

## **‘Why collect the same data twice?’: Exploring the possibilities of using VAT data in the estimation of the Danish IIP**

Søren Kristensen  
Statistics Denmark

### Summary

This paper describes a project which has been launched by Statistics Denmark which aims to explore the possibilities of using administrative data (VAT) instead of survey data in the estimation of the Danish Index of Industrial Production (IIP). The use of VAT data is a recurring theme in Denmark, partly because of the political focus on response burden, but also because VAT data is already being used for other short-term indicators. The main objectives of the project are to explore the possible uses of VAT data as a means to reduce the burden on businesses, increase the efficiency of statistical production and increase the quality of the IIP.

There are, however, significant issues to be addressed before VAT data can be employed successfully to meet these objectives. Most importantly, there is a significant difference between production and turnover. In Denmark the production variable is operationalised as a function of turnover and change in inventory. As VAT data is a measure of turnover an alternative measure of inventories is needed if VAT data is to replace the current survey. Other issues relate to differences in units and definitions of variables. The paper will discuss preliminarily how and if these challenges can be met, and consider different possibilities of using VAT data.

As this is work in progress the conclusions and suggestions provided here are only tentative. Any comments, suggestions and recommendations will, of course, be most welcome.

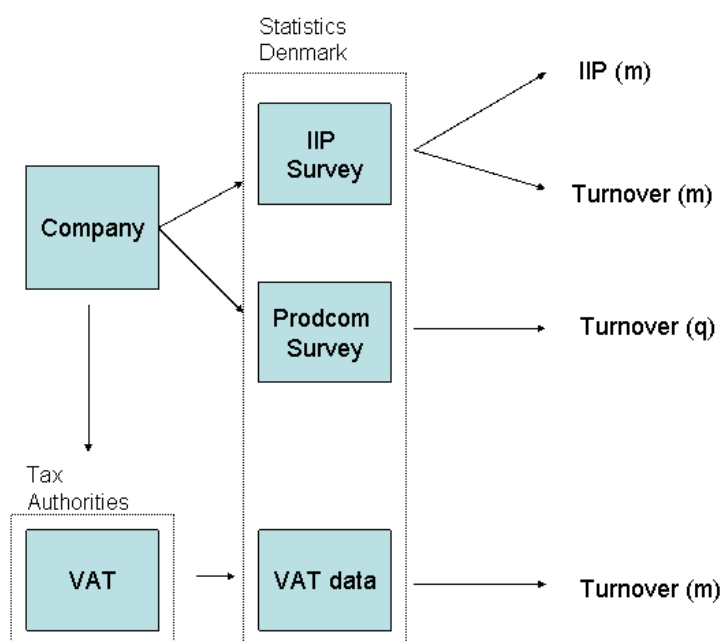
### **1. Why the focus on VAT data? – reasons for exploring its use in the IIP**

As the title of this paper indicates, our data providers – and sometimes we ourselves – occasionally wonder why it is that we collect data which we are seemingly already in possession of. Is it due to a lack of coordination? Is it a matter of not yet having realised the potential of the data? Or is it, as some data providers may believe, simply a matter of bureaucratic red tape?

The IIP has been the focus of such discussion for some years now. It has on several occasions been singled out by businesses and interest organisations as an example of a statistic which puts an unnecessary burden on businesses. One of the largest business organisations in Denmark is recently quoted as saying that whilst the IIP doesn't impose many burdens on businesses as such, it nevertheless creates a lot of irritation amongst businesses because “companies must report information that the authorities already have”.

It is not difficult to see why data providers perceive the collection of data for the IIP as imposing unnecessary burdens on businesses. Every month Statistics Denmark sends questionnaires to a sample of approximately 1200 businesses, or rather KAUs, to collect data for the IIP. We ask for information about domestic turnover, export turnover and inventories (see appendix I), and from these variables the IIP as well as turnover indexes are calculated. At the same time we collect turnover data via the Prodcom statistic, which is a quarterly statistic in Denmark, and we also publish the “sales and purchases statistic”, which is a statistic based entirely on VAT data. All the units in the IIP sample also take part in the Prodcom statistic, and the vast majority of them report VAT data monthly. So, in principle, our data providers report their turnover to us, directly or indirectly, twice each month and thrice each quarter. And we publish three different indicators of industrial turnover, as well as the IIP (see figure 1 below).

Figure 1 **Data sources for the three turnover indicators and the IIP**



From the point of view of businesses, the same data is thus submitted at least two times every month to public authorities, and it seems obvious that the response burden could be lessened by using data from only one of the sources. The focus here is not so much on using Prodcom data instead of IIP or vice-versa, but rather on using VAT data which is generated as a result of businesses settling their VAT accounts. This data is ‘free’ in the sense that no extra work is required by businesses. As most businesses in industry, or at any rate most large businesses, settle their VAT accounts monthly, data is available at the necessary frequency.

However, this is not just a matter of reducing burden for businesses; there are also other equally significant potential gains from making better use of VAT data in the estimation of IIP.

Like most other NSIs, Statistics Denmark is under pressure to produce more for less. There is a budgetary mechanism which cuts a certain percentage of our funding from the national budget every year. This is based on the premise that Statistics Denmark becomes more and more efficient, either through the

increasing use of IT technology or via other means, and that through these efficiency gains the same can be produced for less.

Survey based statistics such as the IIP are expensive. Many resources are spent on trivial operations, such as sending questionnaires out, receiving them, checking addresses, and also on data quality checks, data editing and re-contacting. For short-term statistics, and for monthly indicators, these procedures occupy a very large part of our resources. Although there are economies of scale to take into consideration, a sample of 1200 for a monthly indicator, which is a fairly normal sample size, means dealing with 14.400 questionnaires per year. Data editing and checks will obviously also be carried out on administrative data, but there should be a substantial potential efficiency gain if survey data could be replaced by administrative data.

A significant factor in this context is that VAT data is already being used by Statistics Denmark in short-term statistics. Normally it takes time and effort to transform raw administrative data into data that can be used in statistical production. There are also usually large start-up costs before a reliable system for the correct and regular transfer of data from the administrative unit collecting the data is in place. These investments have been made, and a system has been in place for some time now. So, in this sense, data is almost 'free' for Statistics Denmark as well.

Administrative data can also be used as a means to increase the quality of an indicator as it can supply additional information; either because administrative data can give access to information which for one reason or another cannot be sampled, or because administrative data can be used to increase the quality of data, for instance by using it in grossing up procedures. Today we use the quarterly Prodcom statistic for grossing up, which means there is a significant delay in the grossing up variable. Another possible way of improving the quality would be to use VAT data for imputation.

One must, of course, also take into account the potential negative quality issues which may emerge from substituting questionnaire data with administrative data. The available administrative data may not entirely fit the needs of the statistics production. There could be problems with variable definitions, the reporting units, and data delivery to name but a few of the well-known issues with administrative data. In general it is evident that the less control the NSIs have over the data collection, the less likely it is to be a good fit to needs. A very recent example of this is the decision by the Danish government to allow businesses to defer payment of VAT for some months to alleviate the credit strain imposed by the financial crisis. The ensuing delay in data has had a grave impact on some of our core indicators, and illustrates the vulnerability of statistics based on data collected for administrative purposes.

To sum up, there are at least three different objectives which can be achieved by using VAT data in the estimation of the IIP:

- Reducing the burden imposed on data providers
- Increasing the efficiency of data collection
- Increasing the quality of the IIP

A significant difference between these three objectives is that while the first two are achieved primarily by replacing the current survey data (partially or

entirely) by administrative data, the last objective may be achieved by supplementing with VAT data rather than replacing the survey data. This suggests that the benefits of using administrative data can be viewed in a two dimensional space: a quality dimension on the one hand, and an efficiency/burden dimension on the other.

There is then a compelling case for making better use of VAT data in the estimation of the IIP. As straightforward as it may seem though, there are some issues which will have to be addressed before VAT data can be employed in the production of the IIP. Most importantly, the IIP is not a turnover variable. This is a point which is often overlooked. The IIP is a function of both monthly turnover and changes in inventories (see the table below). While the former part of the equation can feasibly be estimated by using VAT data, there is currently no administrative data available on inventories. Another important issue is that VAT data refers to total turnover, whereas in the IIP survey turnover from trading goods is excluded, as required by the definition for the IIP. Finally IIP data today is collected at KAU level, also as required, whereas VAT data is reported for legal units, which can consist of more than one KAU.

*Table 1* **Comparison between VAT and IIP sources**

|                           | VAT<br>Administrative<br>data | IIP<br>Survey data         |
|---------------------------|-------------------------------|----------------------------|
| Variables                 | Turnover                      | Turnover +<br>inventories  |
| Definition of<br>turnover | all included                  | excluding trading<br>goods |
| Reporting units           | Legal Units                   | KAU's                      |

## **2. How to use VAT data in the estimation of the IIP – exploring the alternatives**

Taking departure in these general ideas, the aim of the project is to consider how to make best use of VAT data and evaluate the feasibility of the different options. Below are listed a number of possibilities which will be analysed further in the course of the project, although much of the work has yet to be done. The list is sorted according to how 'radical' the options are.

1. Replacement of sample data with VAT
2. Partial replacement of sample data with VAT data
3. Use of VAT data for imputation
4. Use of VAT for grossing up

In terms of the objectives outlined above, it is clear that it is only the first two options that will result in a (potential) reduction of the burden on business. The latter two options primarily aim at improving the quality of the indicator

by using VAT data as a supplement. On the following pages I focus primarily on the first two options.

### 1. Replacement of sample data with VAT

This option is the most radical in the sense that it entails a complete substitution of survey data with administrative data. No doubt this would be the solution most preferred by businesses, reducing the burden from something to nothing at all.

At the same time there is a valid argument that using VAT data instead of survey data could raise the quality of the indicator. By using population data instead of sample data, the uncertainties of sampling and grossing up, both important sources of error, can be eliminated.

Moreover, the quality may also be improved through better coverage. In Denmark we use a cut off sample – mainly for the sake of reducing the burden on businesses. Arguably this means that there is an inherent tendency to underestimate new trends in industry in so far as new businesses have a smaller likelihood of being included in the sample. By using VAT data coverage could be expanded without imposing additional costs on enterprises.

The main obstacle to replacing survey data with VAT data is that VAT data only covers one part of what is needed for the calculation of the IIP. Information is not available on changes in inventories from administrative sources. One option that comes to mind is to use VAT as a proxy, but that would defeat the purpose of the IIP as a separate and important indicator. It would also only make sense if the correlation between turnover and production is strong, which it is not

### 2. Partial replacement of sample data with VAT data

In lieu of a complete substitution of sample data with VAT data, another possibility is to carry out a partial substitution. That is, to use VAT data instead of collecting turnover data directly from businesses. For this solution to work it is necessary to combine data from two sources; whilst the VAT data covers the turnover part of the equation another source is necessary for inventories. A simple way to accomplish that would be to continue sending questionnaires to the sample with the questions about turnover omitted. Arguably this would lead only to a minor reduction in response burden, as businesses would still have to complete a questionnaire every month. On the other hand, the experienced burden associated with having to report the same data twice to the authorities would be lessened.

There are, however, a few quality issues that must be addressed if this solution is to work. One relates to the point described above that VAT data is not equivalent to the turnover data used for the calculation of the IIP; VAT reporting units are different from the units reporting IIP data, and whilst turnover from traded goods is omitted in the IIP, it is included in VAT. Both of these definitional differences may lead to substantial divergences between VAT data and the sample turnover data.

Another equally important point is that it is often not the same people who deal with VAT and respond to our questionnaires. Although the VAT turnover

and the IIP turnover should be close to identical for each business, (when the discrepancies described above are taken into account) there will remain some differences simply due to the fact that they emanate from two different sources. This difference might be difficult to estimate, although it can be expected to even out over time. When producing a monthly indicator, this could constitute a problem, but as it would be random error, one would expect that differences at unit level pull in both directions thus evening each other out.

To get a preliminary indication of how much these factors influence data, we looked at data covering 7 months from June 2008 to December 2008. A longer time series would obviously be desirable, but the principal reason for choosing this relatively short series is that first version unrevised data is only available from June 2008. Unrevised data is important as it gives an indication of what kind of quality may be expected at the time of publication of the monthly data. Data was merged at unit level so that VAT data was found for each unit in the sample.

The analysis showed that total turnover from VAT data over all 7 months was 17 pct. higher than the turnover reported through the IIP questionnaire. To cancel out the effect of the differences in the definition of variables we used data from the quarterly Prodcom statistics, which for each unit contains information about turnover emanating from sales of traded goods. When this is deducted from VAT turnover, the difference between VAT and IIP turnover narrows to 6 pct. seen over these 7 months.

Although this might seem to be a relatively modest difference, there is a higher degree of divergence when data is compared month by month. The range of almost 13 pct. points indicate that periodicity influences data, i.e. relating to which month turnover is registered, but also that there might be an element of random error.

*Table 1* **Comparison of turnover data from two sources: VAT data and survey based IIP data**

|       | IIP sample<br>turnover | VAT turnover          | Difference<br>VAT - IIP | VAT ex. Traded<br>goods | Difference VAT<br>(ex. traded<br>goods) – IIP |
|-------|------------------------|-----------------------|-------------------------|-------------------------|---|
|       | <i>billion kroner</i>  | <i>billion kroner</i> | <i>pct</i>              | <i>billion kroner</i>   | <i>pct</i>                                    |
| 2008  |                        |                       |                         |                         |   |
| June  | 40,0                   | 48,5                  | 21                      | 44,7                    | 12  |
| July  | 36,2                   | 42,6                  | 18                      | 38,8                    | 7   |
| Aug   | 35,7                   | 41,8                  | 17                      | 37,3                    | 5   |
| Sept. | 41,5                   | 48,9                  | 18                      | 44,4                    | 7   |
| Oct   | 41,1                   | 47,1                  | 15                      | 42,6                    | 4   |
| Nov   | 35,6                   | 39,5                  | 11                      | 35,3                    | -1  |
| Dec   | 34,1                   | 41,4                  | 22                      | 37,2                    | 9   |
| Total | 264,1                  | 304,7                 | 17                      | 280,4                   | 6   |

As for the issue about the difference in reporting units, this basically has to do with one-to-many relations between units. The majority of the businesses

represented in our sample consist of only one KAU. For these one would expect there to be very little difference between VAT data and sample turnover data. The remaining businesses with more than one KAU can be divided into two types: those whose KAUs all are within industry (here section B, C and D), and those who also have KAUs outside industry. For the first group there should be an overall match between VAT data and sample data, whereas for the latter it can be expected that VAT turnover is higher, simply because some of the units of these businesses are not part of the IIP sample. To investigate this issue we used a dataset from the Statistical Business Register to identify which of the three groups the businesses in our sample belong.

As expected, the analysis revealed that most of the remaining difference between VAT and IIP sample turnover emanates from businesses with KAUs outside industry. This group, which constitutes less than 10 pct. of the sample, accounts for more than 70 pct. of the difference between VAT and IIP turnover. It is worth noting though that for the remaining majority of the sample there is still in total a substantial difference between VAT and IIP turnover. Some of this may be due to the imprecision of the trading goods estimate, but it should also be expected that there is a difference simply due to the fact that these are two different sources. It should be pointed out that it is not a goal in itself to eliminate all differences between the two series – nor should it be assumed that the IIP turnover data is necessarily better than VAT data – but it is important to investigate why these differences exist.

Looking at unit level, further analysis of data showed that for most businesses the difference between the two turnover estimates has little impact on the total aggregate. In fact, the number of significant units which really need to be investigated further can be narrowed down to around 20 out of a sample of 1200.

While there are significant differences between VAT turnover data and the turnover data collected via the IIP survey, these admittedly very initial investigations do give reasons to be optimistic about the possibilities of finding a solution for replacing the IIP turnover data with VAT data.

However, for the calculation of the IIP it is necessary to supplement VAT turnover data with data on changes in inventories. As administrative data on inventory change is not available, the only source of this information is through survey data.

If data is collected from separate sources so that turnover comes from VAT data while data on inventories is collected via a survey, we can foresee a significant challenge in trying to consolidate data. Arguably, when both variables are being surveyed the consolidation of turnover and inventories is to a large degree carried out by the reporting unit itself. If the turnover variable is omitted from the questionnaire, one could expect that the workload associated with quality check and data editing will increase at the NSI.

One way to counteract this consolidation problem would be to adopt a different approach to the estimation of changes in inventories. Instead of collecting data at unit level, so that a match at unit level is necessary, one could estimate the change in inventories at aggregate level for each of the 45 strata.

However, change in inventory is a notoriously difficult variable to measure; there is a high degree of volatility and it is difficult to discern a pattern. We have also observed from our data, that there are large differences in the relative size of the inventories not only between NACE groups but also within them. It is not uncommon to find businesses with similar turnovers but very different sizes of inventory. We are still at the initial stages of analysing that and will have to investigate it further to answer whether it is possible or not to make aggregate estimates and use these for each strata.

But considering these two options for the inventory variable, i.e. consolidating at unit level or producing aggregate estimates, several models for the partial replacement of survey data can be defined.

- One model is to replace all survey based turnover data with VAT data and match this at unit level with surveyed data on inventories. As mentioned there would be some reduction in the burden on businesses, although following this model all the current members of the sample would still receive a questionnaire each month.
- A second model would be to produce an aggregate estimate of inventories for each of the 45 strata. With this model there is no need to match the inventory sample with the VAT data. This means there is a possibility of increasing the coverage for the turnover variable whilst decreasing the sample for the inventories variable.
- A third model is to use a much smaller sample where only the largest and most significant KAUs are included. This sample data could then be supplemented with administrative data.

Unfortunately, at this stage it is too early to conclude whether it is possible to find a feasible solution for a partial substitution of survey data with VAT data. There is still much work to be done in relation to the VAT turnover data, but the prospects for using this instead of the survey based turnover seem quite good. Measuring inventories and combining this with turnover data constitutes a much bigger challenge, which we have only just begun working on. One thing is clear though; without solving the issue of the inventories, the prospects of using VAT data as a means to reduce burden on businesses are very slim.

Appendix I About the IIP

The IIP is one of the key economic indicators (In the EU it has the status of a so-called Principal European Economic Indicator (PEEI)). Its aim is to provide a snapshot of the business cycle.

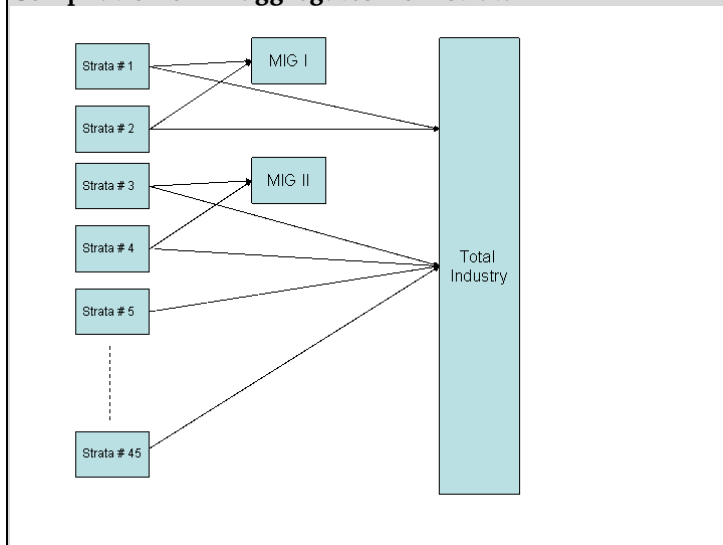
As a member of the EU, Denmark is obliged to live up to the so-called STS regulation, which amongst other things stipulates the demands for the frequency, coverage, and timeliness of the IIP. Although it is an indicator which is regulated, there is room for methodological differences.

In Denmark the index is calculated using two variables: monthly turnover and change in inventories. In this way both sales from stock and production to stock is taken into account.

The statistic is based on a sample of approx. 1200 KAUs representing the largest businesses within NACE rev. 2. sectors B, C and D. It is a cut-off sample; all businesses with 200 or more employees are included in the sample, all with less than 20 are excluded, a stratified sample is drawn within the remaining group of businesses. In terms of value added the coverage is around 85 pct. The quarterly Prodcom statistic is used for grossing up.

The sample is divided into a total of 45 strata. A production index is calculated for each of these strata, and then using the strata as building blocks national and EU aggregates are compiled (see the figure below) .

**Compilation of IIP aggregates from strata**



| Omsætningen af egne varer og tjenester samt lageropgørelse |   |  |  |  |  |
|--|---|--|--|--|--|
| Opgørelsesperiode  | Tidligere indberettet   | Evt. ændring   | Tidligere indberettet  | Evt. ændring   | Angiv periode  |
|  |   | <input type="checkbox"/> Kalendermåned<br><input type="checkbox"/> 4 uger<br><input type="checkbox"/> 5 uger | <input type="checkbox"/> Kalendermåned<br><input type="checkbox"/> 4 uger<br><input type="checkbox"/> 5 uger | <input type="checkbox"/> Kalendermåned<br><input type="checkbox"/> 4 uger<br><input type="checkbox"/> 5 uger | <input type="checkbox"/> Kalendermåned<br><input type="checkbox"/> 4 uger<br><input type="checkbox"/> 5 uger |
|  | Er der ændringer til tidligere indberettede beløb, skriv det nye beløb nedenfor |  | Er der ændringer til tidligere indberettede beløb, skriv det nye beløb nedenfor                              |  |  |
| <b>Omsætning</b><br>ekskl. handelsvarer                    | 1.000 kr.   | 1.000 kr.  | 1.000 kr.  | 1.000 kr.  | 1.000 kr.  |
| Hjemmemarked   | <input type="text"/>  | <input type="text"/>   | <input type="text"/>   | <input type="text"/>   | <input type="text"/>   |
| Eksportmarked  | <input type="text"/>  | <input type="text"/>   | <input type="text"/>   | <input type="text"/>   | <input type="text"/>   |
| <b>Lageropgørelse</b><br>af varer til videresalg           |   |  |  |  |  |
| Varer under fremstilling                                   | <input type="text"/>  | <input type="text"/>   | <input type="text"/>   | <input type="text"/>   | <input type="text"/>   |
| Fremstillede færdigvarer                                   | <input type="text"/>  | <input type="text"/>   | <input type="text"/>   | <input type="text"/>   | <input type="text"/>   |
| Sejrogsade 11<br>2100 København Ø                          |   | Hanne Lange<br>Direkte tlf. 39 17 35 53  |  | e-post khl@dst.dk<br>fax 39 17 34 19<br>www.dst.dk<br>Vend   |  |

1116000-02 A 10.08.46.2006 D56:Tryk

Export turnover

Domestic turnover

Inventories

M - 2 revised data

m-1 revised data

Current month