# 9 Data from taxation systems (Module 9)

### Objectives

This module aims to provide an understanding of:

* Taxation systems as sources of economic data
* The prerequisites for exploiting taxation data
* Three tax systems and their use, in more detail
* How to check the data and deal with exceptional values
* The advantages and disadvantages of taxation data

## 9.1 Taxation systems as a source of economic data

This chapter is not about taxes. It is not about how taxes are recorded in the national accounts. This chapter is about exploiting taxation systems for key (non-tax) data they contain. The statistics derived from taxation systems may be useful for many purposes. And if they can be made available to the national accounts team in electronic form, they are especially valuable for measuring GDP. While the data on the tax itself may also be of interest, it is not the primary focus. These taxation systems include:

* Customs procedures (designed to collect import duties and other taxes on external trade
* Tax returns on business income and profits
* Income tax deductions from pay
* Social security contribution systems
* VAT returns (or similar tax based on sales)

#### Customs procedures

International merchandise trade – the import and export of goods – is one of the oldest sources of tax revenue for governments. In India, Customs duties were introduced in 1786. In most former colonies (including in the United States of America before Independence) Customs duties were the major source of imperial revenue. The Customs clearance documents – “Bills of Entry” in English – have always been the primary source of statistics on this trade (taxable or not). The documents contain several pieces of information, including the type of trade flow, a description of the goods, the country of origin or destination, the weight or quantity, the value for Customs purposes and the taxes due, among other items.

These days, most Customs authorities require traders (or their agents) to supply the information in electronic from, or (if not) capture the data on the paper document. The tax can then be calculated automatically, and the foreign trade statistics produced at little extra cost. [E- learning course on IMTS available from UNCTAD]. The overall values of imports and exports also feature in the balance of (international) payments, and in the expenditure measure of GDP. They are also used for estimating the volumes of these flows, for which the details are necessary. It is more complicated than merely totalling up the net weights [see Module D Ch.3]. The details by commodity are also important in constructing a Supply Use Tables which portray the flow of goods and services in the economy.

This chapter focuses on three other tax systems. Typically, these systems have not been exploited as much as a source of data as they could be. There are obstacles to overcome.

#### Tax returns on business income and profits

Most countries have had systems of income tax returns, as made by both companies and unincorporated businesses, are potentially important sources for measuring GDP. In former Francophone countries, the *Déclaration Statistique et Fiscale* was, and in some countries still is, a key source of data for the national accounts. It is a return that is completed annually for both tax and statistical purposes. In highly developed economies, in which most businesses are taxed, the business income tax system has long been used as a source of data on profits, a key component of the income measure of GDP.

#### Income tax deductions from pay

Total wages and salaries (employment income) is also a component of the income measure of GDP. Tax on this income is often collected using the system known in the United Kingdom as “pay as you earn” (PAYE). Employers deduct tax from employees’ pay depending on their wages (and maybe a code) every week or month. They send the money with a return to the tax authority.

#### Social security contribution systems

Social security contributions are often collected in the same way. The money may go into a social security fund and used to pay benefits to those eligible for them. The returns usually show the number of employees and their total income as well as the total tax and/or contributions being deducted and paid to the relevant authority.

#### Value added tax returns

A more recent type of tax, adopted in many countries only since the 1990s, is Value Added Tax or VAT. It is usually collected every month or quarter from enterprises with more than a minimum annual turnover. Enterprises charge their customers VAT on all their sales (output tax). But if they have paid VAT already on their purchases (their “input tax”) this tax is deductible. Only the difference needs to be paid to the tax authority. Thus (as its name suggests) the tax falls on the “value added” by the enterprise, the difference in value between their (taxable) sales and their (already VAT-paid) purchases. Importantly, VAT is also collected by the Customs Authority on the value of imported goods. In essence, VAT is a tax on the final consumption of goods and services, but it is collected from every registered enterprise involved in their production and distribution along the way. See Section 3 for more details. Returns are completed by registered enterprise to show how the tax payable has been calculated. In particular, they will show the value of sales (both taxable and non-taxable). This information is immensely valuable for measuring GDP.

In some countries, a similar tax based on sales (whether or not it is deductible by registered traders) may be operating, but under a different name, such as Goods and Services Tax (GST), Consumption Tax, etc. If they cover a wide range of activities, the information on the value of sales is of enormous value.

“Excise” duties are another possible data source, but they are only levied on particular goods whether imported or locally produced. Examples are taxes on alcohol of various kinds, petroleum, and tax on airtime.

## 9.2 Prerequisites for exploiting taxation data

In this section, we discuss the main preconditions for exploiting taxation systems for statistical purposes.

First of all, considerable effort and investment are needed to change the way things are done. Many people and organisations are resistant to change. They prefer a quiet life. They prefer to do things in the way they have always been done. This is normal. But the world moves on. Change is here to stay.

#### Taxation systems

The first prerequisite concerns the way the taxation systems themselves are organised. Is there a central database? In large countries the systems are likely to be decentralised. This may make it difficult to obtain the data. In a small country, major taxpayers are likely to be found in the capital city. This is a good starting point.

In the past, income tax returns have usually remained on paper in filing cabinets. There was little chance of exploiting them easily. But if now they are available in electronic form, they could be exploited more easily.

#### Data sharing

Tax Authorities can be very reluctant to share the data they have, for good reason. The tax laws usually say that the data on individual tax returns is confidential and cannot be given to any outsider. There are usually some exceptions, but the National Statistics Office is unlikely to be mentioned explicitly.

The Statistics Law may already require National Statistical Offices to respect the confidentiality of individual data. This Law may also authorise the NSO to require government departments (among others) to provide data. But, if so, the Tax and Statistics laws conflict.

If the parties recognise the benefits of data-sharing, it should be possible to come to an arrangement. A Memorandum of Understanding is useful to ensure the arrangement is sustainable. Second best would be to accept aggregated statistics rather than individual records, but this makes identifying errors less precise.

#### Activity coding

Except for overall income estimates, the taxation data are of little use unless the activity of every business is known. When businesses register for tax purposes, they are usually asked to provide details of their activity. Certain ratios (implicit in a return) depend on the activity. This helps the tax collectors to validate the returns. Many tax departments know about the International Standard Industrial Classification (ISIC) and use it. If not, however, considerable investment will be required to code the individual enterprises according to ISIC.

#### Investment in (IT) systems

If the tax authorities agree, large quantities of tax data will arrive at regular intervals. What is to be done with them? As with surveys, NSOs need processing systems to check the data for extreme values (see Section 6) and to prepare aggregate time-series for use.

## 9.3 Value added tax and its use

This section describes the way VAT works and what the data may be used for.

#### The VAT system

Value Added Tax (VAT) is a very interesting tax. It is worth understanding how it works. As stated in Section 1, registered enterprises (“traders”) must charge VAT to all their customers at the appropriate rate. However, they do not pay this amount to the revenue authority. They deduct the VAT they have already paid either to registered suppliers, or to the Customs authority on goods they have imported. For tax purposes, the value of the goods or services supplied includes all duties and/or excise-type taxes on the products. There is a threshold below which businesses are not required to register. Such businesses are “exempt” from VAT.

To illustrate the way it works, we shall consider a very small country called VATLAND. In this country the people love clothes. They import cotton and some other essentials (A) to make them. There is one factory (B) making cloth from the imported cotton. Another factory (C) is making fabulous clothes. There are two shops. One is a large enterprise (D) that sells the clothes (and a few other essentials) to households (and to other businesses). It also exports a lot of clothes that are made in the clothes factory. Exports are “zero-rated” (the VAT rate is zero). There is also a small shop (E), which is not registered for VAT. In Vatland, if your turnover is less than 500,000 Vatland Rials, there is no need to register: you are “exempt” from VAT. This means you do not charge VAT on your sales, but you must pay it when you import or buy from a registered trader. There is another business (F) which provides all the services the other businesses need. Finally, there is Government (G) which collects the revenue and Households (H) on whom the tax ultimately falls.

Typical transactions of these economic units, together with associated VAT are shown in the Table 9.1. It is assumed the standard rate of VAT is 10 per cent. This example is idealised. In practice, no VAT system will be so simple! There are several points to note about this table:

* The sales of the businesses are shown on the right, while the purchases (costs) are shown on the left. (It could be shown differently.)
* The purchases made by one business from another are also shown as sales of the latter to the former; For example, C is shown buying cloth from B for 1,200 kVR plus 120 VAT and B is shown selling cloth to C for the same amount.
* The VAT that registered businesses charge to customers (“output” VAT) is in the last column.
* The “deductible” VAT that registered businesses have paid to suppliers (their “input” VAT) is shown on the left in the third column of the table.
* The “Net VAT” payable to the revenue authority by each business is calculated by adding up all the output VAT and subtracting all the input VAT.
* The government gets the revenue from two sources: imports and net VAT. It is often the case that in a small open economy more VAT may be collected on imports than on value added within the country.
* The total is ***identical*** to the VAT implicit in the purchases of households: the tax charged to them directly (200kVR) and that paid by small businesses (30kVR).
* The percentage of tax included within the full (“market”) price paid by households to registered traders is not 10%. It is 200/2200 which is 9.1%. Overall the percentage is lower still at 8.7% (230/2650)

### **Table 9.1: Application of Value Added Tax**



#### **The VAT returns**

Every moth or quarter, registered enterprises fill in a VAT return. (There may be other VAT forms for different purposes, but we are focusing on the one used by most businesses to accompany the net VAT payable in each period.) These returns vary from country to country, so it is worthwhile getting copies from the tax department website.

Usually the return is just one page. At the top will be data identifying the taxpayer including the Tax Identification Number (TIN). Taxpayers will usually be allocated a unique TIN when they first register with the tax authority. Ideally this same number will be used by a taxpayer for all tax purposes. It is then possible to link up the data supplied for the different tax regimes.

Then there will often be a column for the value of different types of sales (or, to be more accurate, “supplies” which can have a meaning wider than sales) and another column for the corresponding amount of tax. The types of sales usually include at least

* Taxable supplies (value and output tax)
* Zero-rated supplies (value only)
* Exempt supplies (value only)

Then there may be details of the value of purchases and certainly the amount of input tax paid. The latter may be split between tax paid to Customs (on goods directly imported) and the tax paid to other registered traders. Some input VAT may not be deductible.

There will be space on the return to calculate the net VAT due to the revenue authority, and there may be adjustments to be made from previous periods and so on.

Normally the data will be sorted in a database at the tax department. The data required each period can be extracted in (say) .csv format with a standard query.

#### **What is VAT data used for?**

The main use of VAT data by NSOs is to provide short-term indicators of the value (and after deflation, the volume) of turnover (sales) in the (formal) economy, classified by activity.

Although the indicators are limited to the formal sector, this sector (important from the point of view of government revenue) is extremely well covered, given a full set of data. The returns constitute a Census of the formal sector in every period. They provide a solid base for monitoring overall economic activity.

There are no easy solutions for monitoring the informal sector. And it is by no means easy to measure the *value* of agricultural production. Different methods are needed in such cases. But for very many activities, VAT data may cover most of the production, and assumptions can be made (at least in the short term) about the informal part.

Can the data be used to measure value added itself? There are two things that make this difficult. One is that sales less purchases are only a very rough measure of Gross Value Added as defined in the SNA. For example, purchases may include capital equipment and not just raw materials and consumables. Another problem is the tendency of businesses to minimise their taxable turnover, while maximising declared expenditure. So, while a series of turnover data may give a good indication of trends (the best available), the data may not be so good as a measure of the actual level of value added (see Section 6 below).

Another potential use of the data is to alert the tax authorities to atypical returns that ought to be challenged. Tax audits usually reveal some undercounting of sales.

## 9.4 Annual business income tax (BIT) returns and their use

#### BIT systems

Income tax systems and the way they are applied vary between countries. Normally there will be a distinction between incorporated companies and unincorporated businesses. The system may or may not apply to publicly owned enterprises. Non-profit-making bodies are normally not subject to income tax. Some enterprises may have “tax holidays”.

The tax is charged on profits (net income). The definition of profits for tax purposes usually differs slightly from the definition used in commercial accounts. Some items may be disregarded, but certain allowances may be given instead.

#### **BIT returns**

As mentioned above, in former Francophone countries, the *Déclaration Statistique et Fiscale* was the return on which tax would be assessed. The return would be based on a *Plan Comptable* (chart of accounts), which was enshrined in the law and which specified in detail how to classify all the transactions of an enterprise.

Most countries don’t follow such a rigid accounting plan. International commercial accounting standards do not specify what details to show when presenting accounts. But they do define key items such as total turnover and before tax profits. And tax authorities will usually want to know about certain key items that are included in the accounts

#### **What is BIT data used for?**

There are two potential uses for BIT data, if the prerequisites are fulfilled:

* Improving the coverage (on an annual basis) of the VAT data for measuring turnover.
* Determining key ratios such as
  + margins on traded goods
  + intermediate consumption to total output

Business income tax usually covers many more enterprises than the VAT system. The reason is the VAT threshold, below which a business need not register as a VAT trader. When the annual turnover figures become available from the BIT source they can be compared with the more frequent VAT data. The annual differences can be spread over the VAT periods to maintain consistency and projected forward for more recent periods.

The use of BIT data for determining ratios may be more problematic. In principle more of the necessary data for doing so should be available than is the case with VAT data. But the quality of the details may not be adequate without a lot of editing (see Section 7). Because of this it is unlikely that aggregate data will help. Individual records are necessary to validate the data properly.

9.5 Deductions from employment income

#### PAYE systems for income tax and/or (compulsory) social security

In many countries, compulsory deductions are made from wages and salaries paid to (formal) employees. These deductions are either income tax payments or social security contributions (or both) made by employees. The employer must pay the money to the relevant authority. Again, “formal” in this context means that both the employee and the employer are registered with the tax or social security authorities.

Many employers in these schemes may be exempt from registering for VAT. NGOs are a good example.

#### **PAYE returns**

The exact nature of these pay-as-you-earn schemes varies from country to country. However, normally, every month, employers must make at least a summary return to the tax department and/or social security fund. The return for each scheme probably includes the following three pieces of information:

* The number of employees
* The total wages and salaries paid
* The total deductions made

At the end of the fiscal year, the employer may also have to provide authorities with figures showing the annual pay and deductions for each employee separately. We shall not consider the annual data in this session but focus on the monthly return.

#### **What is PAYE data used for?**

If the summary information specified above exists, and the other prerequisites (Section 3) are met, then it is easy to compile the total number of formal employees, classified by the activity of the employer, and their total employment earnings. By dividing the latter by the former, the levels of average earnings in each activity can be calculated every month. Although not pure price indices, these statistics give an indication of (formal) wage inflation.

These statistics provide estimates of a key part of formal sector value added. But the coverage is limited in countries where much of the employment is informal.

## 9.6 Checking tax data and dealing with outliers

#### Checking at micro level

As we have seen, tax systems are potentially a great source of economic data. But, in common with most numerical information, the data can contain errors. This is especially true of financial data that often have 10 or more digits. One extra digit and the figure will be ten times too big; one digit less and it will be ten times too small. But these errors are not offsetting.

**Exercise: Suppose there are two identical values, one with a zero too many at the end and the other with a zero too few. Verify that the total of two will be 405% higher than it should be.**

Occasionally the errors can be even more extreme. They must be detected and corrected wherever possible. This is perhaps the most important task of a compiler of statistics. Because the datasets are large, an automatic procedure for identifying outliers is essential. In any one period, it is almost impossible to say in isolation whether a turnover figure is correct or not. But when it is compared with a series of values in previous periods, it is much more obvious.

A very simple procedure is to compare the figure with that in the previous period (and/or the same period in the previous year). No need for a long series with his method. However, the biggest (*absolute*) differences may come from the biggest values, which may not be surprising. On the other hand, big *percentage* differences can come from small numbers that have little influence on the total. A useful compromise is to use the *difference of square roots*. This measure focuses attention on questionable changes.

More sophisticated methods are becoming easier to implement. For example, at the time of writing, the latest version of Excel contains time-series forecasting functions that can provide an expected range for the next value, taking account of seasonality. If this next value falls outside the range, it could be highlighted for further examination.

The checking should be done as soon as the dataset is received from the source. The largest outliers should be referred to the source. What is the reason for the big change? Or could it be an error? If no explanation or amended figure is received, it will usually be correct to substitute a more normal value. A record should be kept of such amendments.

Often, in the next period, the number may go back to normal. This would reinforce the view that the previous number was abnormal. If, however the numbers continue to be quite different from what came before, then it is likely a real change has taken place. In this case the statistician should try hard to find out why the change has occurred.

#### Checking at macro-level

The tax databases will contain information on the tax payable. It is worth checking how these figures compare with the government’s revenue account. They will differ, but the differences should not be large. One difference is due to timing. The revenue is not received in the same period to which the tax refers. If the differences are large, there should be an explanation.

In the case of VAT, if in a period the value of taxed purchases exceeds that of taxable sales, net VAT will be negative. In principle, this amount is repayable to the taxpayer by the revenue authority. In practice, the credit may be offset against tax payable in respect of succeeding periods. Again, the way VAT on the imports of VAT traders is accounted for may not always be as shown in the example given in Section 4 above.

## 9.7 The advantages and disadvantages of taxation data

In this section, we summarise the advantages and disadvantages (pros and cons) of using taxation data for measuring GDP. However, surveys are also essential for many purposes. We consider the problem with surveys, as well as their advantages.

#### The advantages of tax data

Given the availability of Customs data, it would be unthinkable to have to do a survey to obtain detailed statistics on imports and exports. (However, some countries supplement the Customs data by carrying out surveys of informal cross-border trade.)

Similarly, if databases of other types of tax returns are available, why not use them? The main advantages are the following

* There are no survey or data capture costs
* The returns contain key variables needed for monitoring the economy
* They cover all business taxpayers (no sampling errors),
* Penalties for non-compliance are severe (little non-response)
* Using the data as control totals, the sample size of enterprise surveys can be greatly reduced

This last point is particularly important when the data provide an *auxiliary* variable for estimating the population totals of the sample values of correlated variables. Put another way, the results of a survey designed to measure ratios (not totals) will be much less variable and so less prone to sampling error (which can be very large for totals).

#### The disadvantages of tax data

The main disadvantages are these:

* The returns are designed for tax purposes. Some data needed for statistical purposes may be missing.
* No direct contact is possible with the taxpayers
* Non-taxpayers are not covered. Because they can be large in number but relatively small, they are a more suitable population for sampling. A household survey can be used to obtain information about their activities.
* Work is needed to maintain good relationships and close cooperation with the tax authority.

Another disadvantage, especially if the tax is new, might be that the coverage of the system could increase over time as more and more enterprises are included in the system, depending on the success of the tax authorities in ensuring compliance. However, there are techniques that can be used to mitigate such effects.

#### The problem with surveys

Surveys have some real advantages (see below). But perhaps the biggest problem for small economies is the cost. It is a well-known statistical law, that accuracy depends on the size of the sample and not on the size of the population. In a small country the cost, relative to the government budget, can be huge. If the same data, for example, quarterly turnover, can be obtained from the tax system, there is no need for a survey.

Accuracy also depends on the variability of the data. In a regular survey of prices for the CPI, the sample size can be relatively small, because the price of a specific item in each location will not vary very much. This is not the case when measuring the value or volume of production. Everyone knows the local price of [a bowl of noodle soup]. But how much [soup] was sold locally (let alone nationally) last month (and how much more or less than in the previous month) is a totally different matter. It would be extremely costly to find out.

Enterprises vary enormously in size (much more than households do). This can be an advantage if we can identify and get information from the largest enterprises on a regular basis, as they will account for a large proportion of activity. For national accounts, the financial statements, particularly the profit and loss accounts, are the main source of the information required. However, some enterprises may be reluctant to disclose such information. This can be a problem, as a National Statistical Office does not same capability or capacity to ensure compliance that the tax authority has.

For more comprehensive surveys of enterprises, we need:

* An up-to-date Statistical Business Register
* A stratified sample (with the largest businesses completely enumerated)
* Survey instruments and enumerators
* The willingness of the respondents to take part and respond (not guaranteed)
* Procedures for dealing with non-response, missing values, data errors, etc

Such surveys are challenging, to say the least.

#### The advantages of surveys

Two main types of surveys are used to compile national accounts: enterprise surveys and household surveys of living conditions.

High quality household surveys of living conditions are a valuable resource. They have their limitations. But they can provide insights into household expenditures and informal activities that are important for measuring GDP as well as for assessing poverty levels and providing weights for the CPI. They are certainly essential for establishing benchmark estimates of the level of GDP and its components using the supply-use framework. For several items needed in the national accounts and the balance of payments, household surveys are the only source.

Surveys of (formal) enterprises work best when complementing the information obtained from tax systems. For example, they can provide a breakdown of the turnover between sales of merchandise (goods purchased for resale) and sales of good or services produced by the enterprise. They can provide details of both the costs of merchandise sales (and hence the mark-ups) and the costs of production. They may be used to identify the extent of foreign ownership and to provide details of their transactions with the rest of the world not available from other sources. They can be targeted at specific activities.

## 9.8 Exercises and discussion points

#### General

* Make a list of the taxation systems operating in the country.
  + Which of them is being used for statistical purposes?
* Draw a diagram showing the flows of VAT in Vatland.

#### If the NSO is not using VAT data

* To what extent are each of the prerequisites preventing the NSO from exploiting VAT data?
* What would it take to overcome these impediments?

#### If the NSO is exploiting tax data

* What procedures are being used to identify extreme values?