A Ninth Application of ASPIRE for Statistics Sweden

Susan Linacre and Stephen Penneck

June 17, 2020

Table of Contents

1. Executive Summary

Three new products were reviewed in 2020: Production Value Index (PVI), Quarterly Emissions Accounts (QEA) and Building Statistics. Section 2 gives the background to ASPIRE, and sets out the changes in this ASPIRE round from previous rounds. The key findings are in sections 3, and are summarised below.

All product areas showed a high level of professionalism in relation to their products and were constructive and supportive despite the difficulties imposed by the remote working arrangements.

The three new product areas all had lower scores than the products reviewed in earlier rounds. We believe this is partly a result of changes in the process, which tightened requirements for achieving higher scores. In addition, these products have had less exposure to developing quality management strategies than the earlier products that have worked through several rounds of ASPIRE.

Rather than reflecting a concern about the actual statistics being published, the lower scores indicate the product areas in this round still have a way to go in understanding how to assess and explain accuracy to their users, and why conversations about the use and accuracy for that use should be driving design and presentation of their products as well as their improvement plans.

There were a number of cross-cutting issues and recommendations, presented in section 4:

1. Statistical leadership is required to make the connections with the users, and drive an understanding of quality and its implications.
2. In particular if quality is defined as fit for purpose, it is very important for statistical areas to have a good understanding of the key uses to which the statistics will be put.
3. Users of economic time series are mainly interested in movements and turning points, and asessments of quality should reflect this.
4. Where statistics are derived from other sources, product areas should have a good understanding of the quality of those sources and convey that to their own users.
5. Time series publications should include an analysis of revisions to help users understand the accuracy of preliminary estimates.
6. The quality issues identified by the product areas should be used to help drive improvement plans.

2. Introduction and changes to ASPIRE in Round 9

In 2011, the Ministry of Finance directed Statistics Sweden to develop a system of quality indicators for a number of key statistical products. ASPIRE was developed to meet this need, and conducted annually from 2011 to 2017.

Following a review of the process in 2018, some changes were implemented in the evaluation conducted in 2019, and these were further refined for 2020. Five of the ten previous products were rotated out in order to provide space for new products. A two year period between evaluation rounds was created in order for the products to have enough time to work on recommendations. The previous period of one year was judged to be a bit too short.

While the general ASPIRE process was much as it had been for the earlier products, the review for ASPIRE round 9 in 2020 differed in the following ways: different products, new to this process of quality management were included; more specific checklists were used in scoring; remote assessment was used due to the Corona pandemic.

Changes to the checklists
In last year's ASPIRE round 8 report, a number of fundamental changes that had been made to ASPIRE in 2018 were listed and described. An evaluation of round 8 with the new external review team showed that it was desirable to make the rating process more comprehensible, transparent and reliable both for the experts and for the product teams. In order to achieve this it was necessary to make clarifications in the so-called checklists that support the rating process. Clarifications were made in collaboration between Statistics Sweden and the review team. Briefly, this involved:

- renaming three of the six criteria to better reflect the cyclical process of improving quality in relation to quality requirements such that the statistics be fit for purpose,
- breaking down the requirements in the previous checklists into the same number of levels as there are ratings i.e. 10 levels,
- specifying in more detail what is required for each of the levels.

Also an accompanying document was compiled called 'Description of the checklists'.

The above changes seemed effective in improving the rating process as needed.

Changes to the evaluation process due to the Corona pandemic
Another change in the ASPIRE-process was that adaptions needed to be made due to the situation with the Corona pandemic. This involved an alternative solution to the normal set-up with meetings between the external review team and the product teams on site at Statistics Sweden.

The product teams did their necessary preparatory work with the checklists and quality documentation as usual and according to plan in January through to March at which time the materials were sent to the review team. However, due to travelling restraints for the experts and the conditions of working at home for Statistics Sweden's staff, an adaption of the process was needed to conduct the actual evaluation. This was done based on two rounds of written questions and answers between the review team and each of the product teams, facilitated by a quality coordinator at Statistics Sweden. A final online meeting was held with each product team together with the review team in order to resolve any outstanding questions and to share the preliminary evaluation results.

The adaption worked well even though the evaluation process stretched over four weeks compared to one week. Even though it was less intense than the normal on-site evaluations during one week, there were a lot of stops and starts in the different exchanges of information for all involved.

Changes in the external review team
The external review team was made up of two experts – Susan Linacre and Stephen Penneck who joined the team in 2018. The team was thereby reduced by one expert compared to round 8.

The review team would like to express their gratitude to staff in Statistics Sweden who have participated in ASPIRE round 9 at this difficult time. We were conscious of the additional challenges that the Corona pandemic would make on their time, and also the different operating circumstances. We are also grateful for the quick responses we got to our questions and for the open way that staff responded.

We think we underestimated the challenges of conducting this review virtually, over an elongated period of time rather than in the usual one week visit. This coupled with the fact that these are new product areas for the review, and the review team is itself relatively new, have added to the challenges. We hope we have understood the quality issues facing these product areas, but it must be recognised that this review has had more limitations than is usual.

3. Product Reviews

3.1 General Observations

There is a natural tendency to compare the overall scores across the products or to rank the products by their total score. However, the ASPIRE model was not developed to facilitate such inter-product comparisons and there are some risks associated with ranking products in this manner. For one, the average score for the component, Sources of uncertainty, for a product reflects a weighting of each single source of uncertainty by their importance to Overall accuracy, which can vary considerably across products. Products with many highly important sources of uncertainty may be at somewhat of a disadvantage in such comparisons because they must perform well in many important areas in order to achieve a high score.

Furthermore, the assessment of low, medium, or high importance to overall accuracy is done within a product, not across products. Thus, it is possible that a highly important source of uncertainty for one product could be of less importance to Statistics Sweden than a medium important source of uncertainty for another product if the latter product carries greater importance to Statistics Sweden or for official statistics. If resources devoted to accuracy improvements are greater for one product than another, this could also explain why some products are able to show greater improvements than others. Further, although we have attempted to achieve consistency in ratings among products, some inconsistencies surely remain.

Finally, the scores assigned to a particular source of uncertainty for a product have an unknown level of uncertainty due to some element of subjectivity in the assignment of ratings as well as other imperfections in the rating process. A difference of 2 or 3 points in the overall product scores may not be meaningful because a reassessment of the product by different reviewers could reasonably produce an overall score that differs from the assigned score by that margin. Thus, any ranking of products would need to acknowledge these inevitable and unknown uncertainties in the ratings.

Normally, a more appropriate use of the product scores is to compare scores for the same product across review rounds as a way of assessing progress toward improvements. However, as this is the first time these products have been rated this is not yet possible. The most important use of the round 9 ratings is to provide a benchmark against which future ratings can be assessed.

The scorings for all three of the products reviewed this year are markedly lower than those reviewed in previous years, and also, in some cases somewhat lower than product areas own self-assessment.

There appear to be three main reasons for these lower scores. Firstly, the changes to the checklists, which were aimed at making them clearer and more specific. This has had the effect of tightening up requirements to achieve a higher score and has thereby made higher scores a little tougher to achieve.

Secondly, these products are all new to the review process, which means that the product areas have not been as exposed to the quality management concepts embedded in the assessment as the earlier products, which have been through several ASPIRE rounds. The process itself helps product areas to identify aspects of quality concern, and also helps them provide more effective communication on quality directly to key users and also through their Quality Declarations more generally.

Thirdly, the lower scores may reflect the nature of the products reviewed. Products in the earlier rounds are significant, well-established and well-resourced areas, whereas the products reviewed in the ASPIRE 9 Round tend to have a smaller resource base for improvement, and newer staff.

Thus, while the product scores are lower than in previous rounds, the review team does not see this as an issue of concern in terms of the actual statistics being published, or the subject matter proficiency of the product areas, which is generally good. Rather it is an indication that the product areas in this round still have a way to go in understanding how to assess and explain accuracy to their users, and that conversations about the use and accuracy for that use should be driving design and presentation of their products as well as their improvement plans.

It should be noted that some of the cross cutting recommendations, specifically those relating to engaging with users, measuring movement and understanding the sources are similar to last year (although in different contexts). We wondered why this was so. It may be that this reflects the particular interest of the reviewers, and other reviewers might have alighted on different concerns, but Statistics Sweden might want to consider whether some cross-agency initiatives in these areas might be worthwhile.

Table 1 shows the summary scores for the three products that were reviewed in round 9.

Table 1. Summary of Average Scores by product

Product	Overall accuracy (average scores)	Sources of uncertainty (weighted average scores)
PVI	35	35
QEA	42	42
Building	28	33

Table 2 shows the average scores per product for each component of Accuracy. The importance of the single sources of uncertainty to Overall accuracy - high, medium, low or not applicable – is indicated by the shaded cells. The average scores for each Accuracy component across the three products are shown in the second last column together with the weighted average scores in the last column. The weights of 3, 2, 1, and 0 correspond to the categorisation of high, medium, low or not applicable regarding the importance to Overall accuracy.

Table 2. Average scores by Accuracy component and product

Sub and sub-subcomponents of Accuracy for statistics	PVI	QEA	Building	Average score	Weighted average score
Overall Accuracy	35	42	28	35	N/A
Sources of uncertainty:	35	42	33	37	N/A
-Sampling	32	40	N/A	36	35
-Frame coverage	28	42	27	32	32
-Measurement	40	42	40	41	40
-Non-response	42	42	35	40	40
-Data processing	43	48	40	44	44
-Model assumptions	32	42	20	31	35
Preliminary statistics compared to final statistics	35	38	27	33	N/A

Importance to Overall accuracy			
N/A	Low (L)	Medium (M)	High (H)
Weights			
0	1	2	3

3.2 Product Ratings and Recommendations

3.2.1. Production Value Index, PVI

Context and Discussion

The PVI is a monthly index showing the economic progress of the business sector, broken down by industrial sector. The quarterly index is inventory adjusted, and is an input into GDP and the national accounts. There are three monthly survey sources: the Business Cycle Statistics for Industry survey (for manufacturing and mining industries) and the Turnover Statistics survey (for other industries). Survey data is replaced by tax data for smaller businesses as that becomes available, so the survey data as especially relevant for the early estimates and for the estimates for larger businesses. In addition, information from the Industrial Inventories Survey is used to adjust for inventory changes. Such adjustment can be significant in measuring movement during times of rapid economic change.

Overall accuracy is about fitness for purpose, and the main purpose of the PVI is an input to the national accounts. There is good dialogue between the PVI product area and the national accounts, whose interest is in quarterly growth rates. Other users are the Riksbank and the National Institute of Economic Research, which are again interested in growth and the identification and assessment of turning points in the economy. The product area told us that they have no indication of how these institutions use the statistics or what their quality requirements are, but users seem to want year on year movements. Our concern was that this approach, while providing a crude seasonal adjustment and trending mechansim, is suboptimal in identifying movement at the current end of the series compared to the application of appropriate seasonal adjustment and trending (Linacre and Zarb 1991, Statistics Sweden 2013).

Good dialogue between statistician producers and users will help to ensure that users have the best advice on how statistics should be used, and help statisticians to have a good appreciation of how statistics are used, so they can ensure that statistical systems are designed right through to be optimal for those uses, and also that quality measure appropriate for those uses are derived.

The PVI is a derived or secondary set of statistics, using survey data as its source. As such the quality of the PVI is heavily dependent on the quality of the source data. We found that the product area had a generally good understanding of the strengths and weaknesses of the sources it used.

The PVI is highly dependent on model assumptions. Other sources of uncertainty in the estimates, sampling, frame coverage and measurement, make only a moderate contribution to overall accuracy, with a low contribution from non-response and data processing.

Sampling has an impact on the early months of the estimates, before the survey is replaced with tax data for smaller businesses. These early months are important in economic assessments as it is when data are being scrutinised for turning points in the economy.

The sample is picked from the new register at the beginning of the year and then is maintained across the year. A better practice would be to pick the sample quarterly or even monthly if a frame were available, but given the frame is only available annually it appears to be appropriately maintained. From a sampling perspective, good practice seems to be applied, but sampling errors (CVs) are not calculated on movements, and there is no plan to develop this work.

The annual selection of the frame means that businesses born during the year cannot be represented in the sample, nor can businesses that grow above the cutoff. This affects the ability to measure rapid changes in growth rates, through the year. This might be an issue, for example as Sweden moves out of the Corona pandemic, if new businesses are a driver of growth, or if smaller businesses grow proportionately more quickly than larger ones.

The frame has a cutoff to exclude the very smallest businesses. This is standard practice, from a pragmatic perspective, but it is important to evaluate the level of the cutoff in terms of contribution to movement, particularly at times when the level of economic activity is changing, ideally across a business cycle.

Some survey data are replaced by tax data after a month or so and there is scope for analysis of the differences to shed light on measurement error in both sources. There has not been any recent work on this, although regular comparisons are done as tax data is received, and the product area believes differences to be small.

Response rates are good. There is a progressive reminder strategy, but no legal enforcement, in line with office policy. The early stages of enforcement can lead to response gains, though we understand the burden of proof in Sweden is high. There is a danger that if legal powers are not used they may disappear.

Revisions analysis is one of the most common tools used by users to assess the quality of preliminary statistics. We noted that the product area evaluates revisions as part of its quality control procedures but there did not seem to be a regular publication of revisions time series, together with assessments of direction of change and reasons for revisions to enable users to make quality assessments.

The improvement plan is driven by, understandably, availability of resources, and also by the requirements of national accounts and Eurostat. Some consideration could be given to assessing how far proposed developments contribute to the most significant sources of uncertainty, and whether any projects which address specific important sources of uncertainty should be prioritised.

Key recommendations for the coming two years

1. Develop a more active dialogue with the main non-national accounts users to ensure they are making the best use of the statistics and that the product area fully understand the uses to which the statistics are being put.
2. Given that the main purpose of the statistics is to assess how the economy is performing, measures relevant to assessing the accuracy of movement in the series should be developed. These should cover uncertainty from all sources, but particularly sampling and the frame. Speed in the detection of turning points and points of inflection through the preliminary data, compared to the final placement of those time points through the PVI and other economic indicator data, should be evaluated.
3. Undertake some quantitative, analytic work to assess the difference between survey and tax data to see whether the differences are systematic.
4. Undertake times series analyses of revisions and publish these regularly.

Other areas for consideration

1. Statistics Sweden could review its policy of not enforcing non-response, bearing in mind that the early stages of enforcement can lead to response gains, and legal powers that are never used can be taken away.
2. Ensure that as a more consistent approach to the modelling of some problematic industries such as construction is developed, any relevant conceptual issues are fully addressed.
3. Some consideration could be given to assessing how far proposed developments contribute to the most significant sources of uncertainty, and whether any projects which address specific important sources of uncertainty should be prioritised.

Figure 1. PVI Ratings, Round 9

Sub and sub-subcomponents of Accuracy	Average Score Previous Round	Average Score Current Round	C1. Available Expertise	C2. Compliance with standards & best practices	C3. Knowledge of requirements, achievements, and improvement needs	C4. Plans for improvement activities	C5. Results of improvement activities and findings from other evaluations	C6. Communication with users and data suppliers	Importance to Overall accuracy (single sources of uncertainty)
Overall Accuracy	-	35	○	◓	◓	◓	◓	◓	
Sources of uncertainty:	-	35							
-Sampling		32	○	◓	◓	○	●	●	M
-Frame coverage	-	28	○	◓	◓	◓	●	●	M
-Measurement	-	40	◒	○	◓	○	●	●	M
-Non-response	-	42	◒	○	○	◓	●	◓	L
-Data processing	-	43	○	◓	◓	○	○	●	L
-Model assumptions	-	32	◓	◓	◓	○	●	●	H
Preliminary statistics compared with final statistics	-	35	○	◓	◓	○	●	●	

Ratings					Importance to Overall accuracy			
●	◓	○	◒	◎	Not applicable (N/A)	Low (L)	Medium (M)	High (H)
Weak	Fair	Good	Very good	Excellent	Weights			
1, 2	3, 4	5, 6	7, 8	9, 10	0	1	2	3

3.2.2. Quarterly Emissions Accounts, QEA

Context and Discussion

The main purpose of the Quarterly Emissions Accounts is to provide a timely early estimate of components of the annual environmental accounts. They also provide an indication of the seasonal pattern of emissions, and are an important satellite to the quarterly national accounts. There is a wide range of users, including the government, international agencies, analysts, journalists and the public. The quarterly estimates are still quite new – first published in 2015 – and there is as yet little understanding of how the quarter-by-quarter estimates are used, what quality requirements these uses have, and how far the quarterly estimates meet these.

The Quarterly Emissions Accounts are a derived or secondary set of statistics. They are complex, dependent on a large number of sources and models. Modelling is a key aspect of their production, and rightly considered the largest source of uncertainty. In most cases, the models multiply activity data by an emissions factor. The factors are updated annually to reflect technical progress. Some factors are agreed internationally, some are more specific to Sweden and are derived in conjunction with other government agencies. It is not clear how sensitive the air emissions statistics are to changes in these factors; this could be established using sensitivity analysis.

Given the wide range of sources of uncertainty, it is important that experts in the quarterly statistics take leadership by distilling the most significant aspects of accuracy and explaining them to their users. This can be done, for example by considering contributions to total and to movements at a level/ sector of interest, and presenting meaningful measures or descriptions of quality for the most significant contributions. While there is good dialogue with users of the quarterly data, this does not seem to extend to dialogue about the accuracy of the data and the highest impact areas of accuracy concerns that might affect data use, or point towards the highest priority elements of an improvement plan.

While the Quality Declaration gives a good description of the main modelling assumptions used in compiling the statistics, including which are thought to be the weakest and where improvements are being sought, these are not clarified or well communicated in terms of their impact on the quarterly data. A possibility raised in discussions is a table showing the most important sources of emission by industry, and the significant accuracy issues for those sources. This might highlight accuracy concerns in areas of particular policy interest for the quarterly data, for example areas where policy was being directed at achieving changing levels of emissions. Measures of the accuracy of modelled data may be available from work done for the annual estimates. In this case, their likely carry through into the quarterly data, at least in terms of the more significant sources, impacting key uses, should be described in the Quality Declaration.

A useful summary quality measure for the quarterly data that is shown in the Quality Declaration is the first column of the table in Appendix 1 that shows the share of periodic data, averaged for a year by industry.

Revisions analysis is one of the most common tools used by users and producers to assess the quality of preliminary statistics. While the group undertakes revisions each quarter and analyses these individual revisions as a quality control process in the quarterly release, a quantitative analysis over time of the revisions to preliminary statistics is not undertaken. Analysis of revisions over time can give an indication of the accuracy of the preliminary estimates, and support effective discussions with users about the impact of accuracy on the use of the preliminary data, and it can help prioritise activities in an improvement program.

The Working Plan is driven by national statistics needs and those of Eurostat, taking account of the availability of resources. The Working Plan is for Environmental Accounts as a whole, and only a small part relates to Quarterly Emissions Accounts (section 1.2). It was not clear whether some of the improvements elsewhere on the plan would also impact on the quality of the Quarterly Emissions Accounts. Nor is it clear how far the improvement plan is driven by concerns about quality, and on which sources of uncertainty they would impact.

Key recommendations for the coming two years

1. Develop a more active dialogue with users to ensure they are making the best use of the statistics and that the product area fully understand uses to which the statistics are being put.
2. Develop more meaningful measures and descriptions of accuracy that support key users in understanding the implications of accuracy, and areas of accuracy risk, most relevant to their particular uses of the quarterly data. These measures should be published where possible, and would contribute to the prioritisation of an improvement plan for the quarterly data.
3. Undertake some sensitivity analysis to establish which emissions factors have the largest impact on the statistics and so where efforts to improve these factors would be most productive. Such sensitivity analysis may be available through work done for the annual series, and if applicable to the quarterly data should be included in the Quality Declaration for the quarterly data.
4. Consider which aspects of the work done for annual statistics, to assess uncertainty due to modelling, are most significant in terms of uses of the quarterly estimates. Refer to these in some way in the Quality Declaration.
5. Publish revision analyses that show the difference between the quarterly data as first published and then subsequently revised over time. The published revisions analysis should indicate any trends in revisions over time, as well as the contribution to revisions of components, such as new data, methodological changes and changes in model assumptions.
6. Clarify in the Working Plan how far it is driven by quality concerns, and making it more explicit how projects impact on sources of uncertainty.

Other areas for consideration

1. Produce a table showing the most important sources of emission by industry, and the significant accuracy issues for those sources. This might particularly highlight accuracy concerns in areas of particular policy interest for the quarterly data, for example areas where policy was being directed at achieving changing levels of emissions.

Figure 2. QEA Ratings, Round 9

Sub and sub-subcomponents of Accuracy	Average Score Previous Round	Average Score Current Round	C1. Available Expertise	C2. Compliance with standards & best practices	C3. Knowledge of requirements, achievements, and improvement needs	C4. Plans for improvement activities	C5. Results of improvement activities and findings from other evaluations	C6. Communication with users and data suppliers	Importance to Overall accuracy (single sources of uncertainty)
Overall Accuracy	-	42	○	○	◔	○	◔	◔	
Sources of uncertainty:	-	42							
-Sampling	-	40	○	◕	◔	●	●	○	L
-Frame coverage	-	42	○	○	◔	◔	●	◔	M
-Measurement	-	42	◕	◕	◔	●	●	○	L
-Non-response	-	42	◕	◕	◔	●	●	○	L
-Data processing	-	48	◕	◕	○	○	◔	●	L
-Model assumptions	-	42	○	◔	◔	○	●	○	H
Preliminary statistics compared with final statistics	-	38	◔	◔	◔	○	◔	◔	

Ratings					Importance to Overall accuracy			
●	◔	○	◕	◎	Not applicable (N/A)	Low (L)	Medium (M)	High (H)
Weak	Fair	Good	Very good	Excellent	Weights			
1, 2	3, 4	5, 6	7, 8	9, 10	0	1	2	3

3.2.3. Building Statistics

Context and Discussion

The Building Statistics series evaluated in ASPIRE Round 9 includes the three collections: Permits, Commencements and Completions. The Building Price Index was taken out of scope given the changed working arangements due to the Corona pandemic.

The Swedish National Board of Housing is a key user of all three collections, although their main focus is on the commencements and completions. The National Board of Housing also disseminates statistics to other users including the Riksbank. The Permits data is used by the National Board and the Riksbank as an economic indicator in their forecasting, but the most important user of the Permits data is considered to be Eurostat. While Statistics Sweden only publishes the quarterly data, and in a non seasonally adjusted form, it provides the monthly permits data to the National Board of Housing which also provides it to the Riksbank. Statistics Sweden also provides seaonally adjusted monthly permit data to Eurostat where it is published as an economic indicator series, with one months delay, in a table with data from other countries.

While Statistics Sweden have a good relationship with the National Board of Housing and Eurostat, there seems to be less communication with other users. Given that monthly permits provide a valued early indicator of building activity, and economic activity more generally, consideration should be given to publishing the seasonally adjusted and trended monthly data for broader use.

The most significant quality issue for the building statistics is the lag in reporting by municipalities which can be of the order of 30 to 40% for the quarterly number of building permits, and is still around 10% after 4 months. The lag in permits affects the commencements and completions data as well since the permits data provides the frame for these collections. An adjustment is made for the lag in reported statistics, with the adjustment being based on an average number of missing permits for the same period over the previous three years. The adjustment made is reviewed quarterly, but the accuracy of the adjustment compared with the final permits data when it is available, is not regularly evaluated. For example there is no evaluation of whether three years of smoothing continues to be optimal, or whether the lag in permits is related to the level of building activity.

Ideally the lags in the permit data would be reduced to lower levels. There are plans to move to more machine to machine interfaces in data collection, but there are no funds allocated for such work. There does not appear to have been any recent evaluation of the source of the lag; whether it is predominantly municipalities failing to report by the due date, or whether the reporting does not include all permits. It appears to be a mixture of the two. Furthermore, it is not clear if the same municipalities have large lags each month. A problem in identifying patterns in relation to this, is that it is not known until later whether a failure to submit a return is because no permits have been issued that month, or is an actual non response.

Given the significance of the lag, it would be desirable to undertake analysis during collection of municipalities that appear to have greater problem with reporting, and work with these to find mechanisms to improve response. One strategy may be to ascertain from non-responders at the deadline for response, whether they are making a nil response, or a delayed positive response, so that adjustments in the estimation procedure can be made.

It appears there has been quite a bit of work undertaken to improve reporting from municipalities with reviewed and revised instruction documents and plans for improved excel spreadsheets. In addition the data collected from municipalities is confronted at the microdata level with data from the ‘Prices for newly produced dwellings’ so that data can be revised as appropriate.

In addition to municipalities failing to fully report permit data, commencement data is also subject to some measurement error in the preliminary data as the municipality reports the date the developer is approved to start building rather than the actual commencement date. This is subsequently corrected by editing the data against information from the developer collected in the ‘Prices for newly produced dwellings’. Other corrections to the initially reported data are also picked up through this micro level data comparison. This comparison is facilitated by the integration of data from this survey in the same production system. As well,

the municipalities have been aided in their reporting through reviewed and revised instruction documents and there are plans in place for improved excel spreadsheets.

Concerns with lags and measurement error are largely with the preliminary data for all three series. An analysis of the difference between preliminary and final data over time would provide insight into the accuracy of the preliminary data at the time it is published and the implications of this for uses of the data.

There is some annual analysis done of the impact of lag adjustment on the data, but information on this is not provided in the Quality Declaration. There does not appear to be any quantitative analysis of revisions over time, nor does the size of revisions appear to be discussed as an issue with users.

Key recommendations for the coming two years

1. Discuss with users their specific uses of the three series. In particular discuss the impact of lags in the permit data, and to a lesser extent in commencements and completions, on these uses, and any implications for lag adjustment. Also discuss the applicability of seasonal adjustment and trend estimation for these uses.
2. Consider if monthly permits data, currently being made available to some key users as an economic indicator, should be published by Statistics Sweden for widespread use as an economic indicator.
3. Analyse the accuracy of the estimated lag at time of first publication of quarterly data, over time, for example by undertaking a time series analysis of the difference between each quarter's lag as first estimated and the actual number of additional permits that flowed in late for that quarter.
4. Analyse the underlying components of the lag by maintaining monthly measures of non-response by municipalities, and for each municipality, maintaining some information on response history, and if possible, some measure of completeness of response.
5. Evaluate the degree of seasonality in the time series for permits, commencements and completions and discuss with users, the benefits of seasonally adjusting and trending the series to support their use as economic indicators.
6. Undertake and publish results from, a quantitative analysis of revisions and how they vary over time, to better understand the impact of lags and initial reporting error on use of the data, as well as allow prioritisation of plans for improvement.

Other areas for consideration

1. Consider querying with each non-responding municipality at the deadline for response, whether they are making a nil response, or a delayed positive response. This information might be used in developing a response history for municipalities, to facilitate targetted interventions.

Figure 3. Building Statistics Ratings, Round 9

Sub and sub-subcomponents of Accuracy	Average Score Previous Round	Average Score Current Round	C1. Available Expertise	C2. Compliance with standards & best practices	C3. Knowledge of requirements, achievements, and improvement needs	C4. Plans for improvement activities	C5. Results of improvement activities and findings from other evaluations	C6. Communication with users and data suppliers	Importance to Overall accuracy (single sources of uncertainty)
Overall Accuracy	-	28	◓	◓	◓	◓	●	●	
Sources of uncertainty:	-	33							
-Sampling		N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
-Frame coverage	-	27	◓	◓	◓	●	●	●	H
-Measurement	-	40	◓	○	○	◓	◓	◓	H
-Non-response	-	35	○	◓	◓	◓	●	◓	L
-Data processing	-	40	○	○	◓	◓	●	◓	L
-Model assumptions	-	20	◓	◓	●	●	●	●	L
Preliminary statistics compared with final statistics	-	27	◓	◓	◓	●	●	◓	

Ratings					Importance to Overall accuracy			
●	◓	○	◒	◎	Not applicable (N/A)	Low (L)	Medium (M)	High (H)
Weak	Fair	Good	Very good	Excellent	Weights			
1, 2	3, 4	5, 6	7, 8	9, 10	0	1	2	3

4. Cross-Cutting Issues and Recommendations

4.1 Statistical Leadership

Statisticians in a National Statistical Office are professionals, with expertise in data, its collection, and analysis, and also in understanding its quality attributes and the implications of these for the key uses of the data. They have a professional set of values associated with ensuring the integrity of statistics and have a duty of care to their users. To be effective they must show statistical leadership in their areas of expertise. This leadership extends outside the organisation to the users, with whom there must be strong engagement, and also inside the organisation to areas of specialist expertise who must be partners in assuring that best practice approaches are being used across the agency, and to areas who produce any source data that they use.

4.2 Understanding the uses and the quality requirements for these uses

Overall accuracy is about fitness for purpose. Good dialogue between statistician producers and users will help to ensure that users have the best advice on how statistics should be used, and help statisticians to have a good appreciation of how statistics are used, so they can ensure that statistical systems are designed right through to be optimal for those uses, and also that quality measure appropriate for those uses are derived. This good understanding will usually be gained through bilateral discussions rather than through wide user council meetings.

4.3 Measuring movements

For many sets of economic statistics, users are predominantly interested in movement rather than level. This interest is particularly strong during periods of rapid change. Given this, it is important that all aspects of the design of economic series be considered in terms of performance in measuring movement, particularly around turning points. This includes the frame and any cutoff used, the frequency of frame updating, the sample design and measures of accuracy, modelling such as for a lag adjustment, and analysis and presentation, such as whether to seasonally adjust and trend estimate. Descriptions of quality should also relate to this key use of estimating movement at the current end of the series. Currently there are a number of practices across the economic series that are not in line with this focus on movements even where it is understood that such movement is of key interest.

4.4 Understanding the sources

Product areas need to have a generally good understanding of the strengths and weaknesses of the data sources they use. Although these will be fully described in the source data Quality Declarations they should be also summarised in the Qualtiy Declarations of the secondary statistics, with their impacts evaluated for users.

4.5 Analysis of revisions

Revisions analysis is one of the most common tools used by users to assess the quality of preliminary statistics. Statisticians often evaluate revisions as part of their quality control procedures but there is often less focus on regularly publishing revisions time series, together with assessments of direction of change and reasons for revisions to enable users to make quality assessments.

4.6 Let quality concerns drive improvement plans

The improvement plan is driven by, understandably, availability of resources, and also by the requirements of users. Often users are most interested in more analysis, or greater detail. An active dialogue with them on quality issues should enable proposed developments that contribute to the most significant sources of uncertainty, to be included in the improvement plan and prioritised.

5. References

Laiho-Kauranne, Johanna, Susan Linacre, and Stephen Penneck. 2019. *An Eighth Application of the ASPIRE Quality Evaluation System for Statistics Sweden*. Unpublished report.

Linacre, Susan, and John Zarb. 1991. "1991 Feature Article - Picking Turning Points in the Economy". *Australian Economic Indicators*, (April 1991 issue), xi-xvi. Accessed June 16, 2020. www.abs.gov.au/AUSSTATS/abs@.nsf/94713ad445ff1425ca25682000192af2/bc7ec6b46d35dcabca256fe9007bfe27!OpenDocument)

Statistics Sweden. 2013. *Consistent Seasonal Adjustment and Trend-cycle Estimation.* Statistics Sweden. Accessed June 16, 2020. https://www.scb.se/publikation/21099