Module 4: Measuring, comparing, and modelling

Learning Goal for Module 4
This Module introduces methods for measuring change in time series data, the concept of an index number, and the simple demand and supply.

By the end of this Module, you will:

- Apply transformations on time series data, such as lags, differences, percentage changes, and moving averages
- Apply and interpret key ratios, such as per capita change
- Understand index numbers as measures of change
- Extend the demand and supply model to understand own and cross-price elasticity
- Understand how to use Excel to model basic demand and supply problems

1. Introduction – references and comparisons

Time series data defines a fixed unit of analysis, with multiple attributes, over more than one point in time. The time units may be short—nanoseconds or one billionth of a second—or long—decades or even centuries. In economics, the common time units are days (stock markets), weeks/months (retail sales, unemployment, prices, etc.), and quarters/years (GDP).

When measuring change (over time or among units at a point in time), we need a reference point. The reference point or base is a fundamental concept in social science. With time series data, this is always a point in time, termed the "base year/month/week/day." Often the reference is the same day last year, which is common with price data and even when comparing the temperature this day to the same day last year.

When comparing other states, the context will define the reference:

- Joan is taller than Jane.
- Jane is shorter than Joan.

In the first statement, Jane is the reference point, and in the second, it is Joan. These are logically the same, but as soon as we move to a quantitative measure, such as height, the reference becomes important. This example shows why the reference point matters.

- If Joan is 67 cm high and John is 65 cm, then Joan is 3.08% taller than John (67/65 – 1) = 3.076%, and
- John is 2.985% shorter than Jane (65/67 – 1) = -2.985%

Other examples where tracking the reference point becomes important.

- If gas costs $1.15/l at one station (A) and $1.09/l at another station (B) then
  A is 5.5% more expensive than B
  B is 5.2% less expensive than A
- Changing the base produces a different answer.

- Total revenue = price x quantity
- If the price of eggs is $2.00 per dozen and a restaurant buys 20 dozen, total revenue is $40.
- If the price rises by 10% ($2.20/dozen) and the restaurant buys 10% more (to 22 dozen), total revenue rises to $48.40.
- Prices and quantities have each both risen by 10%, but total revenue has increased by 21% (make sure you understand why and can calculate this).

- An increase of 25% in the Chinese Yuan (CNY or ¥) relative to the Canadian dollar is the same as a 20% decrease in the value of the Canadian dollar to the CNY (using May 27, 2016, and an exchange of 5 CNY to 1 CAD).
- If ¥100 buys $20, and the value of the ¥ rises by 25%, this means that ¥100 now buys $25. However, the Canadian dollar is now 25% less valuable because one needs more dollars to buy Chinese currency.

- Assume the price of gas at station B increases from $1.00/l to $1.05/l and then to $1.1025/l (which is two successive increases of 5%).
- The total increase of the two price increases is 10.25% (not 10%). (Recall the discussion on the geometric mean.)

2. Transformations

One common early step in data analysis is to explore data transformations. Logs, lags, differences, and percent change are all common ways economists transform data.

2.1. Logs and per capita

Two common data transformations are taking logarithms and calculating per capita values. These apply to time series, cross-section, and pooled data. See

Logs.xlsx

The data on US land area and population offer some interesting lessons. The cross-sectional plot (make sure you understand why this is cross-sectional data) that shows the land area and population for each state appears in Figure 1.

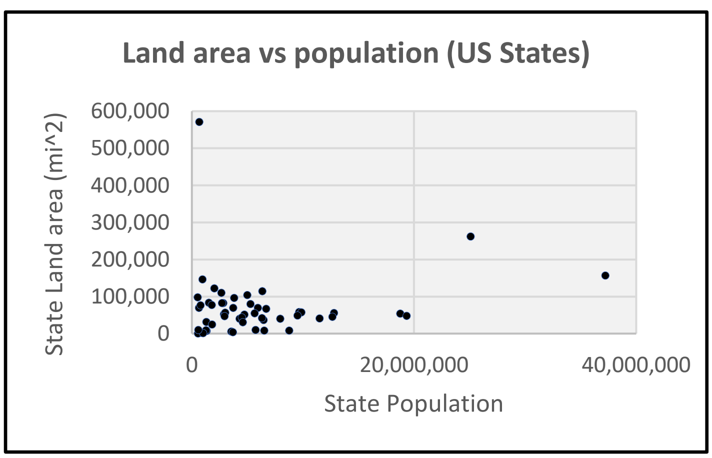


Figure 1: Untransformed data

Compare the raw dataset, downloaded from the web, **[Pop_Land_Area_States_Raw.xlsx]** with **[Pop_Land_Area_US_Final.xlsx]** to understand the data development typical of economic analysis**.**

Now take logarithms, which change the scale, drawing in the outliers (states with large land areas and large populations). Notice that logs change the scale of the vertical and horizontal axes. Applying a log transformation on one or both scales usually "spreads out" the data. In Modules 5 and 6, logarithmic transformations appear in several well-known models from microeconomics.

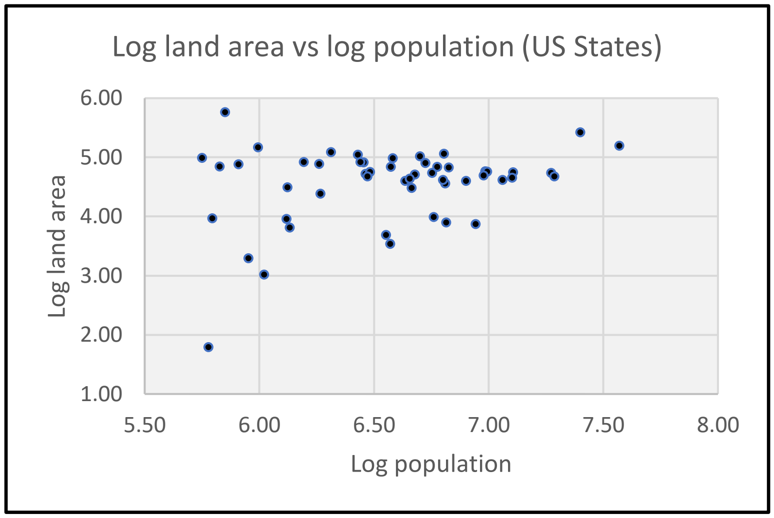


Figure 2: Double log transformation

Another transformation is to plot density by dividing the area by population or number of dwellings. This dataset includes the population per square mile and the number of housing units per square mile.

Transformations Per Capita.xlsx

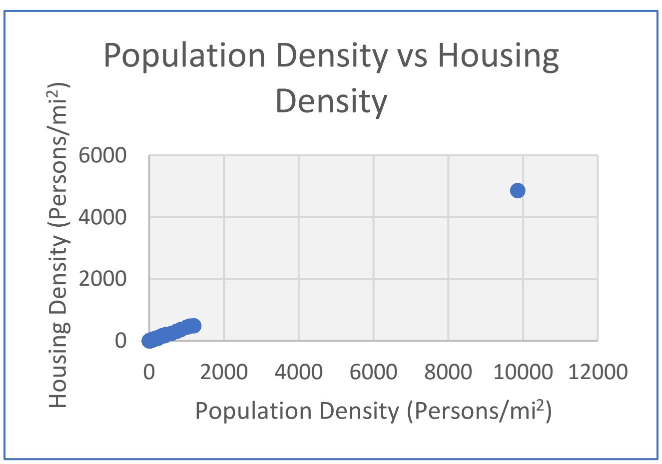


Figure 3: Population Density vs Housing Density

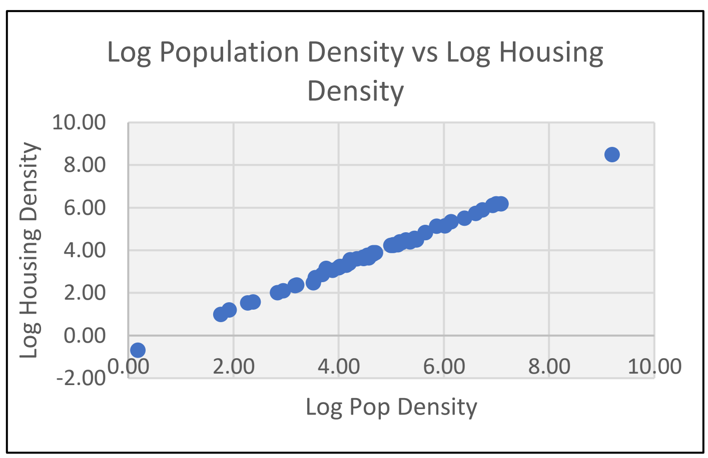


Figure 4: Log Population Density vs Log Housing Density

Applying logs normalizes the plot to make the linear relationship clearer. Note how creating a ratio, persons per square mile, measures housing density. Often an abstract concept, such as crowding in living arrangements, requires a concrete measure, and often more than one choice exists. Persons per dwelling unit is another measure of crowding. Why might this be a better or worse measure than persons per square metre? Developing reliable and valid measures of abstract concepts is an important task in economic analytics.

GDP (gross domestic product)[1] is probably the most common measure of national economic well-being. Many complain that it excludes non-paid and volunteer work, the underground economy (cash transactions), the social costs of environmental degradation, and impacts of unequal income distribution. Official statistical agencies, such as Statistics Canada, are aware of these issues and working on alternative measures. In the meantime, GDP is still the common measure of national economic welfare. See [**REALGDP per CAPITA.xlsx**].

Figure 5 shows the impact of the Great Recession (2009–10) and the COVID pandemic on the Canadian economy. (Note that the vertical axis does not start at 0.) This is technically not a correct method for showing time series, but it does increase the dramatic effects of the recession and COVID.

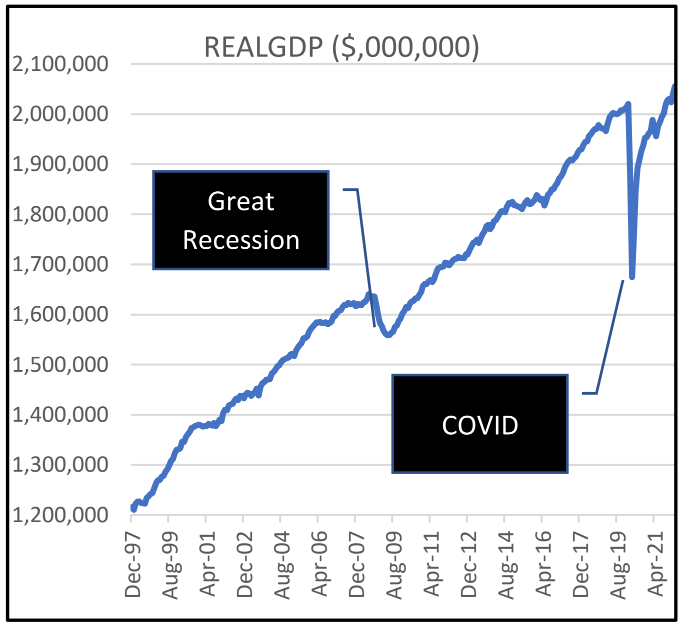


Figure 5: Real GDP - The Great Recession and COVID

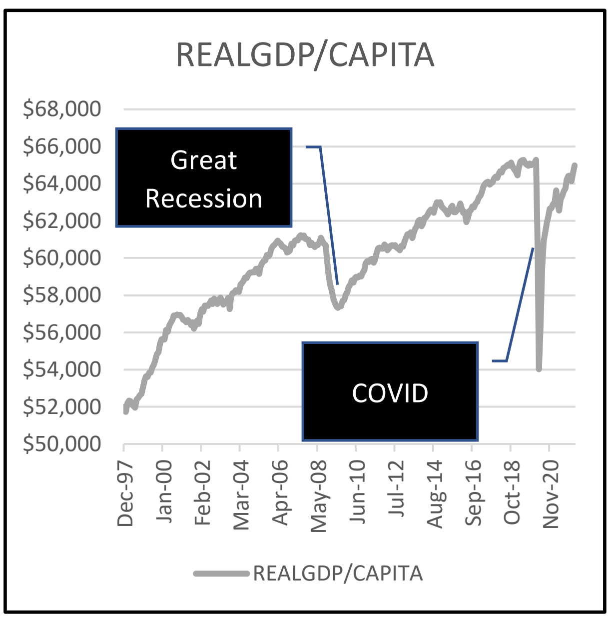


Figure 6: Real GDP per capita

Video: Altering axes for dramatic effect

The economic analyst will face pressures to show data in certain ways to send a political or commercial message. Altering axes is a common technique.

The vertical axis shows GDP in trillions of dollars (annualized monthly values).

Using population (those over 15 years of age) places the impact in a different context that further clarifies the impact that these two events had on the

[1] It is worth understanding the difference between gross domestic product and gross national product. Review a first-year text to refresh your memory.

Canadian economy. Figure 6 shows the dramatic impact of these two events on the economic welfare of individuals over 15. In Module 5, this example will show how trends can tell a more interesting story about an economy. Dividing revenues, costs, or output by a “population” number often supports a deeper interpretation of the data. For example, output per worker is the most common measure of productivity.

COVID offers the opportunity to understand both the effects of log transformations and per capita measures. The most common adjustment when studying the growth of an economy or epidemiological events, such as COVID, uses logarithms to express total cases, as shown in Figure 8. Infectious diseases often show exponential growth. Taking logarithms of the same data still shows changes, but it is often more useful in showing whether public policy, such as social distancing, is working. Study the dataset and embedded notes.

COVID.xlsx

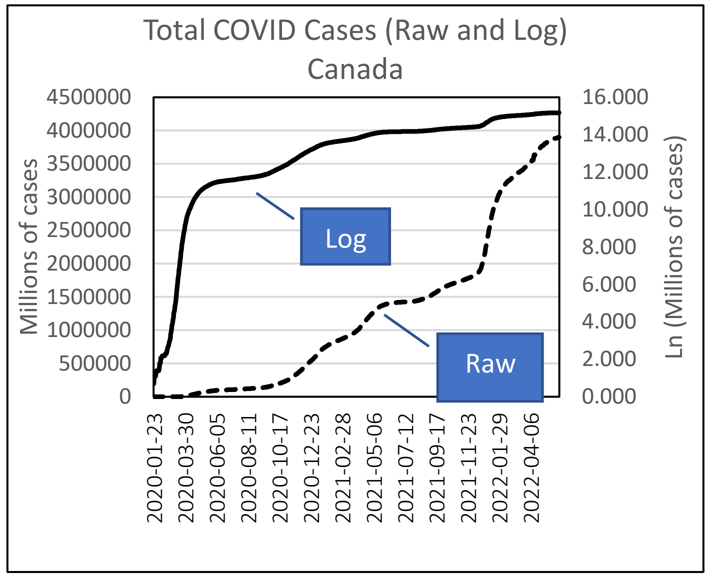


Figure 7: Total COVID cases

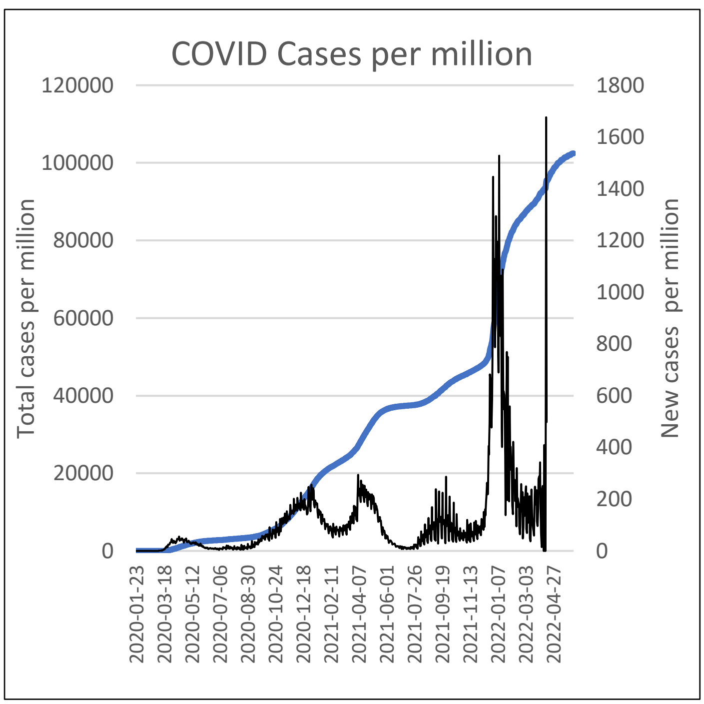


Figure 8: COVID cases

Another common transformation when studying a disease is to measure the number of cases per capita or per 100,000. This is the incidence of the disease. The prevalence is simply the number of people currently infected. Figure 8 shows total and per capita cases (per million.)

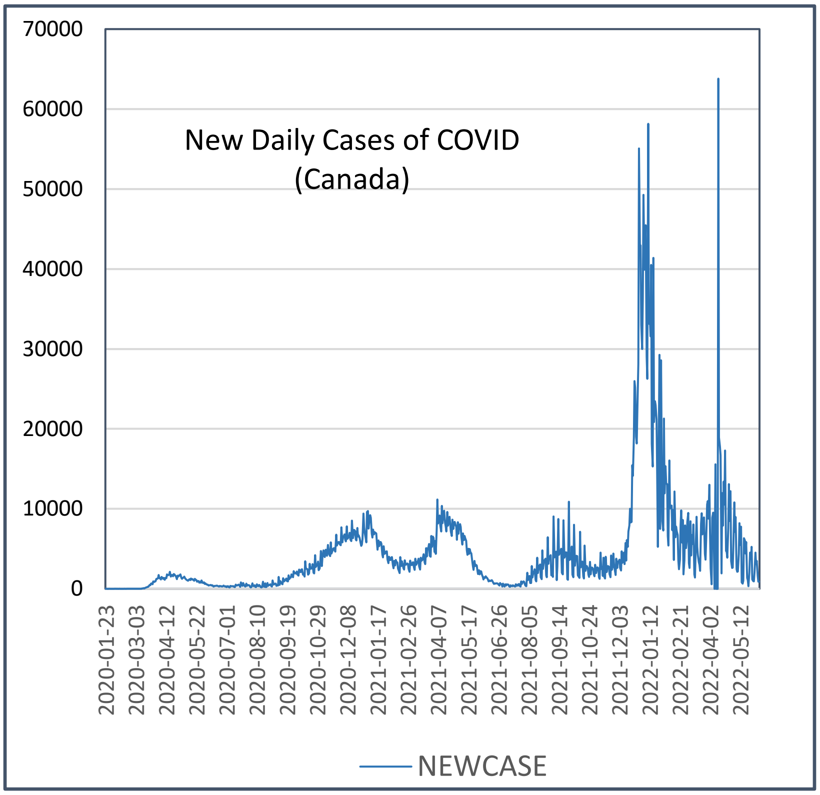


Figure 9: Daily Cases of COVID

Sometimes data show high variability, either because the processes are chaotic or, equally possible, data collection just may be inherently variable (Figure 9). This chart suggests errors in the raw data. Many public health agencies struggled to keep correct case counts during the pandemic and, despite coming from a very reputable data aggregator, errors exist in these data. The data analyst always has prime responsibility for examining the data for errors, flagging the issues for the reader, and taking responsible corrective actions.

> Data aggregator: A repository of data and statistics collected by other agencies. An example being www.statistica.com.

Using a moving average, in this case, a seven-day moving average, smooths out this variability and allows one to see trends a little more clearly, as shown in Figure 10. More on this later in the chapter.

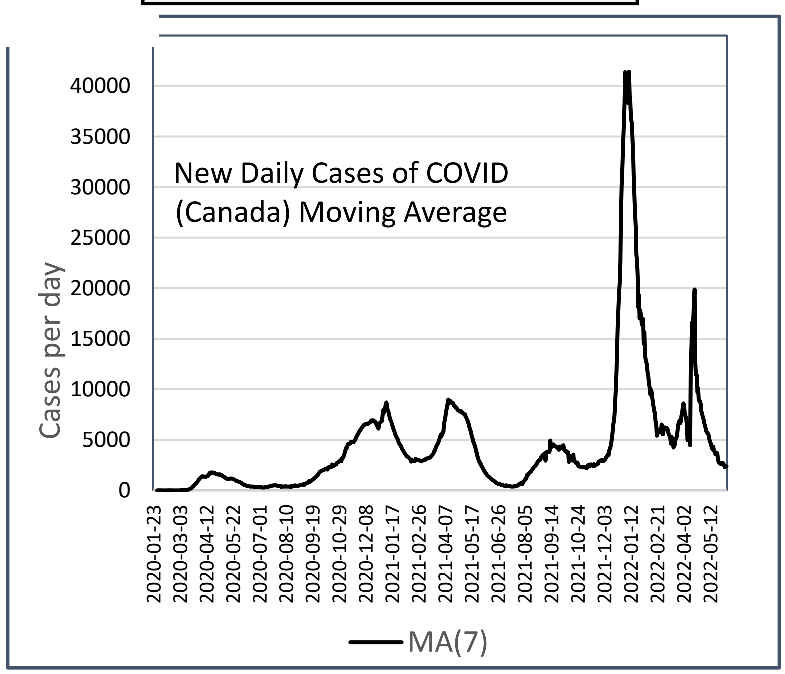


Figure 10: Seven-day moving average of COVID cases

2.2. Lags, differences, and percent change: Time series transformations

Logs and per capita are transformations that apply to both time series and cross-sectional data; this section considers transformations on time series data. The analysis of time series data forms a central task in forecasting. Searching for recurring patterns in a time series supports the projection of past values to the future. Figure 11 shows Canada's GDP as of December 2021. See REALGDP.xlsx

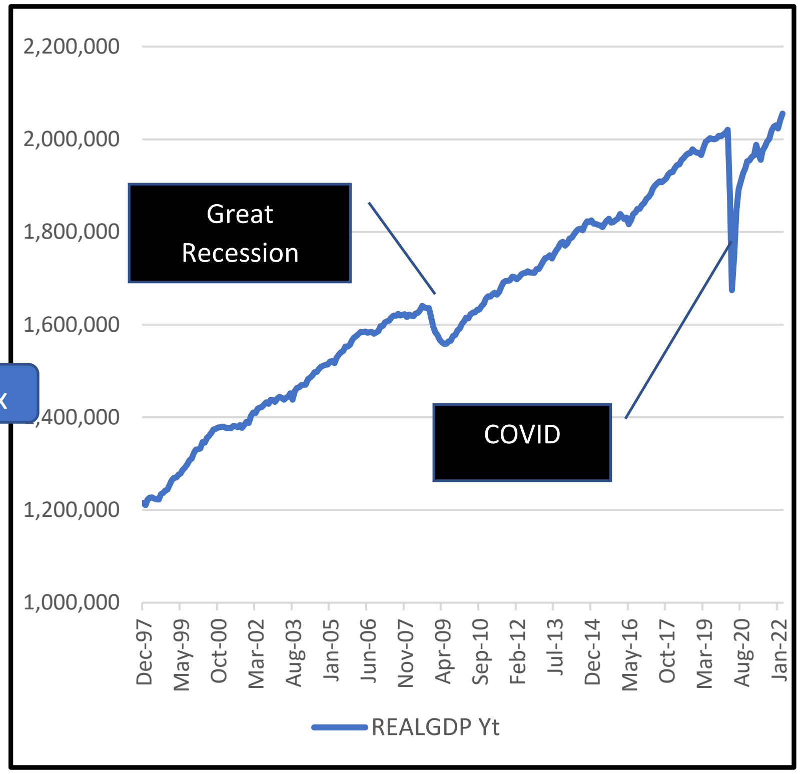


Figure 11: Real GDP

A **lag** stands for the data from the last period. Referring to the spreadsheet, just copy B3:B271 to C4:B271. Note that creating a lagged variable cuts the first observation. The *first difference* is simply the past value subtracted from the present value. The *percentage change* is the first difference divided by the lag.

Series Y_t	
Base	Y_t
Lag	Y_{t-1}
First difference	Y_t - Y_{t-1}
Second difference	$(Y_{t-2}-Y_{t-1})-(Y_t - Y_{t-1})$
Percentage change	$(Y_t-Y_{t-1})/Y_{t-1}$

Inserting a lagged variable into an economic model is one way to express inertia and capture some behavioural effects. First and second differences are common features of **time series analysis** (which is not the same as analysis of time series data). Plotting percentage change can reveal shifts in the data that plotting the raw data may conceal. The goal of "differencing" and percent change is to use a "microscope" on time series data to show shifts and turning points. One approach to inferring causal relations in economics is to align differences and percent changes statistically.

Here is a detailed table showing the various calculations. Make sure you understand the calculation in the spreadsheet [**REALGDP.xlsx**]. Be sure you understand the second differences. Notice that second differences can magnify the first difference. If the Y_t follows some rule that creates a smooth series (such as a parabola or some quadratic), second differences will have a dampening effect.

Raw	First Diff.	Second Diff.
Y_t	$Y_t - Y_{t-1}$	$(Y_{t-2}-Y_{t-1})-(Y_{t-1}-Y_t)$
49		
64	15	
98	34	19
57	-41	-75
37	-20	21
76	39	59
31	-45	-84
33	2	47
10	-23	-25
2	-8	15
64	62	70

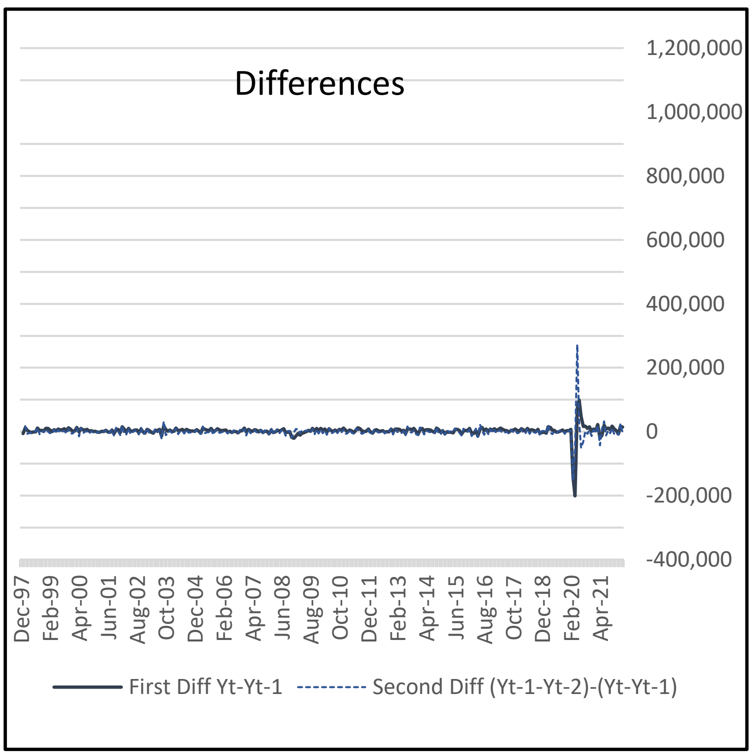


Figure 12: First and second differences

Figure 12 shows the first and second differences. COVID created a major shock that dominates the series. Using the example spreadsheet

REALGDP.xlsx

experiment with the series, stopping it prior to COVID. You will see that second differences tend to have lower variation than first differences, except during unusual events, such as the Great Recession and COVID.

Figure 13 shows percent change, which closely resembles the first difference. The formula shows why. The Great Recession and COVID appear and align with the raw data on real GDP.

Using the spreadsheet, select options to understand the relationship among these transformations. Here are some important points:

COVID.xlsx

- Make the lines thin and avoid colours. Use different dashes instead denote different series.
- Adjust the fonts on axes, legends, and title to fit the text.
- Do not link the file unless you are preparing a report in Word that you will regularly update with new data in Excel.

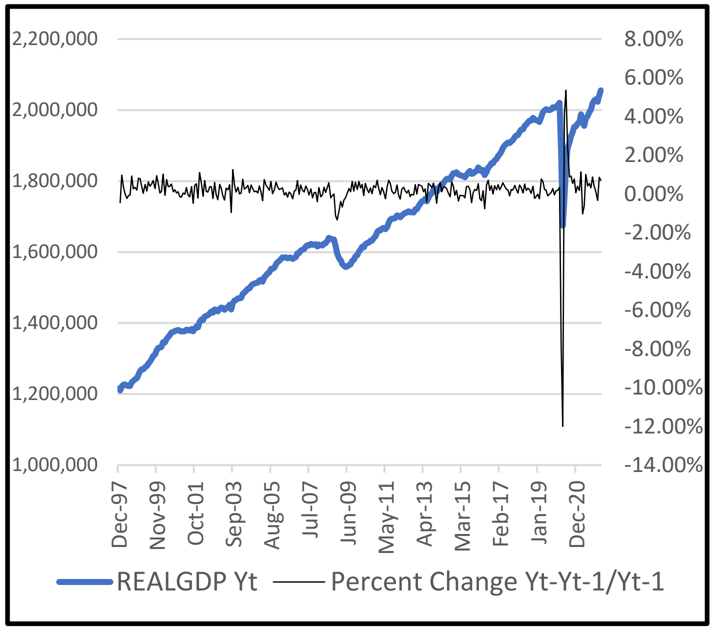


Figure 13: Percent change

2.3. Smoothing and moving averages

Section 2.2 introduced moving averages to smooth data. As with many procedures in Excel, several ways exist to complete a task. One can express moving averages using the =AVERAGE() formula. This section uses the Data Analysis ToolPak to compute moving averages.

Some series are inherently variable, sometimes so much so that underlying trends can be hard to see ("one cannot see the forest for the trees"). For example, retail trade and prices of commodities often show variability around a trend or cycle.

Do not use a moving average or another smoothing technique on differences or percent change. This defeats the purpose of these transformations.

- Motor vehicle sales (cars and trucks) are typically highly variable following a seasonal variation.
- This figure shows sales typically peaking in May/June, with a slowdown in the winter.
- Overall, an upward trend seems clear, with a reversal in 2017 and a steep drop in April 2020 (COVID).
- These data want to tell a story; the job of the economic analysts is to tell that story.
- As an exercise, download these data from the Statistics Canada website and explore care sales during COVID.

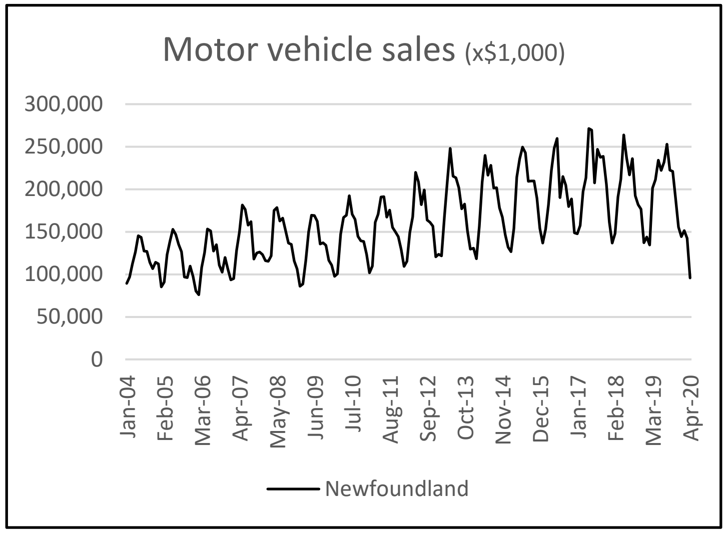


Figure 14: Motor vehicle sales

The standard view of time series data is that it has four components:

- **Trend** (sales move in pace with core demographic and societal trends)
- **Cyclical** (sales change with the fortunes of the economy)
- **Seasonal** (calendar variations with the "season," often referring to months, but could be days, weeks, or quarters)
- **Irregular changes** (measurement error and other factors)

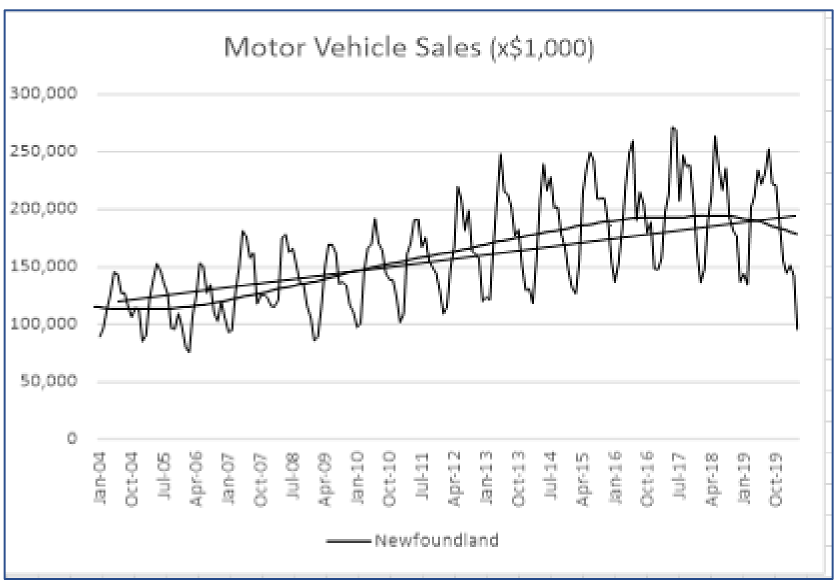


Figure 15:Motor vehicle sales - trends

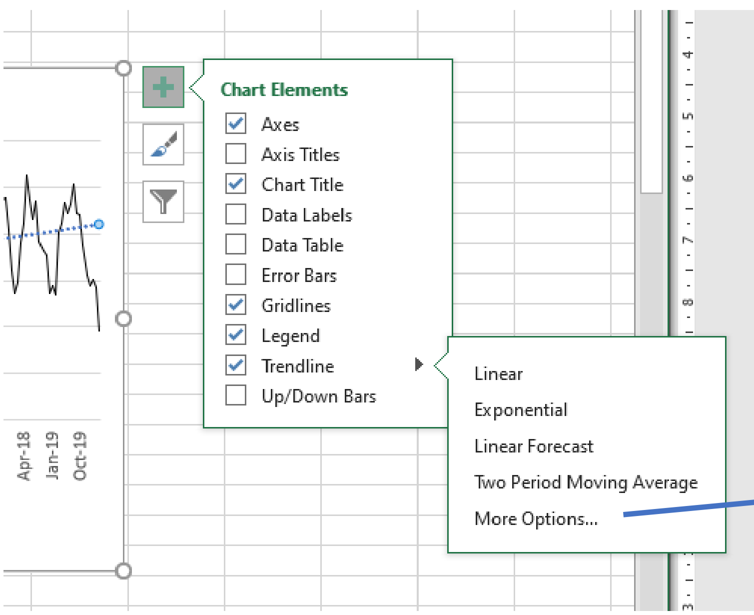
- Seasonal variations usually refer to monthly changes, although this depends on the industry.
- The usual assumption in economics is that trends are linear, but…
- Historically, economists viewed the business cycle—the sequence of expansion and contraction—as being between four and five years, but in the last couple of decades, it has become hard to separate the trend from the cycle. Increasingly, statisticians merge the trend and cycle, and identify the trend-cycle.
- In this figure, the linear trend is not a good summary of what has been occurring with sales.

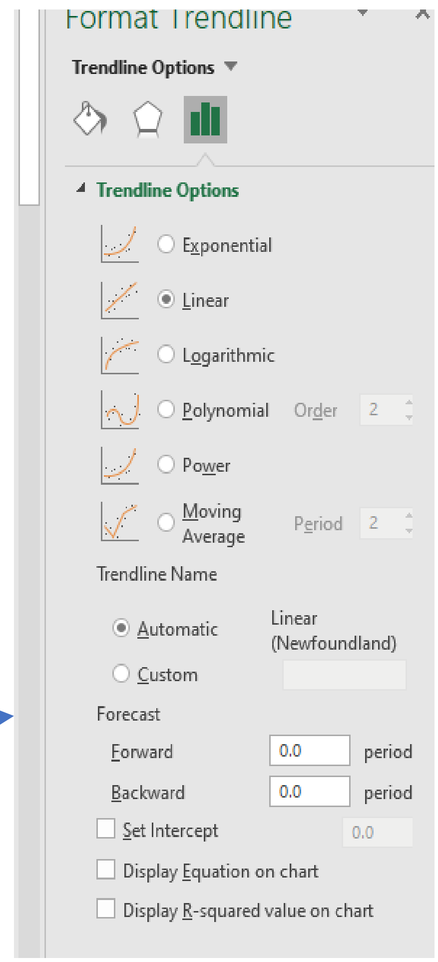
Video: [COVID - Moving Average]

Economic analytics cannot rely solely on statistical manipulation. It is essential to understand and include the social, demographic, political, and even cultural trends in the analysis. The impact of the Great Recession on car sales might be understandable, and a slight dip is clear, but the increase between 2010 and 2017 might be puzzling, unless one were aware of the offshore oil (Hibernia) boom, followed by a slowdown starting in 2018. Economic analysis must tell a credible story about the data, aided by statistical and, increasingly, machine learning and AI techniques. See [**Motor Vehicle Sales.xlsx**].

Common methods for statistical analysis of time series data include regression, covered in Modules 5 and 6, as well as special techniques that fall under the general term "time series analysis" but which are beyond the scope of this course.

You will notice under the graphing functions in Excel some options to insert trend lines. These are fun to play with and can be useful for an informal presentation, but they do not replace a thorough analysis as discussed later in the course.





.

Video: [Moving Average - Data Analysis ToolPak]

When downloading information from many official statistics websites and aggregators, such as OECD, you may see several options for seasonally adjusted data. Study the metadata on that website to select the proper adjustment.

> "A seasonally adjusted time series is a monthly or quarterly time series that has been modified to eliminate the effect of seasonal and calendar influences. The seasonally adjusted data allow for more meaningful comparisons of economic conditions from period to period. A raw time series is the equivalent series before seasonal adjustment and is sometimes referred to as the original or unadjusted time series."
>
> *Statistics Canada*
>
> *https://www150.statcan.gc.ca/n1/dai-quo/btd-add/btd-add-eng.htm#two*

Those wishing a deeper understanding of seasonal adjustment can visit the URL associated with this quote.

Moving averages are a common method for reducing the irregularities of time series data. This can be useful for aiding the "eye" to sell the underlying course of time series data. Consider the COVID data for new daily cases [**COVID.xlsx**]. Moving averages simply average the data for K periods, where K must be 2 or higher. Use the Data Analysis ToolPak to smooth this series in this dataset for cases in 2022.

Experiment with 3-, 5-, and 7-day moving averages. Note that the moving average process will always drop 2, 4, and 6 data points at the start of the series. You need to decide between more aggressive smoothing (higher order of moving average) and preserving sufficient variation to communicate the nature of the data.

3. Index numbers

Economic analytics requires methods to summarize disparate data. The grade point average (GPA) summarizes an academic career. Two important macroeconomic measures are the unemployment rate and the consumer price index. Many refer to the latter as a measure of inflation, although most statistical agencies discourage that interpretation.

3.1. The origin of index numbers

In 1707, a student was denied a full fellowship to attend Oxford because he had an income more than £5 (a threshold set in 1440). The student asked William Fleetwood to help with an appeal. Fleetwood compared 1707 prices to 1440 prices, showing that £5 had lost value in "real" terms. Fleetwood created one of the first price indexes showing the difference between a nominal and a real value; the university awarded a full fellowship to the student.

"The whole is the sum of its parts." That is not the usual way this expression appears, but for index numbers, it is a correct enough statement. A more exact statement might be, "The whole is the weighted sum of its parts," which is exactly what an index number measures.

Definition

An index number is a statistical measure to show changes in a variable relative to a point in time, geography, or some other reference, termed the "base."

Applications

- Measures of living cost (consumer price index)
- Wage indexes (which measure wage changes over time)
- Production indexes
- Unemployment indexes

- Figure 16: CPI with and without energy shows the consumer price index for Canada, all items, and all items less energy.
- Note the inclusion of the base (2002=100).
- See Price Index.xlsx].

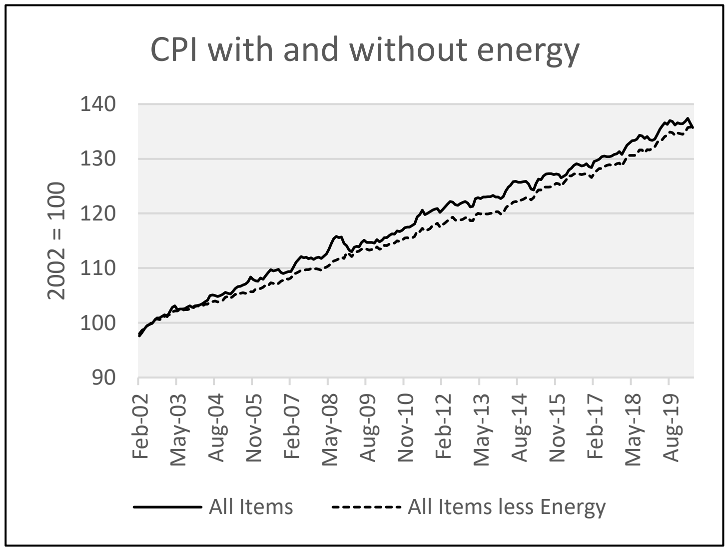


Figure 16: CPI with and without energy

Before delving into the construction of an index, learning how to manipulate indexes offers important skills.

3.2. Real versus nominal

The money illusion experienced by that Oxford student over three hundred years ago persists. In the summer of 2022, when this chapter was written, inflation was running at 6–8% percent; the prices gas, food, and housing have all accelerated, leaving many feeling worse off. At the same time, the labour markets are extremely tight, and many workers switch jobs rapidly, seeking higher wages.

Economists prefer to develop and test theories using real quantities, dropping the "veil of money illusion." The use of real GDP in the examples above shows the true state of the economy. Figure 17 shows money and real wages since 2003. Note that the terms "nominal wages" and "money wages" mean the same thing, as do "real wages" and "inflation adjusted wages."

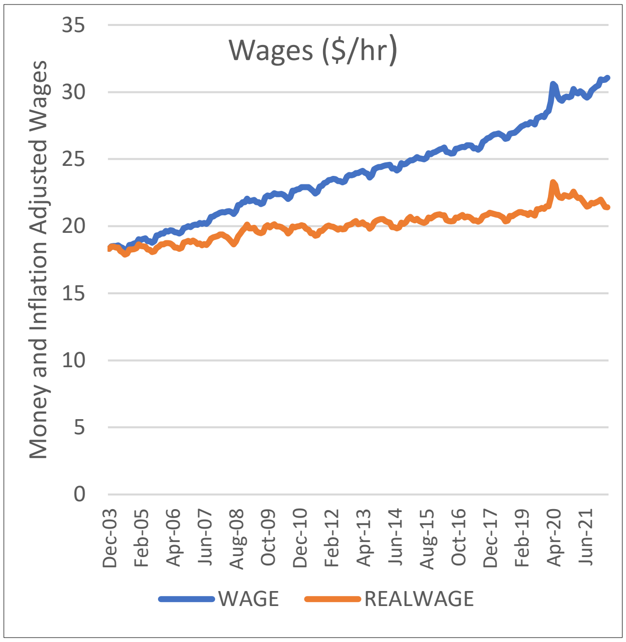


Figure 17: Nominal wages - deflated

This chart shows that real wages grew over this period, peaking in May 2020 and declining through the pandemic. Real wages in April 2022 are about the same as just before the pandemic.

Analysis of wages offers fodder for comment across the political spectrum. Those on the left will point to the decline through the pandemic; those on the right will note that the increase in real wages over the entire period shows that workers are becoming better off.

A challenge for economic analysts is to tell a balanced story. These data pertain to the total economy. Statistics Canada data support detailed analysis by sex, occupation, age, and region. See REALWAGE.xlsx for the data and an illustration of how changing the vertical scale can "bend" the interpretation. Note the structure of this dataset with the Metadata tab.

3.3. Changing base period/reference

In the file REALWAGE.xlsx, the "base" for the original index in 2002 = 100 (Jan 2002). Although not strictly needed, the index used to adjust the CPI was "rebased" to Dec 2003 = 100. Follow the calculations in the example to understand how to do this.

Index numbers most commonly measure changes over time and therefore are time series data, but they can also apply to cross-sectional data. For example, one can imagine a cross-sectional dataset of the value of the average home per capita in Canadian cities. One can create a crude index by selecting one city and dividing its value into all the others. The base city has its value as 1 (since you divided its value by itself), and it is set to 100 as a convention. See

House Price Index.xlsx

Here is a series. The original series has a base of 2002 = 100.

The other two series have had the base changed to 2005 and 2014.

Changing the base is easy

- Using CPI (2002 = 100), change the base to 2005 (Column 3) simply by dividing all entries by the 2005 value when the base year is 2002 (that is 107.00). So, the value of 81.9 in the CPI (2005=100) is 87.6/107.
- To rebase to 2014, divide 125.8 into every entry in the first CPI column to get Column 4. The value of the CPI in 2008, with a new base year of 2014, is 114.1/125.8 = 90.7.

Year	CPI (2002=100)	CPI (2005=100)	CPI (2014=100)
1995	87.6	81.9	69.6
1996	88.9	83.1	70.7
1997	90.4	84.5	71.8
1998	91.3	85.3	72.6
1999	92.9	86.8	73.8
2000	95.4	89.2	75.8
2001	97.8	91.4	77.7
2002	100.0	93.5	79.5
2003	102.8	96.1	81.7
2004	104.7	97.9	83.2
2005	107.0	100.0	85.0
2006	109.1	102.0	86.7
2007	111.5	104.2	88.6
2008	114.1	106.6	90.7
2009	114.4	106.9	90.9
2010	116.5	108.9	92.6
2011	119.9	112.1	95.3
2012	121.7	113.7	96.7
2013	122.8	114.8	97.6
2014	125.8	117.6	100.0

3.4. Splicing two indexes

Series 1 has a base year = 2002, and Series 2 starts in 2003, with a base year between 2003 and 2004.

We can splice the two, creating a new series with a base year of 2003.

The key is to transform the series to the same base year and then just combine the two new series.

- Series 1 is rebased to 2003 = 100 (shown in Column 2). *Divide 102.8 into 87.6 to get 85.2, etc. to get Column 3.*
- Series 2 is rebased to 2003 = 100 (shown in Column 4). *Divide 96.1 into the rest of the series to get Column 4.*
- We now must merge (slide together) series 3 and 4 to become the final series in Column 5.

Year	1	2	3	4	5
1995	87.6		85.2		85.2
1996	88.9		86.5		86.5
1997	90.4		87.9		87.9
1998	91.3		88.8		88.8
1999	92.9		90.4		90.4
2000	95.4		92.8		92.8
2001	97.8		95.1		95.1
2002	100.0		97.3		97.3
2003	102.8	96.1	100.0	100.0	100.0
2004		104.7		109.0	109.0
2005		107.0		111.4	111.4
2006		109.1		113.6	113.6
2007		111.5		116.1	116.1
2008		114.1		118.8	118.8
2009		114.4		119.1	119.1
2010		116.5		121.3	121.3
2011		119.9		124.8	124.8
2012		121.7		126.7	126.7
2013		122.8		127.8	127.8
2014		125.8		131.0	131.0

3.5. Dealing with missing data

Missing data creates problems for economic and business analysis. Missing data is an important topic, which we will take up in more detail later in the course. Sometimes we have just one entry with missing values, and it is often replaced with the average. It is common to encounter information with missing data. If just a few points are missing, then simple steps can offer a good approximation for the missing points.

Sheet A has one missing point, while Sheet B has three missing points.

Many strategies exist for "fixing" the issue; it can be difficult to decide on the best approach. Also, most methods depend on the judgment of the researcher.

1. ***One missing data point***: By far, the most common strategy is to use the simple arithmetic average. In Sheet A, the value for 2005 is 106.9.
2. ***Two missing points:*** Divide the interval into thirds, and to the first missing point, insert the value of the last observation = 33%, and to the second missing add 66%.
3. ***Three missing points:*** With three missing points, as in Panel B, you divide the missing span into quarters.

	A
Year	CPI (2002=100)
1995	87.6
1996	88.9
1997	90.4
1998	91.3
1999	92.9
2000	95.4
2001	97.8
2002	100.0
2003	102.8
2004	104.7
2005	
2006	109.1
2007	111.5
2008	114.1
2009	114.4
2010	116.5
2011	119.9
2012	121.7
2013	122.8
2014	125.8

	B
Year	CPI (2002=100)
1995	87.6
1996	88.9
1997	90.4
1998	91.3
1999	92.9
2000	95.4
2001	97.8
2002	100.0
2003	
2004	
2005	
2006	109.1
2007	111.5
2008	114.1
2009	114.4
2010	116.5
2011	119.9
2012	121.7
2013	122.8
2014	125.8

Using the data from B, it is possible to show this in action for three missing points for 2003, 2004 and 2005.

- The difference between 109.1 and 100 is 9.1. We divide this into quarters (25% of 9.1 = 2.3, rounded).
- We add this to 100, and the first missing number is 102.3.
- We add 2.2375 to this number (or 4.5 to the value 100).
- Finally, repeat, to find the third number.
- We insert the missing values.

2001	97.8
2002	100.0
2003	
2004	
2005	
2006	109.1
2007	111.5
2008	114.1

2001	97.8
2002	100.0
2003	102.3
2004	104.5
2005	106.5
2006	109.1
2007	111.5

3.6. Price/quantity/value relatives: The first step in creating an index

A ***price relative*** is the comparison of a price in one period with the price in some base or reference period. The price relative in "a" referenced to "b" is

$$\text{Price Relative} = Pa / Pb = Pa \mid b$$

The vertical line | means "referenced to."

Note that *a* and *b* can be dates or places applying to time series or cross-sectional data. (Recall [House Value Index.xlsx]).

Verify these properties using Pa=$1.35, Pb=1.15, and Pc=1.25, or whatever values you wish.

Properties of price relatives	
Property	**Definition**
a. Identity	$Pa / Pa = 1 = Pa \mid a$
b. Time reversal	$Pa \mid b \bullet Pb \mid a = 1$
c. Cyclical	$Pa \mid b \bullet Pb \mid c \bullet Pc \mid a = 1$
d. Modified cyclical	$Pa \mid b \bullet Pb \mid c = Pc \mid a$

Example: If P_{2016} is $1.35 and P_{2015} is $1.15, the price relative P_{2016}/P_{2015} is $1.35/$1.15 = 1.17.

A ***volume/quantity*** relative compares quantity change over time or space. This allows us to compare prices in two areas or times.

$$\text{Quantity Relative} = Qa / Qb = Qa \mid b$$

This index has the same four properties as a price relative.

A ***value relative*** applies to price and quantity using total revenue.

$$Value\,Relative = (Pa * Qa) / (Pb * Qb) = P * Qa|b$$

This allows us to track total value in two areas or two times.

A ***link (chain) relative*** is a number that reflects the chain between two periods or states If P_1, P_2, ... is a set of prices, then the link relative is:

$$P_1 / P_2 * P_2 \,|\, P_3 ....$$
$$P_{1\,|\,2}, P_{2\,|\,3}, P_{3\,|\,4} ...$$

Example: If the prices for 2010 … 2012 are 8, 12, 15, 18, the link relatives are:

$$P_{2010\,|\,2011} = 12/8 = 150\% = 1.5$$
$$P_{2011\,|\,2012} = 15/12 = 125\% = 1.25$$
$$...$$

This definition applies equally to quantity and value indexes.

3.7. Constructing an index number

An index number allows us to compare trends for a large amount of information concisely. Many options exist for creating a price index, each with advantages and disadvantages.

An index number must have the same properties as price/quantity/value relatives (see Section 3.6). Only complex index numbers meet all these tests. All price indexes require us to define a *basket of goods and services in a base year*. But first, we must define the summation operator more completely.

The summation operator $\sum$ summarizes a repetitive addition across specified units. The key is that these units have "tags," usually a numeric or alpha designation for a specific commodity (or service) or a period:

$X_1, X_2, X_3 ... X_n$ or

$X_a, X_b, X_c ...$, or

$X_{1961}, X_{1962}, ... X_{2020}$

X may measure prices, quantities, or more complex measures, such as GDP.

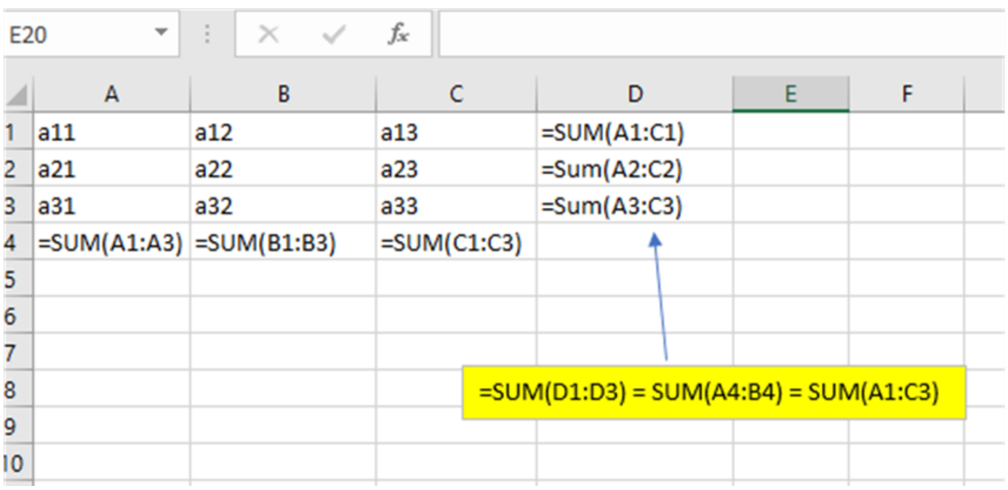
The addition of "m" numbers appears as

$$\sum_{i=1}^{m} Xi = X1 + X2 + \cdots + Xm$$

Where "i" indexes time units (week, month, year), geography (cities, countries), persons, etc.

Here is a 3 x 3 matrix in algebra format. The cells are all labelled using a single variable "a" with subscripts denoting the cell.
Row and column totals use single summation, while the matrix total uses a double summation.
Be careful to track the i and j values.

a11	a12	a13	$\sum_{j=1}^{3} a_{1j}$
a21	a22	a23	$\sum_{j=1}^{3} a_{2j}$
a31	a32	a33	$\sum_{j=1}^{3} a_{3j}$
$\sum_{i=1}^{3} a_{i1}$	$\sum_{i=1}^{3} a_{i2}$	$\sum_{i=1}^{3} a_{i3}$	$\sum_{i=1}^{3} \sum_{j=1}^{3} a_{ij}$



	A	B	C	D	E	F
1	a11	a12	a13	=SUM(A1:C1)		
2	a21	a22	a23	=Sum(A2:C2)		
3	a31	a32	a33	=Sum(A3:C3)		
4	=SUM(A1:A3)	=SUM(B1:B3)	=SUM(C1:C3)			
5						
6						
7						
8						
9						
10						

Annual GNP ($ trillions)				
Year	Country 1	Country 2	Country 3	Total
2000	45	300	135	480
2001	47	301	134	482
2002	48	305	135	488
2003	48	303	136	487
2004	49	307	137	493
2005	50	308	138	496
2006	51	309	140	500
2007	53	310	141	504
2008	55	311	148	514
2009	58	312	145	515
2010	60	315	149	524
2011	61	315	143	519
2012	61	316	149	526
2013	64	319	151	534
2014	64	321	155	540
	814	4652	2136	7602

In the example at the left, each row (year) sums to the right column, each column (country) sums to the bottom row, and the right column and last row sum to the total (7602).

The formula for a simple aggregate price index is, well, simple!

$$\text{Simple Aggregate Price Index} = \sum_{i=1}^{n} P_i \Big/ \sum Po$$

This compares total prices in year "n" and compares them to the base year, usually designated as "O" This index has two flaws. One, it assumes all goods in the basket are bought in the same quantity over time. Two, everything is measured in the same way: cars, cans of tuna, flash memory, and cups of Starbucks can all be measured in the same units.

Example: Computing the simple aggregate price index for university student computing needs (annual). Here is the basket and prices for a typical student.

	Prices ($/unit)			Average quantities bought by students each year		
	2014	2015	2016	2014	2015	2016
6 GB Flash Memory	15.32	9.45	5.50	1	1.5	2
Box of Paper	31.35	32.67	32.50	3	1.5	.5
Cups of Coffee	1.70	2.05	2.25	1500	1550	1617

Over the three years, this student purchases more coffee and flash memory but less paper A simple price relative shows that prices have declined over this period.

$$\sum Pn / \sum P0 = \frac{\text{Sum of prices in current year}}{\text{Sum of prices in first (base) year}}$$

$$= \frac{5.50 + 32.50 + 2.25}{15.32 + 31.35 + 1.70} = 83.21\%$$

Average prices are 83.21 in 2016 referenced to 2014 (which is the base year), which is a reduction largely due to the drop in prices for flash memory. The defect in this index is that it assumes the quantity purchased in each year for each item = 1. Obviously, one approach is to introduce average quantities bought in the base year.

If we average the first two years as the base, we change the denominator.

- Average price of flash memory (15.32+9.45)/2 = 12.39
- Average price of paper (31.35+32.67)/2 = 32.01
- Average price of a cup of coffee (1.70+2.05)/2 = 1.88

Now we use these in a new expression

$$= \frac{5.50 + 32.50 + 2.25}{12.39 + 32.01 + 1.88} = 86.97\%$$

In 2016, average prices are 86.97% of those in 2014, but this formulation, while adding some information, only goes part way. We need indexes that reflect changes in quantities and changes in price.

Combining price and quantity relatives

A price index must include quantity using a fixed basket of goods while tracking the cost of this basket over time. Baskets include physical goods (coffee, pens, …) and services (legal services, Netflix shows, …). Each good/service in the basket is weighted by its “importance,” which generally means the proportion of annual income devoted by the “average” consumer to its purchase. Two formulas (that seem identical, but are not) form the basis for most consumer price indexes:

$$\text{Laspeyres Index} = \sum PnQo / \sum PoQo$$

$$\text{Paasche's Index} = \sum PnQn / \sum PoQn$$

The Laspeyres index uses the basket of goods defined at the start of the years that form the index. It "freezes" consumer preferences, buying patterns, tastes, and technology. For example, a Laspeyres index over 15 years would not include the costs of streaming services. In contrast, the Paasche index uses the basket of goods at the end of the observation period.

Using the data from the student computing needs above , the Laspeyres index appears as

$$\sum P_n Q_0 \sum P_0 Q_0 = \frac{\sum(\text{Prices in 2016})(\text{Quantities in 2014})}{\sum(\text{Prices in 2014})(\text{Quantities in 2014})} = \frac{\sum(5.50*1)+(32.50*3)+(2.25*1)}{\sum(15.32*1)+(31.35*3)+(1.70*1500)} = 1.308$$

and the Paasche index appears as

$$\sum P_n Q_n / \sum P_0 Q_n = \frac{\sum(\text{Prices in 2016})(\text{Quantities in 2016})}{\sum(\text{Prices in 2014})(\text{Quantities in 2016})} = \frac{\sum(5.50*2)+(32.50*.5)+(2.25*1617)}{\sum(15.32*2)+(31.35*.5)+(1.70*1617)} = 1.311$$

	Guns		Butter	
	Quantity	Price	Quantity	Price
2000	405	102.5	2001	0.57
2001	456	103.6	2002	0.58
2002	315	121.6	1999	0.59
2003	500	130	1889	0.6
2004	481	131	2010	0.61
2005	501	134.6	2011	0.62
2006	456	136.6	2500	0.63
2007	402	136.1	2489	0.64
2008	401	136.9	2222	0.65
2009	465	140.1	2117	0.67
2010	500	142.1	2220	0.69
2011	510	142.2	2300	0.71
2012	512	150	2209	0.74
2013	516	151	2117	0.77
2014	520	149.4	2400	0.8

Example: Use these data to compute a Laspeyres index with base year = 2000 and has a Paasche index with a base year = 2014. This is an excellent application for the = SUMPRODUCT function.

4. Summary

This module introduces working with time series data. It shows how logarithms, lags, and indexes reveals patterns in the data that support a deeper analysis of relationships among variables and trends. Economists will need to create specialized indexes to meet specific needs. For example, many central agences created a "stringency index" to summarize the various COVID measures (vaccine mandates, travel restrictions, mask policies, etc.). Any index must try to meet the conditions summarized on page 17.

Annex: Key Excel functons and formulas

Function/Formula	Example	Discussion
=SUM(number1, number2…)	=SUM(A1:A10)	Sums the numbers in an array, which may be rectangular. =SUM(A1:F35) sums 210 numbers.
=SUMPRODUCT(Array1, Array2, Array3…)	=SUMPRODUCT(A1:A10, B11:B20)	This example is the same as writing =(A1*B11)+(A2*B12)+…+(A10*B20). Do you need the brackets? Why/Why not? The arrays must have the same dimension. Try examples with rectangular arrays =SUMPRODUCT(A1:B20, C1:D20).
=SUMIF(range1, criteria, range2)	=SUMIF(A1:A10,”>=9”)	Sum the values in A1:A10 that are greater than or equal to 9). This combines a math formula with a logical formula (=IF).
Moving Average (k)	=AVERAGE(A4:A4+k)	Take the average of the first k numbers in the column, then drag and drop for the rest of the series.
Lags	=B2-B1	Following the standard practice of rows designating time (usually days, weeks, months, years), then a lag is simply subtracting cells (later-earlier).