProfits$$

so that **total income** is divided among the aggregate **return to labor**, the aggregate **return to capital**, and **economic profits** where everything is in real terms, i.e. expressed in units of output.

But, **how large** are these **economic profits**, i.e. how large is the fraction of output that remains after paying capital and labor their marginal products? If the production function has **constant returns to scale** then economic **profits are zero**. This follows from the fact that a constant return to scale production function can always be written as

$$Y = F(K, N) = MPN * N + MPK * K. \tag{5}$$

Finally, notice that if **factors are paid their marginal products**

$$MPN * N + MPK * K = w * N + r * K,$$

and therefore

$$Y = F(K, N) = w * N + r * K, \tag{6}$$

where the first equality is **GDP** from the **production approach** and the second equality is GDP from the **income approach**.

It is important to notice that (5) only requires **constant returns to scale**, while (6) additionally requires **competitive markets and profit maximization.**

Finally, if constant returns to scale, profit maximization, and competition are a good approximation to the real world, then **"profit" in the national income accounts**, i.e. accounting profit, is mostly the **aggregate return to capital**, i.e.

$$AccountingProfit = Profit = \Pi = MPK * K,$$

that is, **economic profit** is roughly **zero**.

2.3 The Cobb-Douglas Production Function

Throughout this course we will use a particular type of production function, the **Cobb-Douglas** production function,

$$Y = K^{\alpha} N^{1-\alpha} \qquad \alpha \in (0, 1) \tag{7}$$

where α is a measure of the **importance of capital** in production and $(1 - \alpha)$ is a measure of the **importance of labor** in production.[15]

Let's take a closer look at some of the properties of the Cobb-Douglas production function. The marginal product of labor and capital are given by,

[15]Formally α is the elasticity of output to capital, i.e. the percentage change in output for a given percentage change in capital, while $(1 - \alpha)$ is the elasticity of output to labor. Constant returns to scale require that both elasticities add up to one.

$$MPN = (1-\alpha) K^{\alpha} N^{-\alpha} = (1-\alpha) \left(\frac{N}{K}\right)^{-\alpha} = (1-\alpha) \left(\frac{K}{N}\right)^{\alpha} = (1-\alpha) \frac{Y}{N} \tag{8}$$
$$MPK = \alpha K^{\alpha-1} N^{1-\alpha} = \alpha \left(\frac{K}{N}\right)^{\alpha-1} = \alpha \left(\frac{N}{K}\right)^{1-\alpha} = \alpha \frac{Y}{K}$$

First, since both marginal products are always positive, then the Cobb-Douglas production function is increasing in each input, since $\alpha \in (0,1)$.

Second, it is clear from the expression after the second equal sign in each line of equation (8) that **marginal products are diminishing** since the exponents are negative on the relevant input. That is the marginal product of labor is decreasing in the quantity of labor and the marginal product of capital decreases in the quantity of capita.

Third, it is clear from the expression after the third equal sign in each line of equation (8) that **capital and labor are complementary** in production, i.e. that the marginal product of labor (capital) increases in capital (labor).

Finally, it exhibits **constant returns to scale** since,

$$F(2K, 2N) = (2K)^{\alpha} (2N)^{1-\alpha} = 2^{\alpha} 2^{1-\alpha} K^{\alpha} N^{1-\alpha} = 2K^{\alpha} N^{1-\alpha} = 2Y = 2F(K, N)$$

And as a result we can verify that output can be written as,

$$MPK * K + MPN * N = \alpha \frac{Y}{K} * K + (1-\alpha) \frac{Y}{N} * N = \alpha Y + (1-\alpha) Y = Y$$

using the expressions after the fourth equal sign in each line of equation (8).

Aside from satisfying all these requirements (increasing, diminishing marginal products, constant returns to scale) the **Cobb-Douglas production** function is consistent with the **empirical evidence of the distribution of income** between capital and labor as we will see in the next chapter.

To sum up, we have used **some desirable properties** of the production process to place **restrictions on (1)**, obtaining a **general characterization**. Then, we have introduced an explicit function, (7), that we will use in this course, the **Cobb-Douglas** production function.

3 The Price Level and Inflation

Denote an aggregate of all prices in the economy, the price level, by P. There are two commonly used measures of the **price level** in an economy. The first is the implicit **GDP deflator** $\left(GDP^{deflator}\right)$and the second is the **consumer price index** (CPI).

As we already know, **nominal** GDP refers to a measure of GDP when prices and quantities both vary from year to year (uses current prices) while **real** GDP refers to the actual quantity of goods and services (since prices are held constant). Since real GDP measures quantities and nominal GDP quantities and prices, from the ratio of both we can derive a measure of prices, known as the **GDP deflator**.

Continuing with our previous example (assuming year 1 is the base year where the GDP deflator is always equal to 100, since it is the ratio of two identical things),

$$\frac{NGDP_2}{RGDP_2} * 100 = \frac{Q_2^o P_2^o + Q_2^a P_2^a}{Q_2^o P_1^o + Q_2^a P_1^a} * 100 = \frac{292}{176} * 100 = 165.9 = GDP_2^{deflator}.$$

The second measure, the consumer price index, is **not as broadly based** as the implicit GDP deflator since it includes **only goods and services that are purchased by consumers**. Furthermore the CPI is an **expenditure-weighted index** (constant basket), which takes the quantities of goods consumed in some **base year** as being the typical basket of goods bought by the average consumer and then uses those quantities as weights to calculate the index in each year. Assuming families consume all the oranges and apples in our example and that year 1 is the base year, so that the basket coincides with production in year 1, the CPI in the second year for our previous example will be[16]

$$CPI_2 = \frac{CostBaseYearQuantitiesCurrentPrices}{CostBaseYearQuantitiesBaseYearPrices} * 100,$$

$$CPI_2 = \frac{Q_1^o P_2^o + Q_1^a P_2^a}{Q_1^o P_1^o + Q_1^a P_1^a} * 100 = \frac{50 * 1.25 + 100 * 1.6}{50 * 1 + 100 * .8} * 100 = \frac{222.5}{130} * 100 = 171.1$$

The CPI is the best measure of the **cost of living for an average Canadian** and it is calculated monthly by Statistics Canada. The current weights used by Statistics Canada are roughly; Food 17%, Shelter 27%, Clothing 5%, Transportation 20%, Health 5%, Recreation 12%, Alcohol and tobacco 3%, Others 11%.

Other expenditure-weighted index is the **Producer Price Index** (PPI), where the relevant basket includes purchases by firms.

Inflation is the **rate of change** or the percentage change **in the price level**, i.e. a measure of **how fast the overall level of prices in an economy rises** (or another way to think about it is **how fast does money lose its value**).[17] The opposite phenomenon is known as **deflation**. As a result, we denote inflation as the growth rate of prices, $g(P)$.

Inflation is simply the rate of change in the price level (either the $GDP^{deflator}$ or the CPI). Continuing with our previous example,

$$Inflation \equiv g(P) = \frac{P_2 - P_1}{P_1} * 100.$$

If we use the $GDP^{deflator}$, the associated inflation rate is 65.9% and we use the CPI the associated inflation rate is 71.1%.

The following figure depicts three, closely related, measures of inflation for Canada over the last half century (the additional measure is based on the CPI excluding food and energy).

There are **some problems** associated with these measures of inflation.

A first problem associated to our measure of inflation is related to changes in **relative prices**.[18] When there is a relative price change consumers typically will purchase less of the goods that have become more expensive and more of those that have become relatively cheap. In computing the CPI the implicit assumption is that consumers do not change their buying habits, i.e. they will always consume the same basket. This induces an **upward-bias** in the CPI which in Canada has been estimated to be in the order of **0.5%**.[19]

[16] Notice that, as before, the CPI in the base year is just 100.

[17] The price of apples has something to say about the value of apples, the price of computers has something to say about the value of computers, but the overall price level is different, rather than saying anything about the value of goods it says something about the value of money.

[18] Changes in relative prices also affect measures of Real GDP and therefore the GDP deflator. In a sense, our measure of Real GDP (and deflator) depends on the base year. Statisticians use chain-weighting to correct for this problem (see the book for a description of this procedure if you are interested).

[19] A potential solution is to derive a price deflator associated with chain-weighted real consumption expenditures.

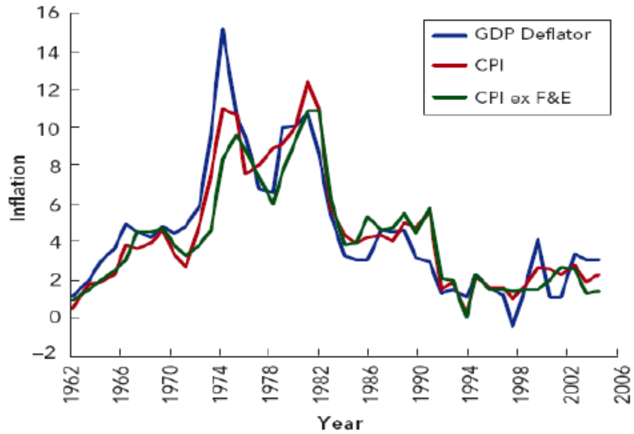

Figure 17: Inflation.

A second problem with our measures of prices (and GDP) is related to the **changes in quality** (or the **introduction of new goods**) that take place through time. As before **ignoring** the quality improvement introduces an **upward-bias on inflation** (and a downward one in real GDP) although most countries, including Canada and the US, adjust for quality with a method similar to the one described in the previous section.

These issues are relevant since in most countries, including Canada, salaries, pension benefits, and other government transfers are **revised** (or indexed) **using the CPI inflation figures**.

4 Nominal and Real Interest Rate

Interest rates are **inter-temporal prices**, in a sense they are the **price of time**, i.e. the price you need to pay to have something today rather than tomorrow. The **nominal interest rate**, denoted by i_t, is the price at which **money** can be shifted from the future to the present –borrowed today with a commitment to pay it back with interest in the future, i.e. the **time-price of money.** The **real interest rate**, denoted by r_t, is the price at which purchasing power or **goods** can be shifted from the future to the present, i.e. the **time-price of goods**.

Example: An economy where a single good is produced, the price of that good today is $100 and the price one year from now is $105, you get a loan from the bank of $1,000 and you have to repay $1,070 in a year.

What is the rate the bank changes you to have the money today (**nominal interest rate**)?

What is the rate the bank charges you to have purchasing power today (**real interest rate**)?

	t	t+1		
Price	$100	$105	5%	Inflation
Loan	$1,000	$1,070	7%	Nominal interest rate
Goods	10	10.2	2%	Real interest rate

Figure 18: Nominal and real interest rates.

In this example you borrow money at a rate of 7% and you borrow goods at a rate of 2% (i.e. next year you will have to return 2% more goods that you borrowed today).

Since we have **data** on the **nominal interest rate** (this are the ones charged by banks) and **inflation** (the growth rate of prices) let's see if we can exploit this data to **construct** a measure of the **real interest rate**. Along these lines let's express the real interest rate as a combination (function) of the nominal rate and inflation.

If you borrow \$1, at a nominal interest rate, i, then one year from today you will have to return of $\$1(1+i)$. If we denote the price level today and one year ahead as P_t and P_{t+1}, respectively. Then, the amount of goods the dollar buys today is $1/P_t$ and the amount of goods you will need to return a year from now will be $1\left(1+i\right)/P_{t+1}$. As a result, the (gross)[20] **real interest rate**, a measure of the **cost of shifting future** (one year ahead) **goods to today**, is given by,

$$\frac{Goods-to-pay-back-at-t+1}{Goods-received-at-t}=1+r=\frac{\frac{1(1+i)}{P_{t+1}}}{\frac{1}{P_t}}=\frac{(1+i)}{\frac{P_{t+1}}{P_t}}=\frac{(1+i)}{1+\frac{P_{t+1}}{P_t}-1}=\frac{(1+i)}{1+g\left(P\right)}.$$

Since $log\left(1+x\right)\approx x$ if x is a small number, then we can take logarithms in both sides of the previous equation to reach,

$$ln\left(1+r\right)=ln\left(1+i\right)-ln\left(1+g\left(P\right)\right)\rightarrow r=i-g\left(P\right)$$

which establishes a relation between the real interest rate, the nominal interest rate and inflation. This relationship is known as the **Fisher equation**.

If we **assume** that the value that people assign to having goods today rather than tomorrow, some sort of a **societal measure of impatience**, does **not change much**, i.e. that the **real interest rate** is relatively **constant**. Then the behavior of the **nominal interest rate** will be mostly driven by the behavior of **inflation**.

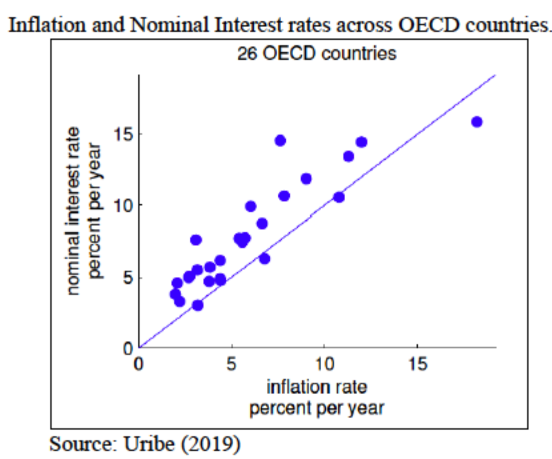

Figure 19: Inflation and Nominal Interest rates across OECD countries.

The previous figure plots the **average yearly nominal interest rate** from 1989 to 2012 for 26 OECD countries (each country is a dot) against the **average yearly inflation rate** in those countries together with the 45-degree line (which reflects a one-for-one change in both variables). The fact that most observations lie a couple of percentage points above this line suggests that **differences in nominal interest rates** mostly reflect cross-country **differences in inflation**

[20]Including repayment of the principal

rates, i.e. that there is **little cross-country variation in the real interest rate** (at least when we consider 20-year averages).

The following figure provides **additional evidence** of the importance of **inflation** as the **main determinant of the nominal interest rate**. The figure compares yearly inflation rates with yearly nominal interest rates for the US from 1955 to 2017. Again, it seems that years of particularly high inflation coincide with years of particularly high nominal interest rates. This relationship becomes tighter over long periods of time, i.e. if we were to take 10-year moving averages of the data rather than use yearly observations.[21]

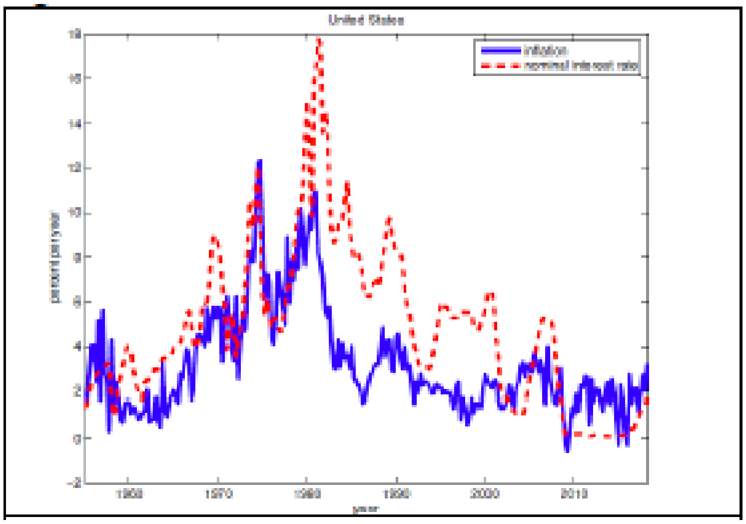

Figure 20: Inflation (solid/blue) and Nominal Interest rates (dashed/red) in the US.

Since to calculate the real interest rate the relevant inflation measure refers to the changes in prices **over the next year** economic agents need to form **expectations about the future evolution of prices**. In most cases we will assume that agents expect future inflation to be equal to inflation over the previous period. The following calculation of the **real interest rate** is done under this assumption.

Notice that although the rate charged by the bank is the nominal rate, the **relevant variable for economic choices is the real interest rate**. The real interest rate is of crucial importance for expenditure decisions (particularly for durable consumption, cars or fridges, and investment, since the revenues and costs associated with investment tend to happen at different points in time). For instance, a firm considering a new investment project, a new factory, will compare costs today (the costs of building and equipping the factory plus the costs of financing it) with the increase revenues coming from this factory (the additional sales at future prices). While the costs include the nominal interest rate, the additional revenues are determined at least partially by inflation, and, as a result, the relevant interest rate in this decision is the **real interest rate**.[22]

Although we keep talking about **the interest rate**, there are **many** interest rates for different maturities, default risks,... but all of them (usually) co-move.

[21] Assume your goal is to **understand the evolution of the nominal interest rate** over long periods of time. Given this evidence, what type of assumption would you make? i.e. at a first pass, what would you want to include in your model and what would you leave outside of the model? It seems that the previous figures suggest that inflation is a crucial determinant of the nominal interest rate.

[22] This is true even if the firm finances the new factory with internal funds, since the interest rate is the opportunity cost of those funds, i.e. the return the firm would get on those funds if it bought a treasury bill instead of building the plant.

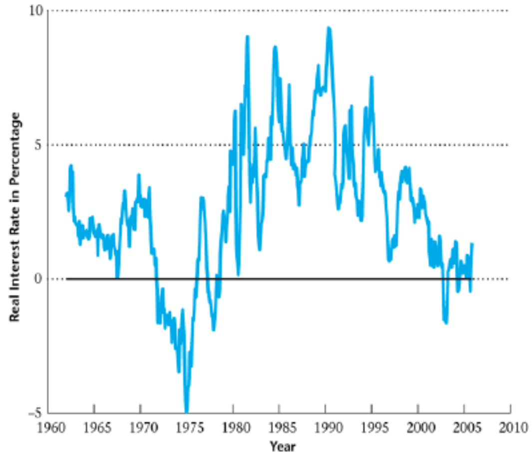

Figure 21: US Real Interest Rate (Nominal Interest Rate - (previous year) Inflation .

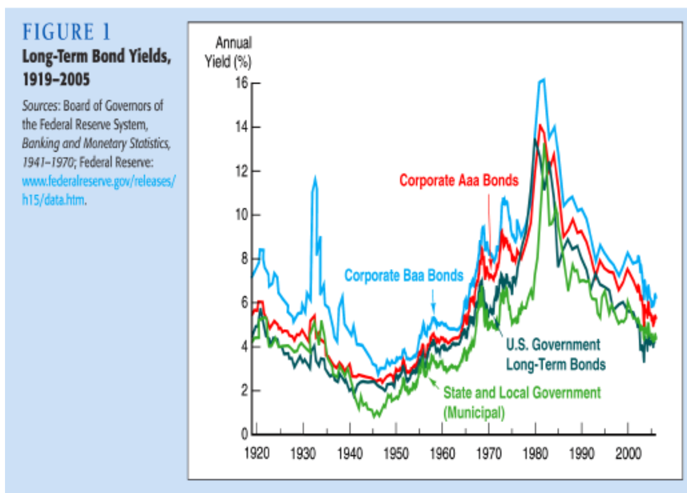

Figure 22: Several Measures of the Nominal Interest Rate.

5 The Unemployment Rate

One important aspect of economic performance is how well the economy uses its resources. Because an economy's workers are its main resource, keeping workers employed is a major concern of economic policy makers. In this subsection we will see how Statistics Canada measures labor market outcomes.

Let's denote by L, N, U and u the labor force, the level of employment, the level of unemployment and the unemployment rate. Someone is **unemployed**, if she/he has no job, wants to work, is **actively looking for a job** but has not yet found one. Statistics Canada interviews (monthly) 60,000 households in the Labor Force Survey. Statisticians classify these households into four categories:

1. Those who **were employed** in some sort of job when interviewed, N.
2. Those not working and who **did not want a job** immediately.
3. Those who **did want a job** immediately but had **not been looking for one** because they did not think they could find one (discouraged workers).

4. Those who **did want a job** immediately, **had been looking**, but had **not found** a job they would take, U.

The **labor force** is the sum of those **unemployed and those employed**. The unemployment rate is equal to the number of **unemployed** people **divided by the total labor force**.

$$L = N + U \qquad u = \frac{U}{N+U} = \frac{U}{L}$$

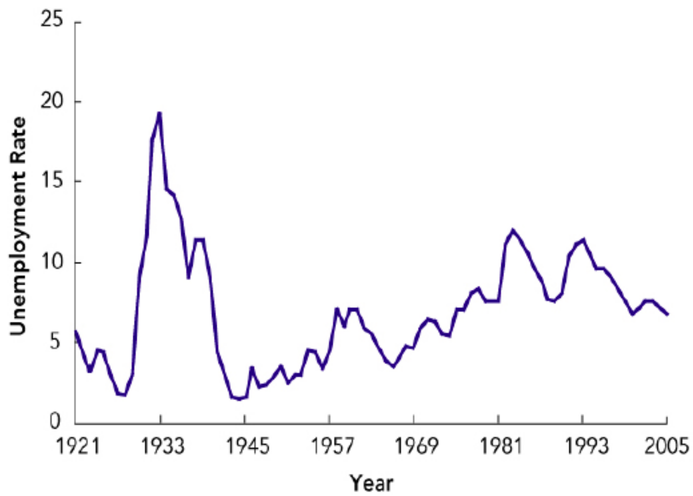

Figure 23: Canadian Unemployment Rate.

Until May of 2004 the University of Washington in Seattle employed me. In August of that same year I began to work for McGill. For those 2 months in between, I was "unemployed". It is an abuse of language to call me "unemployed" at the end of May, even though I really was unemployed according to the Statistics' Canada definition. I was unemployed because I was moving between jobs. This is an example of **"frictional" unemployment** –the fact that I was unemployed did not mean that I was suffering particularly, did not mean that the economy was malfunctioning in any sense, but was instead simply part of the normal process of people finding and changing jobs, just as the fact that every night a large number of full cargo containers are stacked up at the port of Vancouver does not mean that our system for distributing goods to consumers has broken down: "inventories" of goods and of labor that are not "in production" at any particular point in time are a necessary grease to keep the economy operating smoothly. In a dynamic economy, where new goods are introduced in the market displacing some existing products, **large numbers of jobs are created and destroyed in the course of normal economic activity** (in a normal month in the US around 3 million new jobs are created and 2.8 million jobs are destroyed) and therefore some frictional unemployment is not only inevitable, but also a sign of dynamism.

"The Mondays in the Sun" depicts the degrading effects of unemployment on a group of men left jobless by the closure of the shipyards in northern Spain. The only options left to these unemployed workers are to seek employment in local supermarkets or malls. But becoming cashiers requires very different skills than those in a shipyard; they need to be hyper-polite: say sir or madam; start from the belief that the customer is always right; take abuse from cranky customers without getting mad; sympathize. Cashiers need to follow directions: managers have little patience with people who won't do things their way, quickly, without arguing. Cashiers need to have good basic math sense: to figure out--quickly--when the amount of change that the machine is saying you should give this customer is clearly wrong; to be always running sums and differences in their heads to make sure that they have pressed the right buttons. Now spending twenty years working in the highly-skilled --specialized and dangerous-- occupation of shipbuilding is not going to predispose you to be a very good cashier. And if they don't have the possibility of adding value to a business by working in a mall--if the hassle to the manager and their wage are, together, a bigger cost than is the value of the work they do--it is hard to imagine what they are supposed to do. This type of unemployment is **structural unemployment.**

There is a third type of unemployment. In the 2007-10 economic conditions where economic activity is slow, you might loose your job. But you are pretty sure that your unemployment is not the result of low demand in your particular industry because the products it produces are no longer needed but rather because there is a general worsening of economic conditions. Switching to another industry would be pointless: by the time you got retrained and found a job, your old employer (or someone similar) would probably be trying to hire you again. This type of unemployment is known as **cyclical unemployment**. This type of unemployment is associated with the short-term fluctuations in economic activity, the **business cycle**.

The data shows that **most spells of unemployment are short**, i.e. in Canada around 60% of them end within one month, but at the same time 70% of the weeks of unemployment occurred in spells that lasted more than two months.[23] It sounds paradoxical but most people who become unemployed on any given day remain unemployed for only a short time; more than half find a job within a month. Yet of all the **stock of unemployed** on a given day, **around three-quarters of them will be unemployed for more than two months**. So the same households or families tend to feel disproportionally the costs of unemployment. Unemployment is very **concentrated**.

Unemployment in Central Canada (Quebec and Ontario) is more volatile than in other regions. The reason is the relatively large manufacturing sector in these areas, which is more exposed to business fluctuations.

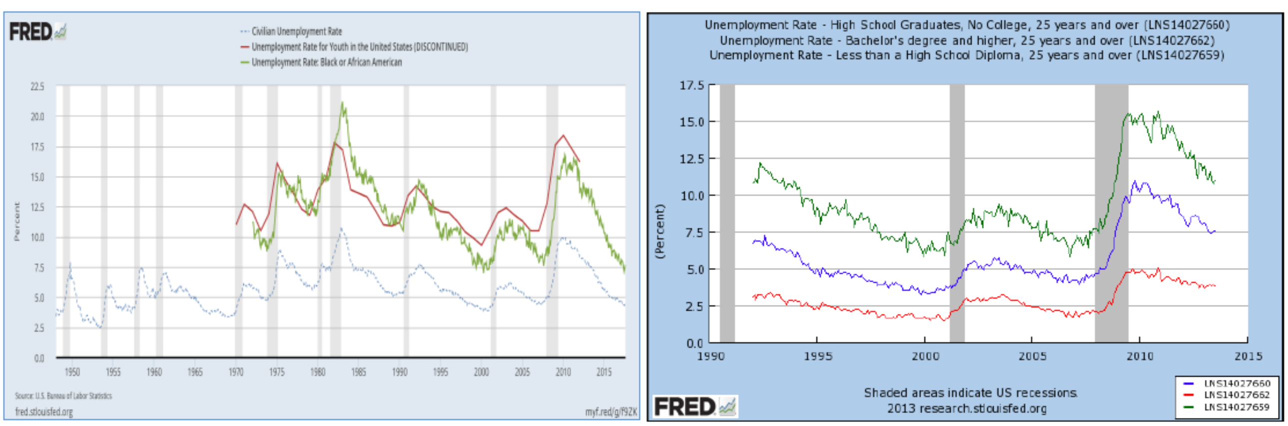

Figure 25: A closer look at US unemployment during the Great Recession II.

As you can see for the first panel in Figure (24), the **unemployment rate exhibits no trend**, it increases in contractions (shaded areas) and falls in expansions, fluctuating around a **level of 5% (structural + frictional unemployment)**. During the Great Recession the rate

[23] Imagine that over a year you have 10 people unemployed, 8 of them 1 month and the remaining 2 for 12 months. So there are 32 months of unemployment. 80% of the unemployed are so for 1 month, but 75% of the months lost in unemployment are concentrated in two guys.

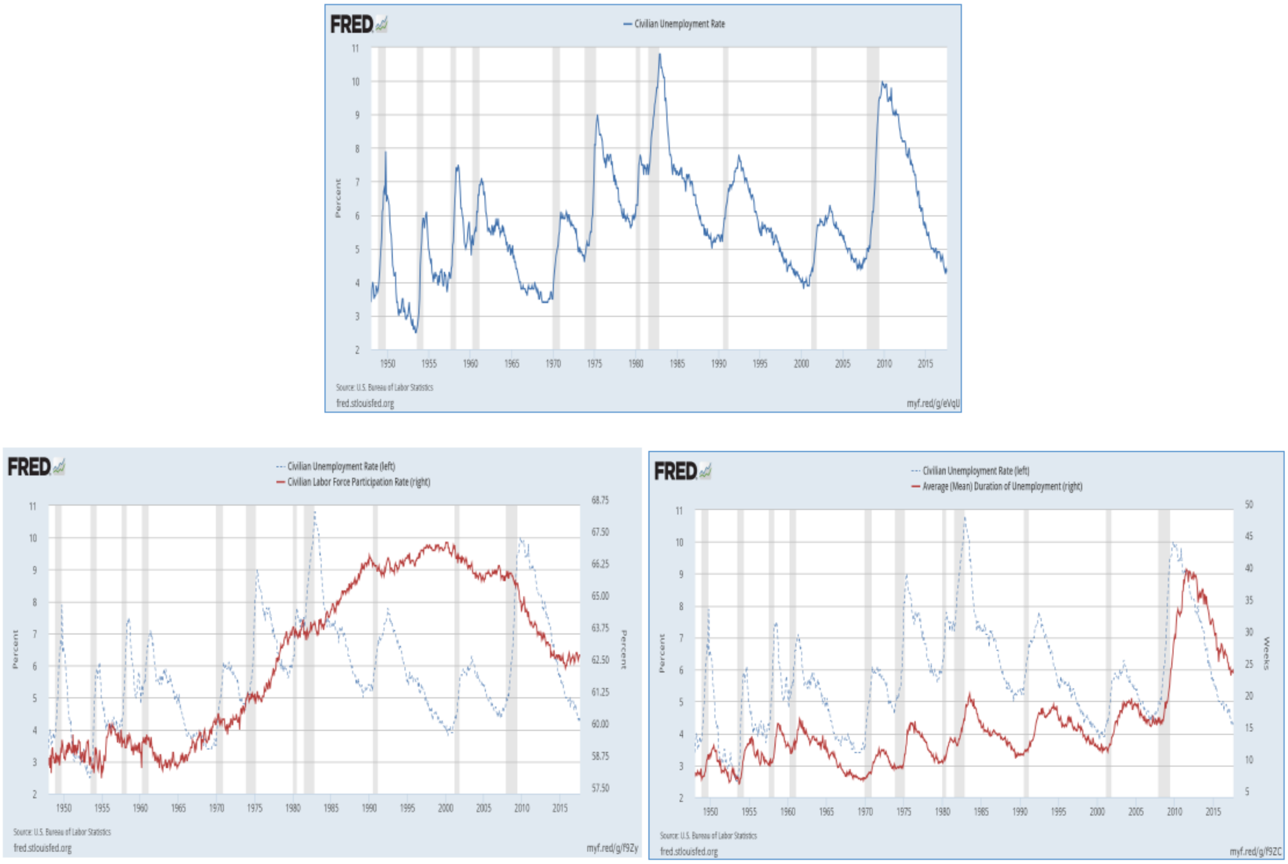

Figure 24: A closer look at US unemployment during the Great Recession I.

of unemployment climbed by almost **6 percentage points**. Using the unemployment rate as a sole indicator of the **labor market performance is quite misleading**. Although by 2012 the unemployment rate was already down to 6%, the **labor force participation** was still low by recent historical standard, **3-4 percentage points below** the post 1980s average. This suggests that the difficulties associated with finding a job have discouraged a large number of workers. Additionally, the length of unemployment spells has almost tripled, increasing from an average of 15 weeks to close to 40 weeks. In terms of socio-economic groups **unemployment** is always higher for **young** workers, **African American** workers, and **less educated** workers.

In the data, there is a clear relation between (the change in) **unemployment and GDP** (growth). Decreases in unemployment are associated with accelerations in GDP growth, while increases in unemployment are associated with slower GDP growth. This relation is known as **Okun's law**.[24] Although this relation varies depending on the country and the period under consideration, as a **rule of thumb** a **one percentage point increase** (decrease) in **unemployment** is associated with a **two percentage point decrease** (increase) in GDP **growth**. Figure (26) depicts two examples of this relation.

[24]Okun's law is named after economist Arthur Okun, who first identified and interpreted it in the 1960s.

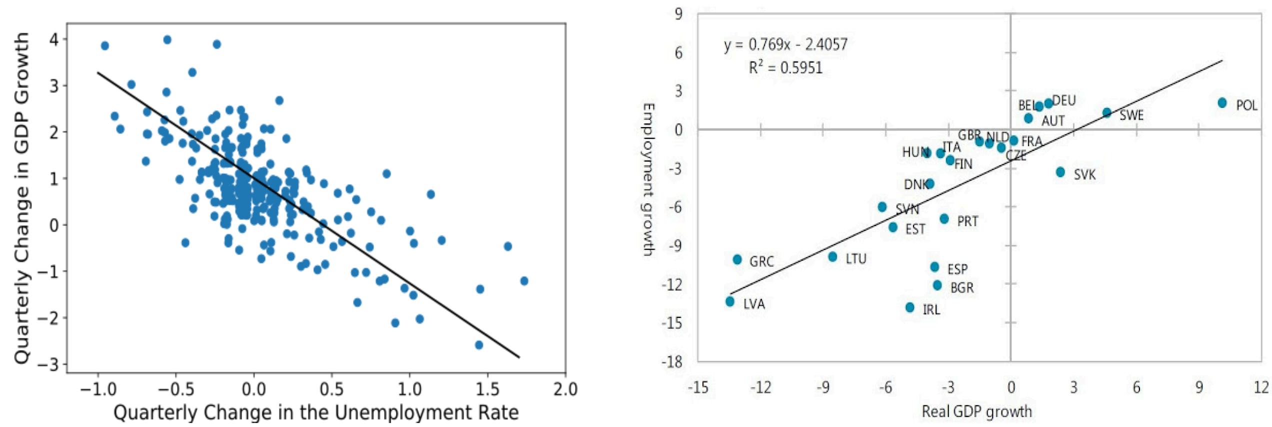

Figure 26: Okun's law for the US (1948-2016) and for EU countries (2008-2011).

Okun's law has a **simple implication**. If the current unemployment rate is *too high*, it will take a period of *above average* growth to reduce it. If, instead, the **unemployment rate is at the *right level*** then **output should grow at the *normal* rate** to keep unemployment constant. In this view, the unemployment rate provides a signal of where the economy stands and what growth rate might be desirable.[25][26]

Finally, there is also a relation between **unemployment and inflation**, the **Phillips curve**, named after William Phillips who first documented it using UK data from 1913 to 1948. When the unemployment rate is high, inflation tends to decline. When the unemployment rate is low, inflation tends to increase.

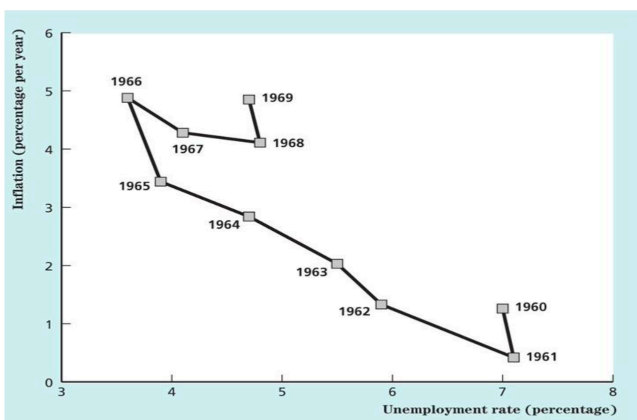

Figure 27: The Phillips Curve for Canada 1960-69.

At first sight, one may think of the Phillips curve as reflecting a **menu of choices** available to policy makers. This menu reflects a tradeoff between two undesirable outcomes, inflation and unemployment. If policy intends to reduce inflation, it will pay a cost in terms of higher unemployment. Conversely, if policy tries to reduce unemployment, eventually prices will rise faster. Nonetheless, as we will see, this relation is **far from mechanical** –varying substantially across time and countries– tending **to disappear** when long periods of time are considered.

[25] Notice that unemployment figures are reported monthly, while GDP figures, aside from being subject to subsequent revisions, are only reported quarterly.

[26] Okun's law can be written as $\left(\bar{Y}-Y\right)/\bar{Y} = c\left(u-\bar{u}\right)$ where $\bar{Y}$ is potential (full-employment) GDP and $\bar{u}$ is the natural rate of unemployment.

6 Money, bonds, stocks, and other assets.

Personal wealth measures the value of **all the assets** owned by an individual. This measure is determined by taking the total market value of all **physical and financial assets owned** and subtracting **all debts owed**. Examples of physical assets are land, houses, or capital. Examples of financial assets are currency, the balance of a checking account, bonds, or stocks. Examples of debts are the balance of a credit card or the outstanding amount on a mortgage.

Sometimes it is useful to think of wealth (or assets) as a **tool to shift income, resources, or consumption through time**. When I take a loan I am using today some money I will earn in the future –in this sense I am shifting income from the future to the present. When I purchase a house for rent (or for my own use) I am using current resources to generate a stream of future rental payments (or housing services for several years) –in this sense I am shifting income from the present to the future.

National wealth, wealth for short, measures the **value of all the assets in an economy**.

As before, it is useful to think of wealth as a tool for an economy to shift resources through time. In this sense, **physical wealth** will **produce output** in the future. Equipment, machines, inventories, plant, and land are different forms of physical wealth. These assets will be used for production in future periods. **Financial wealth** are **claims on someone else's future output**. A **bond** is a claim on the income/output of the issuer, **money** is a claim on the output of someone that is willing to accept it as a mean of payment (or ultimately on the government that is bound to accept it as a way to pay taxes), a **share of a company** represent the ownership of the means of production, physical wealth, of that company and therefore gives access to a fraction of the profits it generates, in the form of dividends or capital gains. Notice that future output will be determined solely by real wealth (and other future conditions). Finally, in a closed economy, the credit of any individual should have as a counterpart a debt for some other individual and, as a result, **financial wealth is in zero net supply**.[27]

Unlike GDP, a **flow variable** measuring the value of final goods and services produced in a given year, **wealth** is a **stock variable** that measures the amount of goods that have been accumulated up to a given point in time.

Figure (28) illustrates the time-series evolution of the composition of national wealth (for the US, left panel) and the cross-sectional composition of personal wealth (for France, right panel).

[27]In this sense there is an asymmetry between personal wealth, which can be negative, and national wealth which cannot be negative.

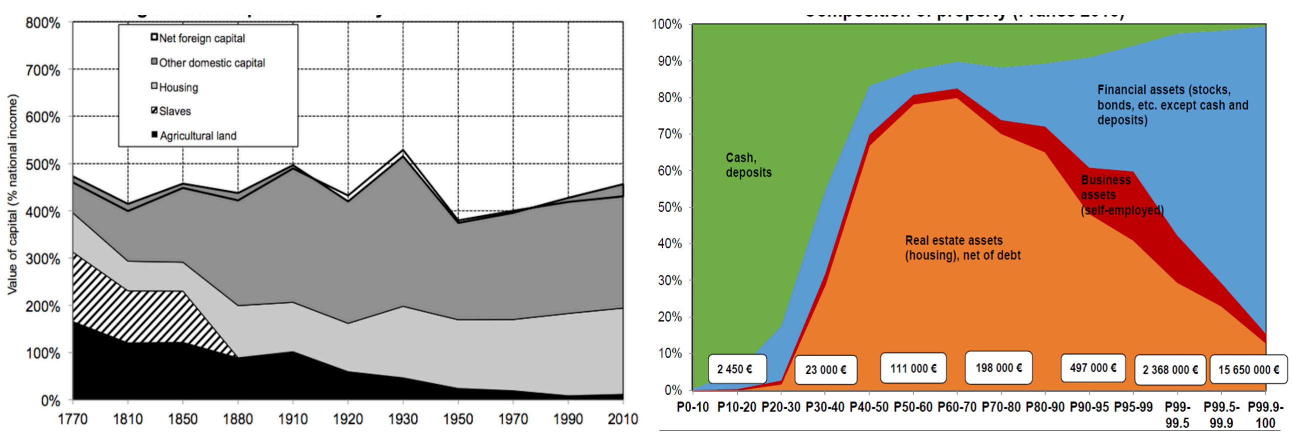

Figure 28: The composition of wealth.

Finally, personal wealth is **very unequally distributed**. For instance, in the US, the 20% wealthiest households own 85% of wealth. Figure (29) compares the evolution of the share of the top 0.1% of income and wealth –two measures of inequality. It is clear that wealth is much more concentrated than (personal) income, which in turn is also very unequally distributed. In addition, both measures of inequality have been increasing since the 1980s.

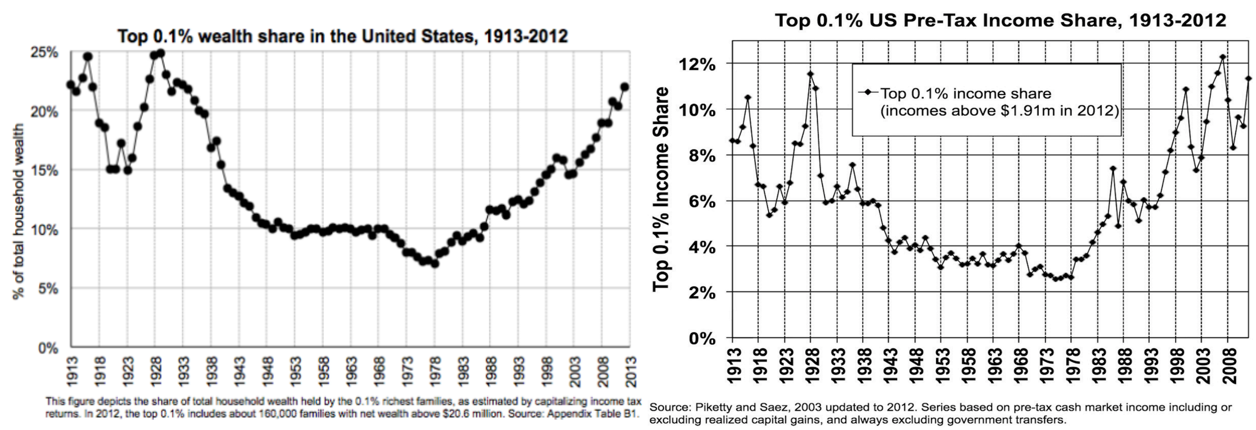

Figure 29: A measure of inequality.

Next, we will briefly discuss three important assets: money, bonds, and stocks.

6.1 Money

Money is whatever is **generally accepted as payment** for a transaction. Money fulfills three functions. Money is a **medium of exchange**, avoiding the double coincidence of needs characteristic of barter economies. Money is a **unit of account** –it is a yardstick used to measure the value of things. And money is a **store of value**, that is you can hold your wealth in the form of money.

Money has evolved from **commodity money** to **convertible money** (fractionally backed paper money) to purely **fiat money**.

We define money in a broad sense including not only bills and coins in circulation issued by the central bank –**currency**– and deposits of commercial banks in the Central bank –**reserves**– but also liquid deposits created by commercial banks –**bank money**. Currency and reserves constitute the **monetary base** which is the monetary aggregate directly controlled by the Central Bank. We will use M to denote the **quantity of money in the economy** –monetary base plus bank money. The following figure depicts the evolution of some monetary aggregates.

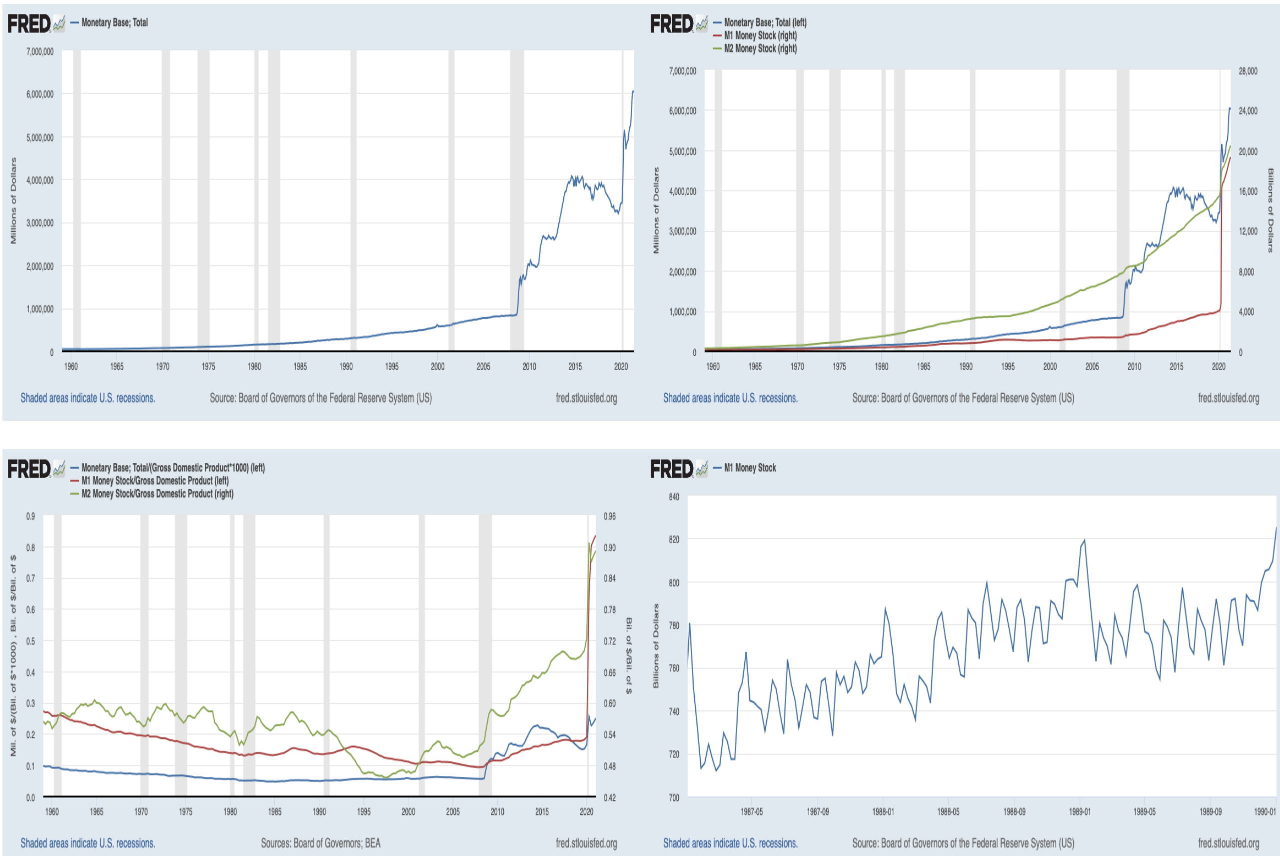

Figure 30: Monetary base, M1, and M2: US 1960-2020.

Commercial banks make profits through its lending and borrowing activities. The terms on which banks lend to households and firms differ from their borrowing terms. The interest they pay on deposits is lower than the interest they charge when they make loans, and this allows banks to make profits. Commercial banks play an **important role channelling** resources **from lenders to borrowers**, from those with resources available for investment to those with opportunities to productively invest them.

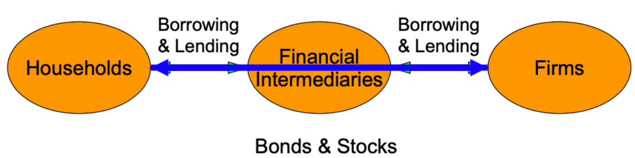

Figure 31: The “official” role of financial intermediaries.

Furthermore, by taking deposits and making loans, commercial banks provide the economy with a service of **maturity or liquidity transformation**. Bank depositors (individuals or firms) can withdraw their money from the bank without notice. But when banks lend, they give a fixed date on which the loan will need to be repaid. This allows borrowers to engage in **long-term projects**. The lenders' deposits are liquid (accessible on demand) whereas bank loans to borrowers are illiquid. In this sense commercial banks **transform illiquid assets**, loans, into **liquid assets**, deposits.

While maturity transformation is an essential service in any economy, it also exposes the bank to a new form of risk, **liquidity risk**, aside from the possibility that its loans will not be repaid, **default risk**. The following figure reproduces the consolidated balance sheet of the Canadian commercial banks.

TABLE 26-2 Consolidated Balance Sheet of the Canadian Chartered Banks, March 2018 (millions of dollars)

Assets		Liabilities	
Reserves (including deposits with Bank of Canada)	40 092	Demand and notice deposits	1 172 163
Government securities	267 598	Term deposits	658 307
Mortgage and non-mortgage loans	2 220 402	Government deposits	17 639
Canadian corporate securities	155 894	Foreign-currency liabilities	2 794 992
Foreign-currency and other assets	2 859 722	Shareholders' equity	319 194
		Other liabilities	581 413
Total	5 543 708	Total	5 543 708

Figure 32: Balance sheet of commercial banks.

Commercial banks finance roughly 5% of their assets with their own equity with the remaining 95% being financed through debt. This implies a **leverage ratio** close to 20, that is each dollar of capital buys twenty dollars worth of assets.

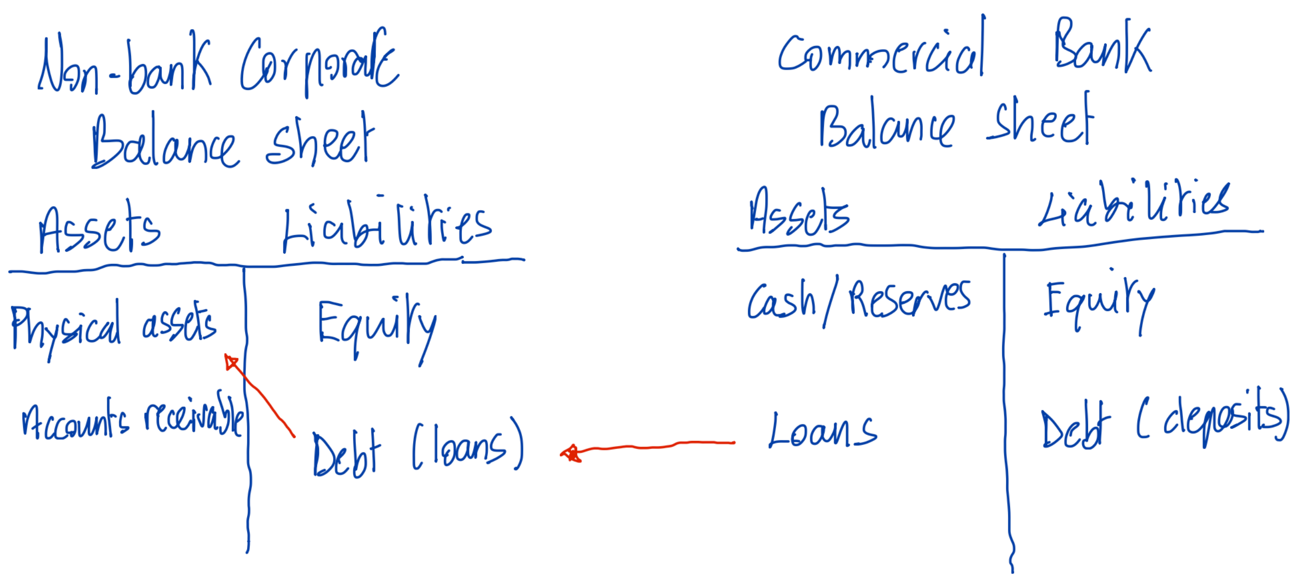

Figure 33: Balance sheet of non-bank and bank corporations.

Figure (33) sketches the balance sheets of non-bank and bank corporations illustrating the **connections between the financial and real sides of the economy**. In this view, it is not surprising that major **economic crisis** are often associated with (and amplified by) **financial crisis**.[28]

6.2 Bonds

Most **borrowing** for large corporations and the government takes place in the form of **bonds**, rather than loans. A bond is a **financial contract** that formalizes the promise of a borrower to **pay** at a pre-determined time **in the future** a certain amount of money (the principal, or **face value**, and interest or **coupon** payment) in exchange for certain **amount of money today** (face value).

Let's consider a one-year bond. Its price, P^B, should **coincide with the expected present value of the stream of income it generates**.[29] As a result, the pricing formula for a one-year bond is

$$P_t^B = \frac{C + FV}{1 + i_t},$$

where C is the coupon, FV is the face value of the bond, which always equals \$100, and i is the relevant (nominal) interest rate.

At the date of of issuance, since the coupon rate ($C/100$) equals the prevailing interest rate, the price of the bond coincides with its face value, that is \$100. For instance, if today the interest rate is 4%, a one-year bond issued today carries a coupon of \$4, and therefore its price would be,

$$P^B = \frac{C + FV}{1 + i} = \frac{4 + 100}{1 + 0.04} = \$100,$$

implying a rate of return or **yield equal** to the **interest rate**,

$$yield = \frac{C + FV - P^B}{P^B} = \frac{4 + 100 - 100}{100} = 4\%.$$

As opposed to loans, bonds are **tradable in secondary markets**. Since the coupon is constant and the price of a bond should always coincide with the expected present value of the stream of income it generates, then **bond prices change as the interest rate changes**.

If the day after the \$4 coupon one-year bond was issued the interest rate increases to 6% then the price of the bond decreases to

$$P^B = \frac{C + FV}{1 + i} = \frac{4 + 100}{1 + 0.06} \approx \$98,$$

[28] Let's put some numbers to illustrate how de-leverage works. Consider the following bank balance sheet: Cash 1, Loans 99, Equity 5, Debt 95. These numbers are broadly in line with the figures in the consolidated balance sheet of commercial banks in Figure (32). Now, consider a shock to the non-bank corporate sector that increases the default rate to 2% of the outstanding loans, i.e. around 2. This reduces by 2 the value of the assets of the commercial banks, let's say Cash 0 and Loans 98. On the liability side we have Equity 3 Debt 95 and if banks try to keep their leverage ratio roughly constant they will need to decrease their outstanding loans by 40% (if they want to keep a leverage ratio around 20, they need to reduce their debt to 57, which implies that their loans should fall from 99 to 60). This eventual reduction in new loans will reduce investment (physical assets) and production.

[29] Why? If the price exceeds this present value, no one will be willing to buy the bond and its price will drop. The converse is true if the price falls short of the present value. As a result, the price is one that coincides at any point in time with the expected present value of the stream of income associated with the bond.

Intuitively, if the prevailing interest rate is 6% a one-year bond with a \$4 coupon will trade at \$98, so that the coupon $(C = 4)$ plus the capital gain $\left(FV - P^B = 100 - 98 = 2\right)$ generate a rate of return equal to the interest rate. As a result, the yield always equals the interest rate and (existing) **bond prices** move in the **opposite direction** than the **interest rate**. Up to adjustments for default risk, maturity risk and the like, yields of different bonds coincide and should be equal to the interest rate.

Finally, similar pricing formulae can be derived for longer term bonds, for instance the price of a three year bond is given by

$$P^B = \frac{C}{1+i} + \frac{C}{(1+i)^2} + \frac{C+FV}{(1+i)^3}.$$

The **yield curve** plots yields of bonds with increasing maturities. Figure (34) depicts the US yield curve towards the end of 2019.

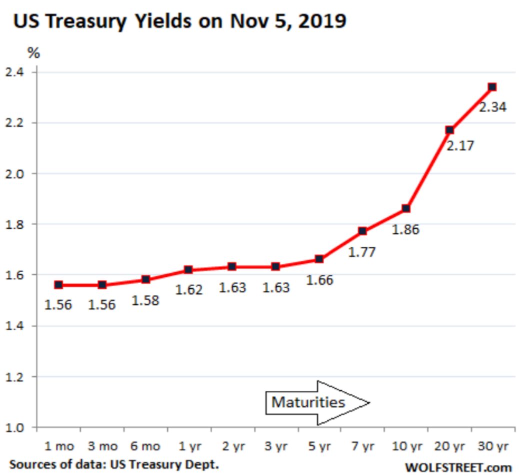

Figure 34: The yield curve.

Economists think of **long-term interest rates** as a weighted average of **current and (expected) future short-term interest rates** –the expectations theory of the interest rate. For instance, if the interest rate today is 5%, and people expect it to remain there for one year and then rise to 10% and remain there. This theory postulates that the yield on a two-year bond should be 7.5%, so investors are indifferent between buying a two-year bond or two consecutive one-year bonds. Additionally, since longer-term bonds are "riskier" than short-term bonds, they carry a risk premium and, as a result, the **yield curve (usually) slopes upwards**.

6.3 Stocks

Aside from loans and bonds, corporations finance their activities issuing shares, **stocks.** The price of a stock, P^S, **"should"** reflect the **expected present value of the stream of income it generates**, including dividend payments, D, and capital gains (changes in the future price) as a result,

$$P_t^S = E_t\left[PV\left(D_t, D_{t+1}, ..., D_{t+k}\right) + PV\left(P_{t+k+1}^S\right)\right]$$

where $E[]$ and $PV()$ stand for expectation and present value respectively.

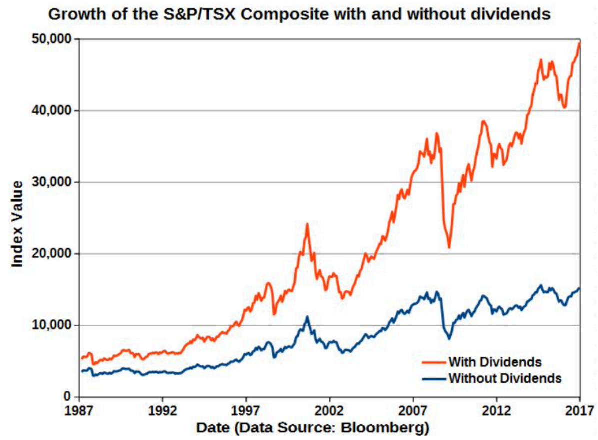

Figure 35: The S&P/TSX Composite.

7 Open economy

7.1 Accounting: The Balance of Payments

The **balance of payments accounts** summarize **Canadian transactions with the rest of the world**. There are two main components in the balance of payments; the **trade account (current account)** balance and the **capital account (financial account)** balance.

Balance of Payments Account = Bal. of Trade Account + Bal. of Capital Account

First, the **trade account** balance is the difference between the sales of Canadian goods and services to the rest of the world, exports (X), and the Canadian purchases of goods and services from the rest of the world, imports (M). The difference between exports minus imports is a **trade surplus** if positive and a **trade deficit** if negative.

Second, the **capital account** balance that is the difference between the sales of Canadian assets (firms, real estate, private or public debt, Canadian currency) to the rest of the world (M^K or capital inflows or borrowing from the rest of the world) and the purchases of assets from the rest of the world (X^K or capital outflows or lending to the ROW). An example of a capital inflow is when a Japanese mining company buys a mine in Alberta. A capital outflow is, for instance, the building by a Canadian manufacturing company of a plant in Bangladesh or the purchase by a Canadian resident of shares of a US company or foreign bonds. The difference between capital imports (sales) minus capital exports (purchases), if positive denotes that Canada is **borrowing** (a capital account surplus, when you sell more assets than you buy) and, if negative it denotes that Canada is **lending** (a capital account deficit, when you buy more assets than you sell) from the rest of the world.

Finally, as a result of accounting conventions, the **balance of payments is always zero**, i.e. the balance of trade account always equals the negative of the balance of capital account,[30]

[30] A more detailed exposition of the balance of payments is as follows,
Current account (trade)
Exports of goods and services +
Imports of goods and services -
Primary et factor income (wages and profits) +/- (+ if you receive income) Less than 5%
Secondary net factor income (transfers) +/- (+ if you receive transfers) Less than 0.05%
Capital account
Sales of assets (inflows) +

Balance of Payments Account = Bal. of Trade Account + Bal. of Capital Account = 0.

This point is best understood looking at a **simple example**: Assume a world populated by two individuals; Domestic, D, and Foreign, F. Both grow apples. D grows green apples and F grows red apples. None of them like the color (or taste) of their own apples so they need to trade to be able to consume apples. For the sake of illustration let's assume apples trade one by one. This year D collects 5 apples and F has a crop of 7 apples. As always they trade them. Let's take D's perspective. She exports (sells) 5 apples and imports (buys) 7 apples, so she is running a **trade deficit** of 2 apples. An alternative way to look at this transaction is from the asset perspective. D is borrowing 2 apples from F, i.e. D is selling an asset (IOweU, i.e. a right over D's future apples) in exchange for receiving from F 2 additional apples today. This asset will deliver 2 apples sometimes in the future. In this view, D is running a **capital account surplus** of 2 apples. **If you buy more than you sell** (i.e. a trade account deficit) **you need to borrow more than you lend** (i.e. a capital account surplus).

In terms of GDP, given that $Y = C + I + G + (X - M)$, if Canada is running a trade deficit, $(X - M) < 0$, then it should be true that domestic (Canadian) expenditure exceeds domestic production, $Y < C + I + G$. As a result, like anyone who spends more than she/he earns, Canada needs to be borrowing from the rest of the world.[31],[32]

For a country as a whole, **openness in financial markets** has an important implication. It **allows** the country to run **trade surpluses and deficits**. A country running a trade deficit is buying more from the rest of the world than it is selling to the rest of the world and must borrow the difference.

Trade in assets (rather than goods and services) dominates international transactions. For instance, in 2005 the recorded daily volume of foreign-exchange transactions in the world was \$1.9 trillion, of which 90% –about \$1.7 trillion—involved US dollars on one side of the transaction. To get a sense of the magnitude of these numbers, the sum of US exports and imports amounts to \$9 billion per day, or about 0.5% of the daily volume of dollar transactions in the foreign exchange market. This computation yields a simple conclusion: **Most international transactions** are associated **not with trade but** with purchase and sales of **assets**. Furthermore, the volume of foreign-exchange transactions in New York doubled between 2001 and 2005. Again, this activity reflects an increase in financial transactions rather than an increase in trade.

Purchases of assets (outflows) -

Change in reserves of foreign currency +/- (- if your reserves increase)

[31] There are alternative, but equivalent, interpretations of a trade imbalance. If $NX = X - M > 0$, net exports are positive, the country runs a trade surplus, its output exceeds its spending, the country holds a positive net foreign investment position, or the country runs a capital account deficit, or the country experiences a net capital outflow, or the country is lending abroad.

[32] Let's look at a simple example to illustrate how both balances, trade and capital, add up to zero. Imagine a world made of two countries: Canada and Europe. Assume one \$ buys one Euro and vice versa. A given year Canada sells production worth \$5 to Europe and buys European production worth Eu7. That same year Europeans buy \$10 worth of Canadian shares, while Canadians buy Eu2 worth of European shares. The balance of trade is the difference between the sales of goods and the purchases of goods, i.e. 5 − 7 = -2, i.e. Canada runs a trade deficit. The balance of capital, includes not only the transactions of shares but also the change in foreign reserves in the Bank of Canada. Let's calculate this item. In order to purchase goods and shares abroad Canadian residents need to buy Eu9 form the Bank of Canada. Now, the Bank of Canada will buy Eu15 from those Canadians selling goods and shares abroad. Overall the foreign reserves of the Bank of Canada (or the amount of Euros in Canadian hands) has increased by 15 - 9 = 6. This increase in foreign reserves is accounted as a purchase of foreign assets. As a result the balance of capital which is the difference between sales of domestic assets and purchases of foreign assets is given by 10 − (2 + 6) = 2, i.e. a capital account surplus. As you can see the balance of payments equals zero, BoP = -2 +2 = 0.

7.2 The saving-investment identity

We can write down **GDP** from the **expenditure side** as

$$Y = C + I + G + (X - M) \tag{9}$$

All income is received by households (they own the firms and therefore they receive the profits), then total household income, which coincides with total production, since **national income equals national product**, is also denoted by Y.

Households use their income to pay taxes, T. **Disposable income** is defined as income after taxes (and transfers), $Y^d = Y - T$. Households can either spend, C, or save, S, their disposable income. As a result

$$Y^d = Y - T = C + S \rightarrow Y = C + S + T \tag{10}$$

Combining both equations, (9) and (10), and rearranging,[33]

$$C + S + T = C + I + G + (X - M) \rightarrow \underbrace{\underbrace{S}_{Private-saving} + \underbrace{(T - G)}_{Public-saving}}_{Domestic-saving} + \underbrace{(M - X)}_{Foreign-saving} = \underbrace{I}_{Investment}$$

So we reach a **very useful identity**, that states that total savings should coincide with investment,

$$Saving \equiv Investment$$

Finally, notice that the saving-investment identity is silent about **what causes what**. It simply states that saving and investment must be equal.[34]

Sample Exam Question (Difficulty: Medium). Use until next lecture to sketch an answer and we will solve it together in class.

Suppose the government deficit is 10, tax revenues are 40, consumption expenditures are 80, the current account surplus is -5 and national saving (private plus public) is 20. Calculate GDP

[33] To be more accurate we should also include in the left-hand side of (1) and (2) the net factor payments from abroad to Canadian residents and instead of trade account balance we should talk about current account balance. This is related to the previous and next footnotes.

[34] There is an alternative way to derive the saving-investment identity that treats explicitly net factor payments from abroad (i.e. payment streams associated with the ownership of assets abroad). In order to calculate private saving we begin with private disposable income, $Y^d = Y + NFP - T$ where Y is GDP, NFP are net factor payments from abroad to Canadian residents and T are taxes. Private saving is simply (defined as) the difference between disposable income and consumption, $S^{Private} = Y^d - C = Y + NFP - T - C$. Then we can calculate government saving as the difference between tax revenues and government expenditures, $S^{Public} = T - G$. If we add both measures of saving we reach domestic saving, $S^{Domestic} = S^{Private} + S^{Public} = Y + NFP - T - C + T - G = Y + NFP - C - G$ that using the expenditure decomposition of GDP becomes $S = C + I + G + (X - M) + NFP - C - G = I + (X - M) + NFP$ and since the current account deficit is given by $CA^{Deficit} = (M - X) - NFP$ (if positive) we have $S^{Private} + S^{Public} = I - CA^{Deficit}$ or $S^{Private} + S^{Public} + CA^{Deficit} = S^{Private} + S^{Public} + S^{Foreign} = I$. Notice that in this case net factor payments from abroad enter in domestic private saving (with a plus) and in the current account deficit (with a minus), so this equation is equivalent to the one in the main text. The same analysis could be extended to allow for government transfers and interest payments on outstanding public debt.

7.3 Prices

In an open economy there is an additional relevant price, the **exchange rate**. The **nominal exchange rate**, e, is the price of foreign currency in terms of domestic currency (i.e. how many Canadian \$ are required to buy one Euro, $CAN\$/Eu$, or the price you pay when you buy Euros in your bank). This definition implies that an **increase in the exchange rate** is a **depreciation** of the domestic currency, i.e. a decrease in the value of the Canadian \$ relative to the Euro.

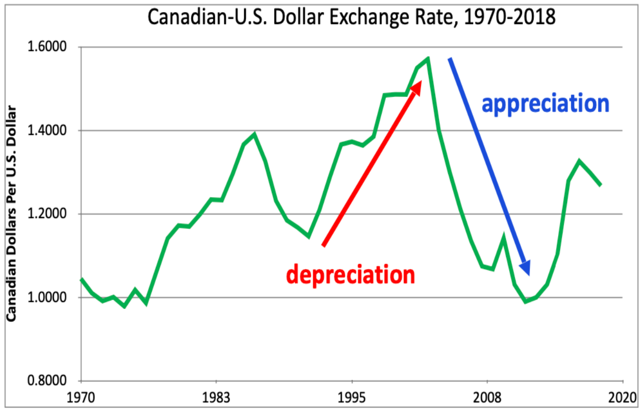

Figure 36: CAD-US\$ exchange rate.

Nonetheless for international transactions it is also relevant the **difference in prices across country**, expressed in a common currency. Denoting by P^* the price level abroad, that is the price of foreign goods in units of foreign currency, the **real exchange rate**,

$$\varepsilon = e\frac{P^*}{P}$$

is the price of foreign goods in terms of domestic goods, this becomes clear when we express each element in terms of its units as

$$\varepsilon = e\frac{P^*}{P} = \frac{CAN\$}{Eu}\frac{Eu/EUgood}{CAN\$/CANgod} = \frac{CANgood}{EUgood}.$$

Similarly, a **higher real exchange rate**, a real depreciation, implies that **foreign goods become more expensive relative to domestic goods**.

7.4 The equilibrium nominal exchange rate

Like any other price, the **nominal exchange rate** is determined by the supply and demand of the good in question, in this case the **supply and demand** of the **Canadian \$** for international transactions. Canadian \$ are supplied by domestic residents who want to buy foreign goods, who want to have a vacation abroad, or that want to buy foreign assets. Similarly, foreigners interested in Canadian goods, vacations or assets determine the foreign demand for Canadian \$.

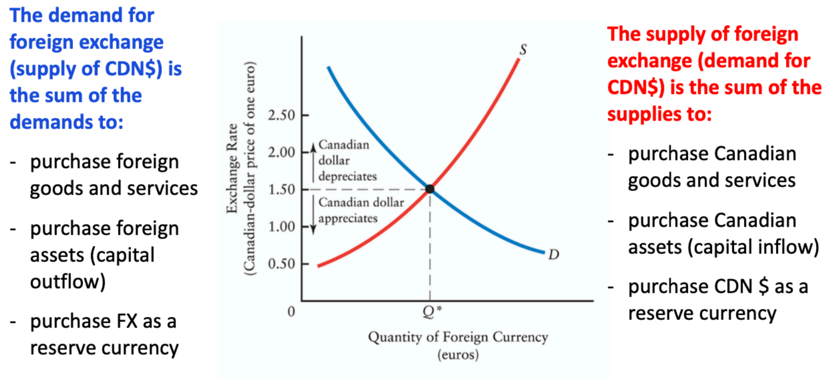

Figure 37: The determination of the exchange rate: supply and demand for CAD.

This suggests a **couple of approache**s to evaluate the **"correct" level for the nominal exchange rate** or its likely future evolution. The first approach is based on the current account, i.e. **trade of goods and services,** while the second is based in the capital account, i.e. **trade of assets**.

For the sake of illustration let's focus in the value of the Canadian \$ in terms of US \$. Let's denote this exchange rate as $e^{CAD/US\$}$ and remember that it is the price of one US\$, i.e. the number of Canadian \$ you need to pay of one US\$.

The first approach is on a **non-arbitrage condition for goods** and services, known as the **law of one price.**[35] If this law holds between Canada and the US then the price of Canadian goods should coincide with that of US goods expressed in Canadian dollars, i.e.

$$P^{CAN} = P^{US} e^{CAD/US\$} \tag{11}$$

The **intuition** for (11) is simple if Canadian goods are more expensive than US goods evaluated in Canadian dollars, $P^{CAN} > P^{US} e^{CAD/US\$}$, agents in both countries can make **risk-less profits buying goods in the US and selling them in Canada**. On the one hand this will increase US prices relative to Canadian ones. On the other hand, since buying US goods requires US dollars, Canadians will have to sell Canadian dollars increasing the supply of Canadian dollars for foreign transactions. As a result the Canadian dollar will depreciate, an increase in $e^{CAD/US\$}$. Both forces –the increase in the US price and the depreciation—work to restore the previous inequality.[36]

According to this first approach, **differences in prices** across countries are helpful to **determine the nominal exchange rate**, while **differences in domestic inflation rates** should be informative about the evolution of the exchange rate. Log-differentiating the law of one price we have

$$g\left(P^{CAN}\right) = g\left(P^{US}\right) + g\left(e^{CAD/US\$}\right) \rightarrow g\left(e^{CAD/US\$}\right) = g\left(P^{CAN}\right) - g\left(P^{US}\right) \tag{12}$$

that suggests that the **Canadian dollar depreciates** whenever **Canadian inflation exceeds US inflation**.

[35] A non-arbitrage condition simply states that if there are price differences for the same good across different locations there will be arbitrators that will buy in the cheap place and sell in the expensive one. As a result of these actions those price differences will tend to disappear.

[36] The presence of transport costs, tariffs, or market power are some of the reasons that lie behind the deviations from the law of one price.

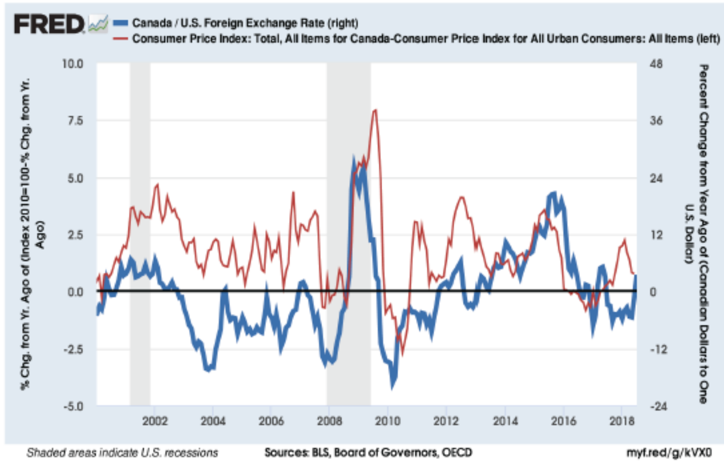

Figure 38: Nominal exchange rate and inflation differential: US and Canada.

The figure above compares the rate of change in the nominal exchange rate, $g\left(e^{CAD/US\$}\right)$, with the inflation differential between both countries. In line with the theory both variables move at unison. Nonetheless, the co-movement has become weaker in the last years of the sample.

The second approach is based on a **non-arbitrage condition for assets**, known as the **uncovered interest-rate parity**. This condition requires that the (expected) rates of return, expressed in a common currency, coincide in both countries. If this is not the case international investors will only buy assets in the country that yields a higher (expected) return. This condition requires that

$i^{CAN} = i^{US} + E\left[g\left(e^{CAD/US\$}\right)\right] \rightarrow E\left[g\left(e^{CAD/US\$}\right)\right] = i^{CAN} - i^{US}$

where E is the expectation sign.[37] If the Canadian interest rate exceeds the US interest rate, international investors must expect that the Canadian dollar will depreciate enough so that the decrease in value of the CAD relative to the US$ compensates for the Canadian interest rate advantage.

The previous figure compares the Canadian nominal exchange rate with the interest rate differential between Canada and the US for 10-year government bonds.

[37] It might be useful to plug some numbers in the interest-rate parity condition to verify how it works. Imagine we have $i^{CAN} = 10\%, i^{US} = 5\%, e_t^{CAD/US\$} = 1$. If international investors are to be indifferent between buying Canadian and US bonds they will have to expect the Canadian dollar to depreciate by 5% over the term of those bonds, i.e. $e_t^{CAD/US\$} = 1.05$. A Canadian \$ invested in Canada yields 10 Canadian cents. A Canadian \$ invested in the US yields also 10 cents (of CAD), since with one CAD you buy one US\$ today, that yields US\$1.05 at maturity that converted back into CAD is CAD1.1025.

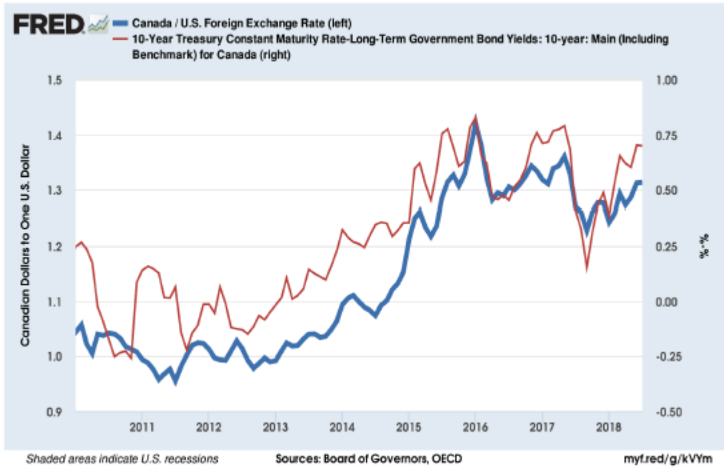

Figure 39: Nominal exchange rate and interest rate differential: US and Canada.

8 Applications

8.1 Trade and capital accounts in a two-country world: A coincidence

There is a narrative of European sovereign debt crisis that followed the 2007 financial crisis that traces back its origins to the excessive borrowing of the **countries in the South**. The basic story is that these countries were **living beyond their means**. Since this is not sustainable on a permanent basis it will eventually create all sort of troubles, as it did in in 2010. Consider a two-country EU (or for the sake of it, a two-country world) with Spain and Germany. In 2008 Spanish per capita real GDP was roughly 85% of its German counterpart. So our world is made of two relatively rich economies. In line with the previous narrative, has **Spain been borrowing** (and consuming) **too much**? The following figure reproduces the stock of debt to GDP for Spain. It almost doubled during the 2000s.

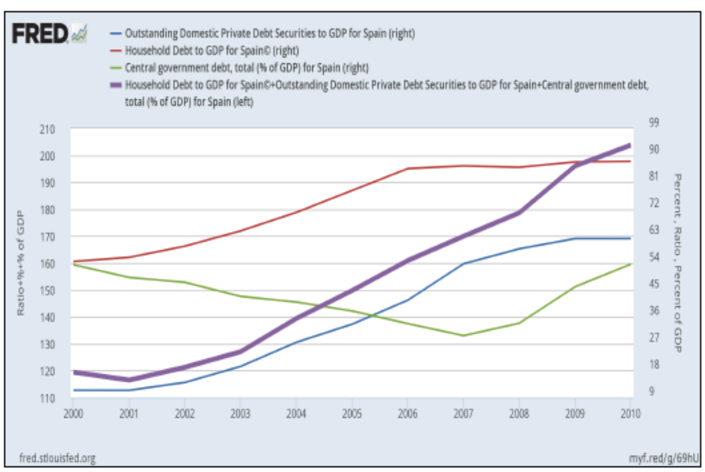

Figure 40: Several Measures of Debt: Spain.

By 2010 the Spanish private and public sectors had accumulated debt equivalent to two-year worth of production. Our previous discussion of the trade and capital accounts would suggest that **Spain should have been running important trade deficits** over the years leading to 2010.

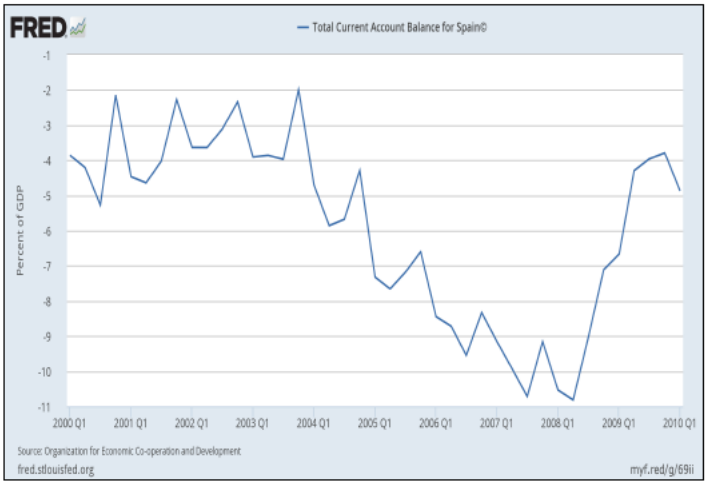

Figure 41: Spanish Current Account Balance.

And as the previous figure suggests, Spain has been running an average trade deficit of 6% per year over that decade. But if Spain is systematically running trade deficits, someone else should be systematically running capital account deficits to finance the excess of expenditure over production in Spain.[38] In our two country world, it has to be Germany. The following picture plots together the trade balance for our two countries (a coincidence).[39]

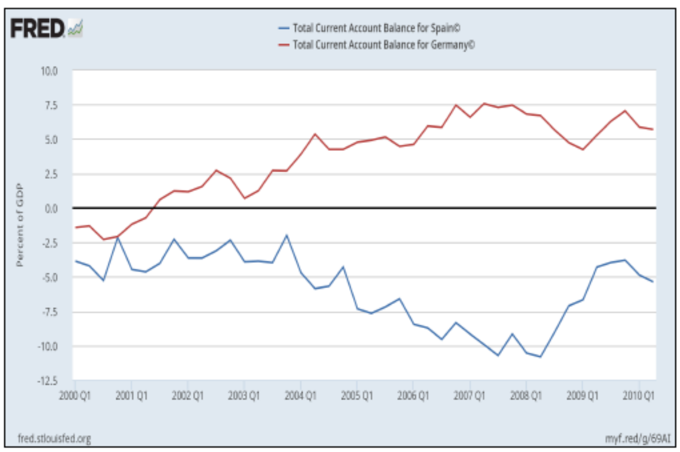

Figure 42: Current Account Balance: Spain and Germany.

This analysis suggests an **alternative interpretation of the imbalances** that lead to the sovereign debt crisis. Instead of the problem being that Spain has been borrowing too much, the problem could have been that **Germany has been spending systematically too little**, well

[38]Notice that this figure overstates the total debt of Spain with the rest of the world. Since most of the securities issued by the private and public sector are held as assets by Spanish residents.

[39]It is a coincidence since with more than two countries the bilateral trade account is not informative of the bilateral capital account. For instance, imagine we extend our world to include Japan. Spain runs a trade deficit with Germany of 100. Germany has a trade deficit with Japan of 150. Japan has a deficit with Spain of 50. What do we know about the capital flows? We know Spain is borrowing 50, Germany is borrowing 50 and Japan lending 100. But the trade deficit of Spain with Germany of 100 does not imply that Germany has a capital account deficit of 100 with Spain. Notice that I use current and trade account interchangeably, although as stated in a previous footnote they are slightly different objects.

below the amount it produces every year.[40]

8.2 The saving-investment identity: Deficits and investment

Recall the savings-investment identity,

$$\underbrace{\underbrace{S}_{Private-saving} + \underbrace{(T-G)}_{Public-saving}}_{Domestic-saving} + \underbrace{(M-X)}_{Foreign-saving} = \underbrace{I}_{Investment}$$

During the Trump administration there have been growing concerns with the impact of **Chinese trade policy and currency manipulation** on the US balance of trade, the decline of US manufacturing capacity, and the rise in outsourcing. The Administration has introduced tariffs and quotas on Chinese products and services in an attempt to eliminate the trade deficit. There were similar concerns in the early 2000s. At that time, most commentators blamed the aggressive export-oriented policy followed by China over the previous (two to three) decades as the main reason behind the American trade deficits.

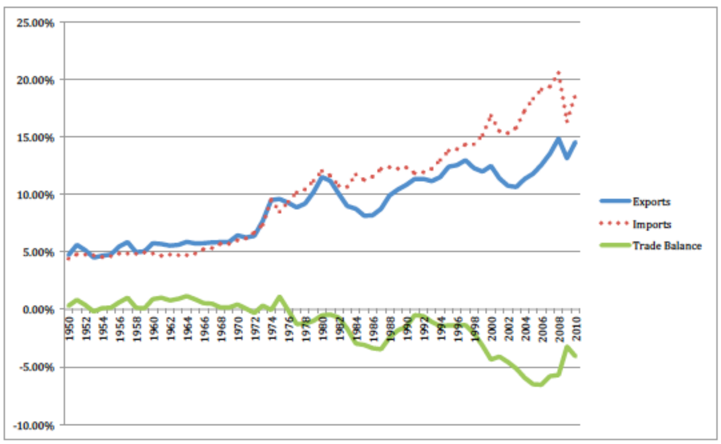

Figure 43: US Exports, Imports and Trade Balance.

As the following figure illustrates, beginning in the **80's the US** began to run **important trade deficits**. Up to that date **domestic saving** was roughly equal to **domestic investment** and therefore **trade** was basically **balanced**. Since then, the US domestic saving rate has been falling, but at a faster pace than the US investment rate. As a result, the arising gap was filled with foreign saving. For instance, in the Bush junior administration (2001-2009), average domestic saving fell below 2% while the average investment rate remained close to 7%. The difference was financed by continuous trade deficits. On average, every year the US economy was spending 5% more than it was producing or, equivalently, it was borrowing abroad the equivalent of 5% of its national product. In terms of the savings-investment identity

$$\underset{\downarrow\downarrow}{S^{domestic}} + \underset{\uparrow}{S^{foreign}} = \underset{\downarrow}{I}$$

[40]Before the existence of the Euro, decreases in spending relative to production in Germany lead to appreciations of the Deutsche Mark, which in turn led to reductions in German trade surpluses. Unfortunately, under a common currency this automatic adjustment mechanism is no longer available.

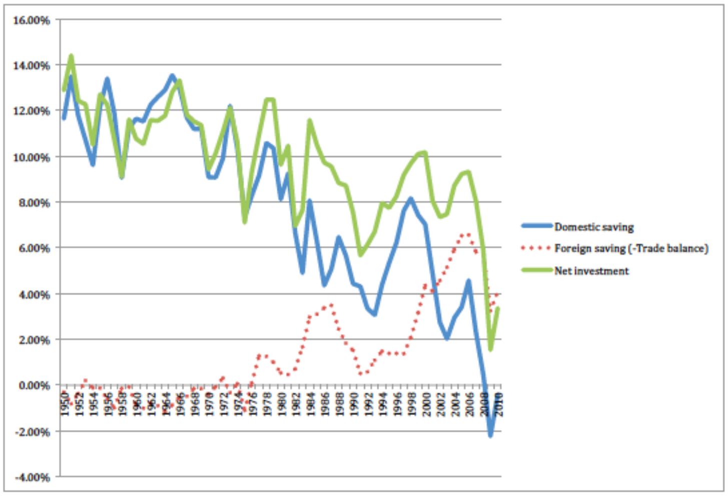

Figure 44: US Domestic Saving, Foreign Saving, and Net Investment.

The savings-investment identity already suggests that with a declining domestic savings rate the only way to attenuate the decline of investment is through continuous increases in trade deficits. At this stage, it seems sensible to explore what lies behind the declining domestic savings rate in the US. The following figure reproduces the **US budget deficits**, i.e. the excess between government expenditures and tax revenues, and compares it to the **trade deficit**. It is interesting to notice that periods where the government deficit was particularly large coincide with periods characterized by large trade deficits.[41]

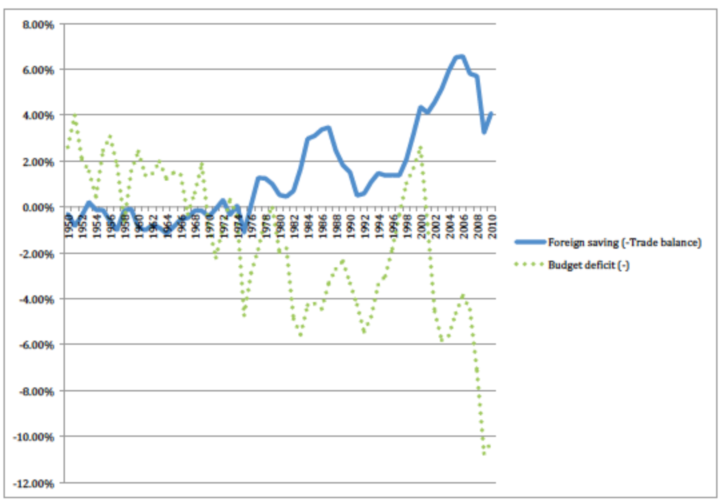

Figure 45: US Foreign Saving and Budget Deficit.

So it seems that to some degree government deficits and trade deficits are connected, it is what some economists have dubbed as the **"twin deficits"**. The savings-investment identity

$$\underset{\approx}{S^{private}} + \underset{\downarrow\downarrow}{S^{public}} + \underset{\uparrow}{S^{foreign}} = \underset{\downarrow}{I},$$

suggests that the continuous government deficits may lie behind the continuous trade deficits experienced by the US since the 1990s.

[41]The correlation coefficient between those two series is -0.55.

A potential narrative for these twin deficits goes as follows. For given levels of private domestic saving and investment, if the government increases the budget deficit borrowing the difference between revenues and expenditures. This will tend to increase the domestic interest rate, since it creates an excess demand in the market for (loanable) funds. The higher domestic rate creates a differential relative to the world interest rate. Capital flows into the US in search of higher returns and the US dollar appreciates. As a result of the domestic appreciation US exporters find more difficult to sell abroad while US imports become relatively cheap. As a result, the trade balance deteriorates.

Now, we are in a position to evaluate the effects the elimination of US trade deficits advocated by the Trump administration. The elimination of trade deficits requires one of the following: a large in private saving, that is a decrease in consumption, a large increase in public saving, that is an increase in taxes and/or a decrease in government spending, or a large decrease in investment, that is a decrease in the future productive capacity of the US economy. As you can see, none of the options seems economically attractive or politically feasible.

To sum up, a potential interpretation of the data is that **large budget deficits** in the US lead to **important decreases in investment** that were only partially **compensated by** foreign borrowing, i.e. running **large trade deficits with China**. In this view, the trade deficit is more a reflection of US internal imbalances than the cause of these imbalances.