On the use of Value Added Tax data in creating a timely monthly Production Value Index

Annika Lindblom[1]

[1]*Statistics Sweden, Sweden, annika.lindblom@scb.se*

Abstract

The Production Value Index (PVI) is a survey that measures the change of the production of goods and services in the business sector. Turnover is the variable which serves as an indicator of production value, except for two industries where production volumes are used.

Detailed final PVI are published quarterly, approximately 45 days after the end of the reference quarter. A preliminary and more aggregated monthly PVI is published no later than 40 days after the end of the reference month.

Statistics Sweden strives to improve accuracy of the statistics, reduce response burden as well as production costs. Final PVI are, since reference month April 2015, based on turnover values for small and medium sized enterprises taken from the Swedish Tax Agency's Value Added Tax (VAT) register. However, when it comes to the preliminary monthly PVI the problem of timeliness arises. Many small and medium sized enterprises declare VAT too late to be part of the production process needed to meet the publication dates. In addition, small and medium sized enterprises often are allowed to declare VAT quarterly which means that distribution by month must be modelled.

During recent years focus on the use of administrative data has increased at Statistics Sweden. Consequently, a project started to explore the possibility of replacing direct collection with VAT-data also in the preliminary monthly indicator. In a first step, the manufacturing sector was studied since the structure of the service sector with many small enterprises makes it more complicated in this context.

Conclusions from this project led to a preliminary PVI where turnover values from small and medium sized enterprises in the manufacturing sector are taken from VAT-data. Preliminary monthly estimates, based on this new method, have been published since reference month April 2023. This paper focuses on experiences and results from the project as well as on an evaluation of the accuracy of estimates produced by the new method. Furthermore, Statistics Sweden has decided to continue with the service sector and hopefully be able to identify industries where replacing direct collection with VAT-data would be feasible. Experiences from the service sector is included in this paper.

Keywords: replacing direct collection, administrative data, timeliness

1. Introduction

Development of official statistics is to a large extent influenced by the needs of producing timely and cost-effective estimates, especially in the domains where some alternative data sources are available. This paper concerns the survey Production Value Index (PVI) where

only one variable is collected, monthly turnover and the key parameter is year-on-year changes based on index. A preliminary monthly PVI are published no later than 40 days (29 days in the retail trade) after the end of the reference month. Each quarter, final turnover estimates are produced for each month within the reference quarter[1], and they are to a large extent based on Value Added Tax (VAT) - data. However, directly collected turnover values in preliminary PVI are used for large and important KAUs in final PVI. Final monthly estimates replaces the preliminary, and revised monthly estimates are published 45 days after the end of a reference quarter. Preliminary PVI was until reference month March 2023 produced by a sample survey with direct collection from Kind of Activity Units (KAUs). In Sweden, the vast majority of KAUs coincide with the Legal Unit (LU). Furthermore, small and medium sized KAUs coincides. Conclusions from a project[2] at Statistics Sweden (SCB) showed that it would be feasible to replace direct collection with VAT-data for small and medium sized KAUs also in the preliminary monthly indicator, with acceptable level of accuracy. In a first step, this new method has been implemented in the manufacturing sector from reference month April 2023. The structure of the service sector, with many small KAUs, makes it more complicated in this context. This paper will focus on conclusions from the project as well as on an evaluation of the accuracy of estimates, mainly in terms of size of revisions between preliminary and final PVI. Small revisions between preliminary and final estimates are desirable among users. At present, published estimates by the new method are available for the reference period 2023M04-2024M02.

2. A new method for preliminary PVI

In this section, conclusions from the project mentioned above are discussed. Those conclusions were decisive when moving towards a practical implementation.

Direct collection from large and important KAUs. Turnover from large and important KAUs will continue to be directly collected in preliminary PVI (and used in final PVI). Mainly because turnover, with high accuracy, is difficult to obtain from VAT-declarations due to the complexity of those KAUs. Furthermore, turnover specifications according to activity is required from large and important KAUs and this information cannot be obtained in a VAT-declaration. And finally, due to their importance, it is important for SCB to have the possibility to contact them.

---

[1] Lindblom (2014)

[2] Elezović, Suad. Lindblom Annika (2019) and Elezović, Suad. Lindblom Annika (2020)

Micro- versus macro-approach. Two plausible general approaches were studied. The first approach was a micro-approach based on estimation of turnover at KAU level. The second approach was a macro-approach based on a methodology to estimate turnover at aggregated levels, such as different industries. Missing values for a single KAU (due to timeliness and other factors) are compensated by using models for turnover estimation at aggregated levels. In both cases, survey estimates for large and important KAUs, are combined with VAT-estimates for small and medium sized KAUs. This study showed that both approaches gave comparable results, and the recommendation was to proceed by using similar methods in preliminary and final PVI. Using similar methods is a good starting point when it comes to small revisions between preliminary and final PVI as well as the fact that much of the framework for the production process in this case already was in place.

Complete enumeration versus a sample. There are two reasonable alternatives how to proceed with the micro-approach. The first one would be to use a sample of KAUs each year and use turnover collected from VAT-data for small and medium sized KAUs. Then, the estimates are produced by using the estimation procedure already implemented for preliminary PVI. The second option is to base the estimation on the whole frame population by assigning a turnover value to each KAU included the frame population. The final estimation for small and medium sized KAUs is performed by combining VAT- data and imputations.

Each alternative has its own advantages and disadvantages. However, the second alternative has a strength when it comes to small revisions between preliminary and final estimates. Furthermore, each KAU only represents itself, which is always a strength when it comes to eliminating sampling error(-s) and impact from individual values. The recommendation was to proceed with the second alternative.

Imputation method. Different imputation methods were studied. Based on the results the decision was to use ratio imputation in the preliminary PVI. The auxiliary x variable is observed VAT-values the same month previous year. Imputation is done within non-overlapping imputation groups based on industrial activity and size in terms of annual turnover. Five different sets of imputation groups (hierarchical) are used, and the same method is used in each imputation group. The imputation groups must consist of at least ten KAUs with observed VAT-values on both occasions. Otherwise, the next imputation group in the hierarchy is used.

In those rare cases where ratio imputation is not possible imputation is done by the mean value in stratum.

Parallel run. A parallel run has been conducted for the period 2022M04-2023M03 and estimates were produced by old and new method. In addition, in retrospective estimates for the period 2021M04-2022M03 according to the new method were produced. A comparison

between level estimates and estimates of change showed that the new method produces estimates with improved or at least maintained level of accuracy. The main concern was timeliness (and to some extent declaration period) of VAT-data, see section 4. However, declaration period turned out to be a minor problem in the manufacturing sector.

3. Types of level estimates produced within the framework of PVI

As mentioned before, each quarter final PVI level estimates replaces preliminary PVI level estimates and revised monthly estimates are published. Final PVI level estimates are used as input to the index calculation. Except for the latest three months where preliminary PVI level estimates constitute the input (before final PVI replaces preliminary PVI). To overcome the methodological differences and obtain small revisions between preliminary and final PVI a more comparable level estimate for preliminary PVI is modelled. This modelled level estimates are used as input to the index calculation. Modelling is done by using the final PVI level estimate the same month previous year in combination with the corresponding estimate of change based on preliminary PVI from both occasions. This evaluation includes calculated size of revisions between level estimates for preliminary PVI as well as for modelled preliminary PVI. Table 1 shows the different types of estimates included in the evaluation.

Table 1 Different types of estimates included in the evaluation.

Type of Estimate	Method	Denoted
Level Estimate	Old method	SampPrel
Level Estimate, modelled	Old method	SampmPrel
Level Estimate	New method	VATPrel
Level Estimate, modelled	New method	VATmPrel
Final PVI		VATFinal

4. Challenges using VAT-data in the preliminary PVI

Timeliness is the major challenge when it comes to use VAT-data as observed turnover values in the preliminary monthly PVI. Other challenges, such as measurement errors, are of significantly less importance. Many small and medium sized LUs often declare VAT too late to be part of the production process needed to meet the fixed publication dates. In addition, small and medium sized LUs are allowed to declare VAT quarterly which means that distribution by month in those cases must be modelled.

A LU can declare VAT to the Swedish Tax Agency monthly, quarterly or yearly. The expected annual turnover decides how often a specific LU is obligated to declare VAT:

Table 2 General time points for declaring VAT to the Swedish Tax Agency

Large, annual turnover > 40 million SEK[1]	**Small, annual turnover between 1-40 million SEK[1]**	**Small, annual turnover between 1-40 million SEK[1]**
Refperiod month (T)	Refperiod month (T)	Refperiod quarter (Q)
T + 26	T + 40	Q + 40

1 40 million SEK is about 4 million EUR

LUs with expected annual turnover less than one million SEK declares VAT on a yearly basis. To be part of the production process, VAT-data must be available no later than T+35. As table 2 indicates, VAT-data from small and medium sized KAUs can, at that time point, be expected to be missing. Table 3, see appendix A, shows by the domains used for publication the proportion of imputed turnover, at the time for production as well as at current date (April 2024). The size of the proportion at production date versus corresponding current date clearly shows that timeliness is a challenge. Table 4, see appendix A, shows the proportion of imputed turnover by month, at the time for production as well as at current date. The proportion of imputed turnover is quite the same all months but increases slightly the latest months. The same pattern applies to fourth quarter 2021 and 2022. One reason could be that turn of the year and upcoming final accounts means an intensive period for businesses. Note that all directly collected turnover values from large and important KAU are included in table 3 and 4.

Nevertheless, it is important to bear in mind that non-response in the former sample survey with direct collection also meant a quite large proportion of imputed turnover each reference month.

5. Evaluation and results of the new method

The accuracy of estimates is evaluated mainly in terms of revisions between preliminary and final PVI (level and change estimates). At present, published estimates by the new method are available for the reference period 2023M04-2024M02. However, for evaluation purpose, the period 2022M04-2023M03 is of most interest because the parallel run provided estimates by both methods. And, as mentioned before, in retrospective estimates according to the new method were produced also for the period 2021M04-2022M03.

In this evaluation, directly collected turnover values from large and important KAUs is identical in preliminary and final PVI. In practise, this is not always the case because corrections of specific turnover values can be necessary between preliminary and final PVI.

5.1 Mean Square Error

MSE is a measure of the size of the error when comparing a model calculated value (here preliminary PVI) with the true value (here final PVI):

$$MSE = E\left[\left(\hat{\theta} - \theta\right)^2\right] = Bias^2\left(\hat{\theta}, \theta\right) + Var\left(\hat{\theta}\right)$$
$$\widehat{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{\theta}_i - \theta_i\right)^2 \quad (1)$$

where n = number of periods
$\hat{\theta}$= model calculated value (preliminary PVI)
θ= true value (final PVI)

MSE measures the average squared difference between the model calculated value and the true value. RMSE measures the average difference between the model calculated value and the true value.

$$\widehat{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{\theta}_i - \theta_i\right)^2} \quad (2)$$

5.2 RMSE - level estimates

RMSE by domain, based on level estimates from reference periods 2022M04-2023M03, between preliminary and final PVI has been calculated. This means a domain average based on twelve months, according to formula (2). RMSE has been calculated for the types of estimates listed in table 1.

Table 5, see appendix A, shows that the new method means a smaller RMSE (denoted +) in most domains. Table 5 shows that *VATPrel* has smaller RMSE compared to *SampPrel* in all domains. However, this is not a quite fair comparison because *SampPrel* and *VATFinal* is produced by different methods whereas *VATPrel* and *VATFinal* is produced by similar methods.

The “fairest” level estimate comparison would be to compare modelled estimates produced by the two methods. Table 5 shows that *VATmPrel* has larger RMSE compared to *SampmPrel*

in three domains which implies that the model calculation is unnecessary when the method for preliminary and final PVI are similar. A comparison between *SampmPrel* and *VATprel* (a fair comparison) shows that *VATprel* has smaller RMSE in all domains.

5.3 RMSE - estimates of change

Monthly year-on-year change estimates based on turnover has been calculated for the period 2021M04-2023M03 where level estimates by both methods were produced/obtained in retrospective. In addition, corresponding change estimates has been calculated for the period 2023M04-2023M12 by the new method. The calculated RMSE constitutes of a domain average according to formula (2) based on twelve months for the period 2022/2021 and nine months for the period 2023/2022 (because final PVI is not finalised for the first quarter 2024). Table 6 below shows available estimates for different periods.

Table 6 Estimates avaiable for different reference periods

Period	SampPrel	VATPrel
2021M04-2022M03	Yes	Yes
2022M04-2023M03	Yes	Yes
2023M04-2023M12	No	Yes

Table 7, see appendix A, shows that the new method also means smaller RMSE for estimates of change (except in one domain). Furthermore, a comparison between the size of RMSE for *VATPrel* 2022/2021 and *VATPrel* 2023/2022 shows the same order of magnitude. Perhaps slightly larger 2023/2022 but RMSE for 2022/2021 may be interpreted with some caution because level estimates for the period 2021M04-2022M03 were produced in retrospective.

6. The service sector

A feasibility study concerning the service sector has been completed at SCB. As mentioned before, the structure of the service sector, with many small and medium sized KAUs, makes it more complicated to replace direct collection with VAT-data. Timeliness and declaration period impacts the service sector much more compared to the manufacturing sector. In addition, the service sector includes industries with a significant amount of activity exempted from VAT. As well as industries with a significant amount of KAUs declaring VAT on a yearly basis. Nevertheless, the conclusion from the feasibility study was that at lest some of the industries have a potential for replacing direct collection with VAT-data. It was decided to conduct a

parallel run for the period 2024M04-2025M03 where estimates will be produced by old and new method. Estimates from this parallel run will provide the basis for an evaluation and future decision on implementation in the service sector.

7. Conclusions

This evaluation indicates that estimates with improved or maintained accuracy is obtained by the new method. However, it is important to continuously evaluate revisions between preliminary and final PVI to make sure that they do not increase. There are some areas of improvement to consider, like use AGI (PAYE tax return) as additional auxiliary information in the imputation method. AGI is a quite new data source at SCB which is timelier compared to VAT and declaration period is basically always month. In addition, consider omitting the modelled estimate for preliminary PVI. Results from this evaluation indicates that *VATPrel* gives smaller RMSE compared to the modelled V*ATmPrel*. This is quite reasonable since the new method for preliminary PVI is more similar to the method used in final PVI compared to the old method for preliminary PVI based on a sample survey.

References

Lindblom, Annika. (2014). Användning av omsättningsuppgifter från Momsregistret i den kvartalsvisa Konjunkturstatistiken (in Swedish).

Elezović, Suad. Lindblom Annika (2019). On the use of VAT data in creating estimates of monthly turnover. Discussion paper Advisory Scientific Board of Statistics Sweden

Elezović, Suad. Lindblom, Annika. (2020). Projekt effektivare granskning – Momsuppgifter Del 1 (in Swedish).

Appendix A – Tables and Figures

Table 3 Proportion of imputed turnover at production date and current date, by domain

	Production Date		Current date	
Domain	TurnOver Estimate[1]	ThereOf Imputed, %	TurnOver Estimate[1]	ThereOf Imputed, %
B07-09	5 864 174	4,0	5 969 450	0,6
C10-12	20 550 799	10,2	17 565 005	2,7
C13-15	1 229 358	21,5	1 098 585	2,7
C16	9 618 280	13,0	10 660 632	3,1
C17	12 321 652	1,2	11 664 551	1,6
C18	1 304 528	29,4	1 290 357	4,8
C19[2]	14 333 532	0,0	14 333 532	0,0
C20-21	24 812 374	2,3	21 136 698	1,0
C22	4 897 829	14,5	4 621 876	3,4
C23	4 699 204	6,9	4 454 151	2,1
C24	19 619 411	1,4	18 428 210	0,6
C25	15 189 239	26,5	14 525 313	4,2
C26	6 724 479	8,1	5 454 394	2,2
C27	8 300 937	5,9	6 827 138	3,6
C28	24 791 153	11,7	21 540 191	4,8
C29	32 213 411	2,4	26 077 714	1,6
C30	3 639 933	4,1	2 993 269	1,5
C31-33	8 444 541	29,8	7 575 660	5,2
Total	218 711 935	8,1	218 501 354	0,6

1 Average of turnover estimates 2023M04-2023M12, million SEK

2 This domain consists only of large and important businesses (direct collection)

Table 4 Proportion of imputed turnover at production date and current date, by month

		Production Date		Current Date	
Reference Period	Production Date	TurnOver Estimate[1]	ThereOf Imputed, %	TurnOver Estimate[1]	ThereOf Imputed, %
2023M04	20230601	213 050 277	6,1	213 710 024	0,1
2023M05	20230630	231 159 668	8,2	231 124 280	1,6
2023M06	20230806	233 784 231	7,4	233 696 760	0,6
2023M07	20230904	171 871 594	7,4	171 344 995	0,0
2023M08	20231003	196 997 604	8,6	196 271 175	0,7
2023M09	20231103	238 361 269	6,8	237 657 392	1,2
2023M10	20231203	239 322 436	8,5	239 703 325	0,0
2023M11	20240104	231 453 251	9,1	230 861 859	0,4
2023M12	20240202	212 407 080	10,7	212 142 374	0,9

1 Average of turnover estimates based on all industries, million SEK

Table 5 Size of RMSE when comparing level estimates, preliminary and final PVI for the period 2022M04-2023M03

Domain	SampPrel	VATPrel	New meth.	SampmPrel	VATmPrel	New meth.	SampmPrel	VATPrel	New meth.
B07-09	55 348	15 068	+	71 183	23 563	+	71 183	15 068	+
C10-12	235 957	152 836	+	381 082	203 922	+	381 082	152 836	+
C13-15	53 672	28 044	+	51 531	32 853	+	51 531	28 044	+
C16	219 675	112 977	+	282 489	168 698	+	282 489	112 977	+
C17	25 792	8 986	+	64 997	42 280	+	64 997	8 986	+
C18	74 839	24 395	+	75 039	53 468	+	75 039	24 395	+
C19	0	0	=	0	0	=	0	0	=
C20-21	91 540	20 958	+	100 977	97 180	+	100 977	20 958	+
C22	75 741	50 912	+	88 189	80 967	+	88 189	50 912	+
C23	69 279	21 563	+	155 175	79 667	+	155 175	21 563	+
C24	85 573	43 341	+	225 150	126 912	+	225 150	43 341	+
C25	335 561	211 189	+	360 315	170 705	+	360 315	211 189	+
C26	174 474	48 936	+	215 686	103 295	+	215 686	48 936	+
C27	186 660	74 340	+	184 213	71 367	+	184 213	74 340	+
C28	374 547	83 482	+	367 990	597 595	-	367 990	83 482	+
C29	387 940	85 408	+	299 822	323 050	-	299 822	85 408	+
C30	140 841	26 916	+	171 225	60 591	+	171 225	26 916	+
C31-33	380 430	119 269	+	220 520	907 788	-	220 520	119 269	+

Table 7 Size of RMSE when comparing estimates of change, preliminary and final PVI

	Change 2022/2021[1]		New	Change 2023/2022[2]
Domain	SampPrel	VATPrel	method	VATPrel
B07-09	0,89	0,58	+	0,59
C10-12	1,41	1,21	+	3,25
C13-15	4,86	3,59	+	1,94
C16	2,49	1,47	+	2,60
C17	0,60	0,36	+	2,64
C18	7,15	5,53	+	2,11
C19	0,00	0,00	=	0,00
C20-21	0,67	0,75	-	1,25
C22	2,52	2,38	+	2,17
C23	4,36	2,24	+	3,23
C24	1,72	0,57	+	3,19
C25	2,94	1,42	+	2,81
C26	4,28	5,33	-	1,51
C27	3,83	1,60	+	3,00
C28	3,86	3,34	+	5,41
C29	2,27	1,23	+	0,69
C30	5,89	2,56	+	1,76
C31-33	4,45	2,07	+	1,93

1 Based on estimates of twelve months

2 Based on estimates of nine months. Final PVI for first quarter 2024 is not yet published