**CMT212 COURSEWORK 2 – DATA REPORT**

**SECTION 1: Links to Data; Introduction.**

https://www.ons.gov.uk/economy/nationalaccounts/supplyandusetables/datasets/inputoutputsupplyandusetables

The linked data represent the UK supply and use tables (SUTs) for the years 1997 to 2014. The data in these tables underpins the annual GDP estimates for the UK, illustrates the process used to compile those estimates, and provides a more detailed picture of the economy than GDP alone. Moreover, study of the SUTs can reveal potential problems in compiling this extremely ambitious dataset.

That said, from the perspective of much econometrics (and statistics in general), the data has important limitations which make it vital to understand exactly what the data is saying and how it can legitimately be used if any analysis is to be valid.

Below, we first take a brief look at the structure and purpose of the SUTs (what data is included, and how this data is linked) and illustrate this using a simple visualisation. We then describe some of the limitations inherent in this structure and some of the practical issues involved in populating it with estimates. This theoretical analysis indicates some quite restrictive limits on what can be “legitimately” analysed, which allows us to select analyses on the basis of this theoretical discussion rather than try many different ones until we find statistical significance. This is interesting in its own right, and allows us to avoid the undesirable practice of “p-hacking” (insert refs).

Based on this theoretical discussion, we select two pairs of sub-datasets of the SUTs, subject them to simple statistical evaluations, and describe the results verbally and with further visualisations.

**SECTION 2: General Data Report.**

**SECTION 2.1: THE STRUCTURE OF THE SUTs.**

The SUTs contain the following economic variables, broken down by “industry” and/or “product” as shown:

|  |  |  |
| --- | --- | --- |
| **Variable** | **Industry?** | **Product?** |
| Output | Yes | Yes |
| Intermediate Consumption | Yes | Yes |
| Compensation of Employees (COE) | Yes | No |
| Gross operating surplus (GOS) | Yes | No |
| Taxes (net of subsidies) on production | Yes | No |
| Household final consumption expenditure (HHFCE) | No | Yes |
| Non-profit institutions serving households (NPISH) final consumption expenditure (NPFCE) | No | Yes |
| Central government final consumption expenditure (CGFCE) | No | Yes |
| Local government final consumption expenditure (LGFCE) | No | Yes |
| Gross fixed capital formation (GFCF) | No[[1]](#footnote-1) | Yes |
| Changes in inventories | No | Yes |
| Changes in valuables | No | Yes |
| Imports[[2]](#footnote-2) | No | Yes |
| Exports | No | Yes |
| Taxes (net of subsidies) on products other than those on production | No | Yes |

**SECTION 2.2: WHY THIS STRUCTURE?**

This may seem like a random assortment of economic variables, but in fact these variables are closely linked. Definitions of the terms above, the rationale for assembling them in one place, and the economic reasons why they are linked are explained in detail in the Appendix. Here, we can summarise the conclusions reached there as follows:

* The variables broken down by industry show how firms (and firm-like units) in each industry produce UK GDP (output, intermediate consumption), and how this production is distributed (COE, GOS).
* These industry variables are of interest partly because they make up the terms in the *production function*, a basic concept in economics which informs much empirical work in the field.
* The variables broken down by product show how each product used in the UK is supplied (output, imports) and used (all other variables broken down by products).
* These product variables are of interest partly because they allow more accurate deflation of these nominal values to produce constant-price or real GDP and GDP components.

These linkages can be visualised as follows:

[insert vis 1]

**SECTION 2.3: LIMITATIONS OF THE SUTs**

In the previous section, we asserted that the variables shown in the SUTs are linked and that these linkages are the primary reason for the structure of the SUTs. Investigating these linkages is therefore an obvious focus for the analysis. Before going further, however, we should note some limitations of the SUTs and some features of the analysis which might mitigate them.

* *Nominal Values.* All values in the SUTs are nominal, not real. As is well-known, this makes them unsuitable for time-series analysis without taking this into account (Tufte 1983). However, this difficulty can be avoided by analysing *ratios* of two variables, rather than the variables themselves – because the units which were in nominal £s disappear, so the series can be analysed over time. More intractably, we cannot say anything about deflation using the SUTs alone, as there is no price information.
* *Annual Values*. All values in the SUTs are annual; there are no quarterly or monthly versions of the SUTs. In effect, this means that series without industry and/or product breakdowns will have insufficient sample sizes for worthwhile analysis, so the analysis should be of these breakdowns rather than the totals.
* *Little Metadata.* Most professionally produced surveys are published with quality metadata such as coefficients of variation, as recommended in the Statistical Quality Framework. The SUTs are not. Moreover, they were stripped of their National Statistics status in 2014, primarily because the main source of product detail for intermediate consumption, the Purchases Survey (PS), was discontinued in 2005; 2004 patterns have been used ever since insert ref. (The PS was reinstated in 2016 and will be used for this year’s SUTs for the first time since 2005.) Since Limitation 2 means that we should be analysing industry and/or product detail, analysis should avoid intermediate consumption.

Combined with the economic linkages described earlier, these limitations impose severe restrictions on the choice of series to analyse. We want an analysis that tells us something about the production function (since deflation analysis cannot be performed using only the SUTs); which uses ratios of variables, rather than the variables themselves; and which avoids intermediate consumption.

**SECTION 2.4: ANALYSIS CHOICE 1: WAGE SHARE.**

These restrictions suggest a particular variable to analyse: *wage share*. Wage share is defined in Schneider (2009) as compensation of employees divided by GVA. It is, therefore, a ratio. Moreover, it relates to the shape of the function defined by the right-hand side of the production function equation (Equation (3) in the Appendix), so it deals with the production function without involving intermediate consumption. Moreover, an industrial analysis is possible. It is reasonable to assume that manufacturing and non-manufacturing industries have different technologies, and the economics literature on the wage share includes industrial studies. We follow de Serres et al (2002) who argue that over 1975-1995, in several developed countries, industrial change contributed to the wage share change – specifically, a decline in manufacturing led to a decline in whole-economy wage share because manufacturing industries had higher wage shares. Although the SUTs do not have the data to replicate their analysis, we can investigate whether the basis of this claim holds in the period 1997-2014 – that is, whether manufacturing industries have higher wage shares than non-manufacturing industries.

**SECTION 2.5: ANALYSIS CHOICE 2: EFFECT OF THE PURCHASES SURVEY.**

Limitation 3 also suggests a further area of possible analysis. The last reference year in the dataset is 2014; 10 years before that, the last PS used in these SUTs was collected. To assess the effects of this on the 2014 dataset, it may make sense to analyse intermediate consumption product detail for two reference years while the PS was still being collected. This is of interest because it provides a look behind the sparse metadata for the SUTs at a major quality issue with a key economic dataset.

**SECTION 4: Data Report.**

**SECTION 4.1: WAGE SHARE DATA REPORT.**

1. The wage share is defined as follows (Schneider 2009): for industry i in period t, the wage share WS is given by
2. As mentioned above, WS has no units because both the numerator and denominator of the LHS of the equation in 1 are in £ of the year t, so they cancel. It can therefore be analysed over time.
3. The wage share for the whole economy is shown in the following (non-interactive) visualisation.
4. The whole-economy wage share timeseries is remarkably flat, which confirms the general empirical observation Schneider (2009) refers to as “Bowley’s Law.”
5. The wage shares for each year, by manufacturing/non-manufacturing industries, are shown in the following visualisation. Following de Serres et al (2002), we classify mining industries as “manufacturing.”
6. The scaled dots in the visualisation are for interest only: in the actual analysis we don’t scale for industry size, as we are not trying to explain whole-economy trends but to look for the effect of being manufacturing.
7. The mean for the manufacturing industries looks generally higher than the mean for the other industries (except in 1997). There is a clear outlier year in 1998 where this is marked, but in all subsequent years, a mild effect looks visible.
8. In the visualisation, also note the large non-manufacturing industry with WS = 0 for all years. This is imputed activity of owner-occupiers – additional GVA (and output) representing the accommodation services provided by home-owning households to themselves. For obvious reasons, this always has zero employment and therefore a wage share of zero.
9. Another industry of interest is oil, gas and metals extraction (the larger black dot in the bottom left segment). This has a low WS, presumably because of the profitability of the oil industry. It may therefore decrease any effect measured by the analysis.
10. Finally, note the construction industry (near the middle of the chart). This also has a fairly low WS, which highlights an issue with the analysis: we do not adjust COE to include mixed income (in the UK, the profits of self-employed small traders), which in the SUTs is part of GOS. Most economics do make this adjustment (Schneider 2009), including de Serres et al (2002). This potential error has the opposite effect to that in the previous bullet point.
11. Also note a few outliers with wage “shares” higher than 1; this occurs when COE is higher than GVA, because of negative GOS and/or net taxes.
12. There are a few cases of negative wage shares, where GVA is negative. These do not have economic meaning and have been cleaned from the regression below.
13. We test whether this is statistically significant using the simple linear regression on a single dummy variable set to 1 for manufacturing industries in the python file (see python file wageshare\_manuf\_regression\_cleaned). This is equivalent to a t-test except that it gives the intercept and slope (interpretable in this case as the expected wage share difference between a manufacturing and non-manufacturing industry).
14. The p-values for all years except 1998 are well above 0.05, indicating that there is **no** statistically significant difference between the wage share of manufacturing and non-manufacturing industries.
15. We can conclude that there is no evidence that the production functions of manufacturing and non-manufacturing industries are different with regard to the use of labour – a surprising result.

**SECTION 4.2: INTERMEDIATE CONSUMPTION SHARE DATA REPORT.**

1. We want to investigate the effect of not running the Purchases Survey using the product proportions of the total. For clarity, the product proportion for product p is defined as, where there are n products
2. To assess the effect of not using the Purchases Survey, we look at the change in PP across the period since it was run (2005-2014) and compare it to the longest possible time span when it was available (1997-2004).
3. If not using the Purchases Survey is a problem, we would expect the change in PP to be significantly smaller in the later period, when product proportions were not updated from the 2004 PS result.
4. We cannot use PP changes directly, as the changes would obviously be autocorrelated with the general economic importance of the product.
5. We therefore use the percentage change in the share – that is, the change in PP divided by the PP in the start year.
6. We further take the absolute of this value – as we are interested in the size of the change, not its sign. Clearly, any positive or negative share error would have to be “compensated” by changes of the opposite sign.
7. Clearly, the true position of the economy changed over the two examined periods, probably in different ways. We need to adjust the PPs to account for this – as, otherwise, the difference is unrelated to the phenomenon we’re trying to measure.
8. We do this by a simple shift-share adjustment inspired by insert ref. For each period, the IC values by industry and product for the end year (2004 or 2014) are divided by the total IC value for each industry for that year (to give a series of shares equivalent to PPs but for specific industries) and multiplied by the corresponding value for the start period (1997 or 2004). The results are then aggregated to product totals to give the adjusted end-period data. This procedure models what the end period results would have been if the industry distribution of the economy had remained as it was in the start period.
9. This adjustment may not cover all the relevant economic changes over the period – but if we could do that, there would be no problem, as we would know the IC data to be accurate!
10. Again, we perform a simple linear regression of the absolute percentage change on WITHPS, which is 0 for the later changes and 1 for the earlier ones.
11. As the p-value is well above 0.05, we can conclude that the analysis has not detected any significant effect.
12. This is reinforced by the fact that the coefficient for WITHPS is actually negative; that is, if there is any effect the changes in the earlier period are smaller than the later ones – the opposite of the predicted effect if a problem existed.

**APPENDIX: DETAILED THEORETICAL DISCUSSION.**

**SECTION A.1: GENERAL THEORETICAL BASIS.**

The SUTs are described in detail in Eurostat (2008). To summarise, the SUTs rely on the fact that GDP is an accounting balance which has been defined such that it can be calculated (from different surveys and administrative data) in three separate ways, with the true values of the results always being equal (Eurostat 2008 pp. 46-62).

**SECTION A.2: DEFINITIONS OF VARIABLES**

|  |  |  |
| --- | --- | --- |
| **Variable** | **Definition** | **Reference** |
| Output | The sum of all products created during the reference period, valued on a defined basis (see next section) | Eurostat (2010), p. 55, section 3.14 |
| Intermediate Consumption | The sum of all products consumed during production processes (ie, raw materials), valued on a defined basis (see next section) | Eurostat (2010), p. 68, section 3.88 |
| Compensation of Employees (COE) | Remuneration of employees for labour inputs into the production process | Eurostat (2010), p. 87, section 4.02 |
| Gross operating surplus (GOS) | The residual from output after IC, COE and taxes on production have been deducted | Eurostat (2010), p. 200, section 8.15; and p. 201, table 8.3 |
| Taxes (net of subsidies) on production (D.29-D.39) | Taxes incurred as a result of production that do not vary with the amount or price of production (eg, business rates) *less* subsidies on the same basis (eg, CAP payments) | Eurostat (2010), p. 93, section 4.22 |
| Household final consumption expenditure (HHFCE) | Expenditure by households to satisfy own needs/wants, not for production | Eurostat (2010), p. 70, sections 3.95 and 3.96 |
| Non-profit institutions serving households (NPISH) final consumption expenditure (NPFCE) | Expenditure by NPISHs (most charities, unions, religious organisations, etc) to satisfy individual or collective needs or wants | Eurostat (2010), pp. 70-71, section 3.97 |
| Central government final consumption expenditure (CGFCE) | Expenditure by central government (inc. devolved administrations) to satisfy individual or collective needs or wants | Eurostat (2010), p. 71, section 3.98 |
| Local government final consumption expenditure (LGFCE) | Expenditure by local government (inc. devolved administrations) to satisfy individual or collective needs or wants | Eurostat (2010), p. 71, section 3.98 |
| Gross fixed capital formation (GFCF) | Expenditure on goods and services used in production but not consumed in one reference period, net of sales of such goods or services | Eurostat (2010), pp. 73-77, sections 3.122-3.145 |
| Changes in inventories | Net increase in the value of goods in inventory and goods or services work in progress from the last ref. period, not including increases/decreases resulting from price changes of existing goods/services (if there was a decrease, this value is negative) | Eurostat (2010), pp. 77-78, sections 3.146-3.153 |
| Changes in valuables | Net increase in the value of goods with intrinsic (non-production-related) value (if there was a decrease, this value is negative) | Eurostat (2010), pp. 78-79, sections 3.154-3.157 |
| Imports[[3]](#footnote-3) | Goods or services purchased by resident units from non-resident units | Eurostat (2010), p. 79, section 3.158 |
| Exports | Goods or services purchased by non-resident units from resident units | Eurostat (2010), p. 79, section 3.159 |
| Taxes (net of subsidies) on products other than those on production (D.21-D.31) | Taxes incurred per unit of production or per £ value of production, net of subsidies on the same basis (eg, VAT) | Eurostat (2010), p. 92, sections 4.16-4.18 |

**SECTION A.2: DETAILED DISCUSSION – WHY THESE VARIABLES?**

The material in this section is summarised and extensively clarified and rendered as non-technical as possible from Eurostat (2010), pp. 55-61, sections 3.14 to 2.52; and pp. 275-276, sections 9.01 to 9.14.

Consider a firm which makes money from selling products. After paying for labour (salaries) and raw materials, and paying taxes required to produce, the rest is profit.

(1)

This can be rearranged to

(2)

In terms of National Accounts concepts that are broadly similar to the intuitive concepts in (2), this can be restated as:

(3)

Before adjusting for certain taxes and subsidies, GDP is defined as either side of this equation. This pre-adjustment version is known as gross value added (GVA). The left-hand side is known as GVA measured by production (GVAP); the right-hand side is GVA measured by income (GVAI). We have discussed a single firm, but an economy can be understood as the sum of a set of firms or producing units that can be treated the same way; for units that do not sell their output, or where the sale price does not value the output accurately, “sale price” must be (and is) imputed. The SUTs for the UK aggregate data covering all such units resident in the UK. GDP can also be defined in a third way. In economics textbooks, the usual GDP equation is:

(4)

The right-hand side of (4) is linked to (3) as follows. Using (2) in (4), and restating the RHS in National Accounts terms,

(5)

This can be restated as:

(6)

The LHS of (6) can be understood as total supply, the RHS as total use. Intuitively, everything that is used in one of the categories on the RHS must be supplied in some way; it can’t just appear out of nowhere. Either it is supplied from the domestic economy (output) or it has to be imported. Similarly, anything that is sold must be purchased by someone for some purpose, and the variables in (6) are defined such that the RHS of (6) covers all possible reasons for purchase.

Comparing equations (3) and (6) to the table in the previous subsection, it is apparent that the SUTs simply present all the terms in these equations – that is, the components of the three measures of GDP, recalling that “gross capital formation” is defined as (Eurostat 2010, p. 73, section 3.122)

(7)

It remains to justify the product and industry breakdowns.

**SECTION A.3: WHY THE BREAKDOWNS BY PRODUCTS AND INDUSTRIES?**

In this section, we show that presenting the variables described in the previous section by products and industries as they are presented in the SUTs is both conceptually valid and of interest from the perspective of economics. This will allow us to select analyses that provide useful information for economic analysis.

First, we need to define “product” and “industry.” “Product” refers to a type of good or service: for example, banking services (CPA division 64) (Eurostat 2013). “Industry” refers to the sum of all units whose dominant activity is producing the corresponding product. For example, the banking industry (SIC division 64) consists of the sum of all units with the dominant activity of producing banking services (Eurostat 2010, p. 50, section 2.150).

Product breakdowns are available for all components of total supply and total use: all GDP components except GOS and COE.

* This is *valid* because the supply and use identity expressed in (6) holds for each product separately as well as for the whole economy, with the same logic. For example, all purchases of banking services must also be sales of banking services – it wouldn’t make sense for them to be purchases of motor vehicles, or agricultural products! On the other hand, industries may supply goods or services other than their dominant activity; for example, banks might supply management consultancy (CPA division 70), so the equations don’t hold for them.
* This breakdown is *of interest* for two main reasons. Firstly, the breakdown by product gives information on what the UK is producing, what it’s importing and exporting, and what those products are used for. Secondly, it allows more detailed deflation (conversion of nominal to real values) because different products have experienced drastically different price changes over time (eg, computer equipment vs. oil).

Industry breakdowns are available for all components of GVAP and GVAI: all terms in (3).

* This is *valid* because, recalling that the intuition behind (3) is a simple model of a firm, the identity in (3) should hold for the whole economy as the sum of all firms, and also for industries as subsets of firms.
* This is *of interest* because the equation (3) reflects the *production function* that is fundamental to microeconomics. The production function is *Y = f(K, L)*, that is: value-added (GVA) is a function of capital and labour inputs. (GOS is distributed to the owners of the capital involved, just as COE is distributed to the “owners” of the labour.) In theory, each industry represents a production “technology” that we can measure by measuring all the values in (3) by industry.

1. A table of GFCF by industry is in fact published in the linked dataset. This is not, however, strictly part of the SUTs framework and we do not reference it elsewhere. [↑](#footnote-ref-1)
2. In the tables, imports of goods and services are identified separately. This is not relevant to the analysis here and we do not reference it elsewhere in the paper; “imports” can be understood as the sum of these values. [↑](#footnote-ref-2)
3. In the tables, imports of goods and services are identified separately. This is not relevant to the analysis here and we do not reference it elsewhere in the paper; “imports” can be understood as the sum of these values. [↑](#footnote-ref-3)