Improving the quality of statistics of manufacturing turnover growth: timeliness and accuracy

Discussion paper 05002

R.H.M. van der Stegen

Statistics Netherlands

Voorburg/Heerlen, March 2005

Explanation of symbols

.	= data not available
*	= provisional figure
x	= publication prohibited (confidential figure)
–	= nil or less than half of unit concerned
–	= (between two figures) inclusive
0 (0,0)	= less than half of unit concerned
blank	= not applicable
2004–2005	= 2004 to 2005 inclusive
2004/2005	= average of 2004 up to and including 2005
2004/'05	= crop year, financial year, school year etc. beginning in 2004 and ending in 2005

Due to rounding, some totals may not correspond with the sum of the separate figures.

Publisher
Statistics Netherlands
Prinses Beatrixlaan 428
2273 XZ Voorburg
The Netherlands

Printed by
Statistics Netherlands - Facility Services

Cover design
WAT ontwerpers, Utrecht

Information
E-mail: infoservice@cbs.nl

Where to order
E-mail: verkoop@cbs.nl

Internet
http://www.cbs.nl

ISSN: 1572-0314
Key figure: X-10
Production code: 6008305002

Statistics Netherlands

IMPROVING THE QUALITY OF STATISTICS OF MANUFACTURING TURNOVER GROWTH: TIMELINESS AND ACCURACY

Summary:

The objective of the research described in this paper is to improve the quality of the statistics of the turnover growth of the manufacturing industry. The quality parameters studied are timeliness and accuracy. Without adaptations of the compilation process of the statistics, improvement of one parameter will usually lead to a deterioration of the other. Methods are developed which improve one parameter without a deterioration of the other.

The quality of the statistics is determined in a simulation over 60 months. For each month, the (time dependent) difference is calculated between the final value of the growth of the turnover and earlier values as a function of the number of days after the reporting month. These monthly error functions are combined in a root mean square error, and a maximum overestimation and underestimation as a function of the number of days after the reporting month.

The turnover growth is studied for the aggregate manufacturing industry and a division into six branches. Currently the data for the first publication of the enterprises are downloaded 30 days after the end of the reporting month.

Three methods are explored for improving the quality:

1. *Alternative imputation methods for non-response within the existing approach. Using larger aggregates and a different reference month for imputation result in a better quality. For example, timeliness can be improved with one week without sacrificing accuracy.*
2. *The selective response chasing method assumes that the largest enterprises are chased and respond within a certain number of days. The number of the largest enterprises is limited and they account for a large percentage of the total turnover. The response of the other enterprises is used as it is currently received. Most criteria are as accurate as on day 30 without response chasing using a complete response of the largest enterprises on day 14, whereas response chasing until day 21 results in more accurate and timelier results.*
3. *The expected value is a prediction of the turnover for the next month. This prediction is obtained with a time series model. The expected value is the starting value for unknown response and is replaced by response as soon as it is received. The use of an expected value improves the accuracy of the estimation of the turnover growth during the first 3 weeks after the reference month. However it is not accurate enough for a timelier publication. The main problem is the irregular behaviour of the turnover which cannot be captured by the time series model.*

Keywords: manufacturing industry, timeliness, accuracy, imputation

1. Introduction

1.1 Incentive

Timeliness and accuracy are important quality parameters of statistical information. These two parameters cannot be analyzed independently. Without a redesign of the process, improved timeliness will usually result in less accurate results and vice versa. As users of statistics are asking for timelier statistics, a redesign of the compilation process is needed, because improved timeliness should not result in less accurate results.

The quality of several important short-term statistics can be improved. To start with, this paper focuses on the timeliness of the turnover growth of the manufacturing industry. It was chosen because it is one of the primary sources for the industrial production index and quarterly GDP.

1.2 Objective

The objective of this study is to develop methods that improve the quality of the publication of the turnover growth of the manufacturing industry. This will result in more accurate results in the current publication or a more timely publication with the current accuracy.

1.3 Approaches

Several redesigns are investigated for improving the quality; they can be divided into three categories:

1. The "traditional" approach. "Traditional" refers to the fact that no additional information (leading indicators, related indicators, business tendency survey outcomes) is used and that no special action is undertaken to increase early response. The emphasis is on looking for better methods of imputation for non response.
2. The selective response chasing method. The aim of is to get the response of the most important enterprises, i.e. the largest enterprises, as early as possible. If selective response chasing is applied to produce timely indicators it may be an option to contact the enterprises in an early stage of the data collection period and ask them to make a best guess or a first estimate of their turnover. The realization of the method in daily practice is not discussed. It is assumed that all contacted enterprises are able to supply the requested data on a certain day after the reporting period.
3. The Expected Value Approach. In this approach the starting point is not "observed values", but "expected values", at an (intermediate) aggregate level. These expected values or "nowcasts" could for example be obtained

using historical information, information from tendency surveys or related indicators. Expected values are then replaced by observed values from enterprises once these become available. The real observations can also be used to improve the imputations for the remaining missing data. In this approach, preliminary outcomes can be obtained at any moment after the reporting period. They gradually evolve from "pure" nowcasts into truly statistical estimates, as more real observations become available.

These three approaches were explored for improving the quality of the short-term statistics of the growth of the turnover of the manufacturing industry.

1.4 Structure of report

The next chapter describes the present situation. A simulation model is developed for testing the present compilation method and several new methods. After that attention is paid to the new approaches for improving quality. In this chapter the results are also described on an aggregate level. The results on branch level are described in the appendices. Some conclusions are drawn in the final chapter.

2. Current method

Turnover growth statistics of the manufacturing industry are an important link in the production chain of the gross domestic product (GDP). This is described in the next section. Then some characteristics of the response of the enterprises are presented. Finally the compilation of the industrial turnover growth is described.

2.1 Publication of related statistics

The growth of the turnover of the manufacturing industry is an important source of information for the industrial production index. This is one of the most important short-term economic indicators. To a certain extent, this indicator gives a first impression of economic growth/decline. The industrial production index is commonly presented as a value added quantity index which refers to a particular base year. The production index is one of the most important inputs for the quarterly flash estimate of the Gross Domestic Product (GDP), which is released 45 days after the end of the reference quarter. The first release of the production index is around 40 days after the end of the reporting month, using the data collected until day 30[1, 2].

---

1 The production index is subject to revision in each of the subsequent four or five months as a result of more complete response. Additionally, the production index is adjusted annually to the three most recent years of National Accounts available.

2 Aelen(2003) assumes that the data of day 37 are used for publication. This is not correct because publication is at day 37 and one week is needed for processing the data. Therefore the data are downloaded from the database at day 30.

Manufacturing industry turnover growth is used as the primary source for the production index, if possible in combination with stock data. Producer price indices are used to arrive at volume measures. Statistics on industrial turnover growth are published monthly. The first release is around 37 days after the end of the reporting month using the data received until 30 days after the end of the reporting period. While this is not particularly slow, compared to other European countries, it is considerably slower than in the United States, where quarterly GDP is published after 25-30 days, and the first monthly industrial output estimate around the 15th after the reporting month.

2.2 Turnover response

In this paragraph some characteristics of the source data under consideration are investigated. First some general information of the respondent database is presented. This is followed by the response of the enterprises as a function of two parameters, namely the size of the enterprises and the branch. In the last subsection, the growth distribution of individual enterprises is presented.

With regard to turnover, turnover data from some 6500 enterprises in the manufacturing industry are collected on a monthly basis, mostly by mail. Data are available from January 1993 till December 2002. These enterprises are selected according to a cut-off mechanism: all enterprises with 20 employees or more are selected; all enterprises with less than 20 employees are left out completely.

The enterprises can be divided according to their size category (see Table 1) or SIC branch (see Table 2).

Size category	Number of employees
5	≥20 and ≤ 49
6	≥50 and ≤ 99
7	≥100 and ≤ 249
8	≥250 and ≤ 499
9	≥500

Table 1: Size categories

SIC	Branch
15/16	Food, beverages and tobacco
17-19	Textiles, clothing and leather products
20/26	Products of wood and building materials
21/22	Paper products, publishing, printing and reproduction
23-25	Petroleum/chemical/rubber/plastic
27-36	Metal products and electrical equipment
15-37	Aggregate manufacturing industry

Table 2: SIC-branches

Table 3 presents a representative example of the number of enterprises and the corresponding percentage turnover of the enterprises per SIC and size category. While size 9 accounts for 42 % of the total turnover of the manufacturing industry, it is only 233/6988=3% of the enterprises. The selective response chasing method uses this fact in section 3.2.

SIC	Number per size category						% Turnover per size category					
	5	6	7	8	9	Total	5	6	7	8	9	Total
15/16	499	244	155	129	67	1094	2%	2%	3%	7%	9%	23%
17-19	170	84	37	19	1	311	0%	0%	1%	0%	0%	2%
20/26	316	118	64	27	12	537	1%	1%	1%	1%	1%	4%
21/22	513	236	106	72	39	966	1%	1%	1%	2%	3%	9%
23-25	294	181	127	70	49	721	3%	2%	5%	4%	17%	30%
27-36	1921	784	373	188	65	3331	5%	4%	5%	7%	12%	32%
15-37	3732	1655	863	505	233	6988	12%	10%	14%	21%	42%	100%

Table 3: Number of enterprises and corresponding percentage turnover per size and SIC category.

Secondly, the response is studied as a function of days after the reporting period. Table 4 (on page 10) shows the response for the aggregate manufacturing industry. The response as a percentage of the turnover of the enterprises is depicted in Figure 1 for size categories 5, 7 and 9.

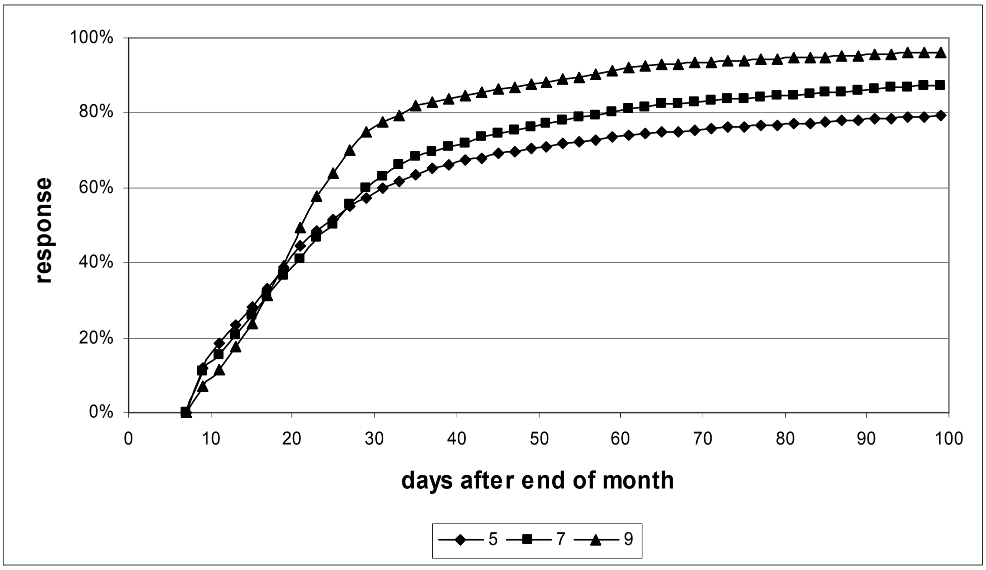

Figure 1: Response as percentage of turnover for size categories 5, 7, 9.

The largest enterprises have the lowest response a few days after the end of the reporting month, while the smallest enterprises have the highest response. After 30 days, the situation has turned around; the largest enterprises have a higher response than the smaller ones. After one year almost 100% of the turnover generated by the largest enterprises (size 9) is reported compared to 84 % of smallest (size 5) enterprises. Response of the enterprises that haven't responded after one year is expected not to be received. Size categories 6 and 8 are not depicted in Figure 1, their response rate is between 5 and 7 for size 6 and between 7 and 9 for size 8. Approximately 95 % of the total turnover is reported, the other 5 % imputed during the compilation.

The response as a percentage of the turnover is presented as a function of the branch in Figure 2. All branches show a similar behaviour except SIC 23-25 (Petroleum/chemical/rubber/plastic). This branch is a fast responder because large enterprises respond fast and this branch contains relatively many large enterprises.

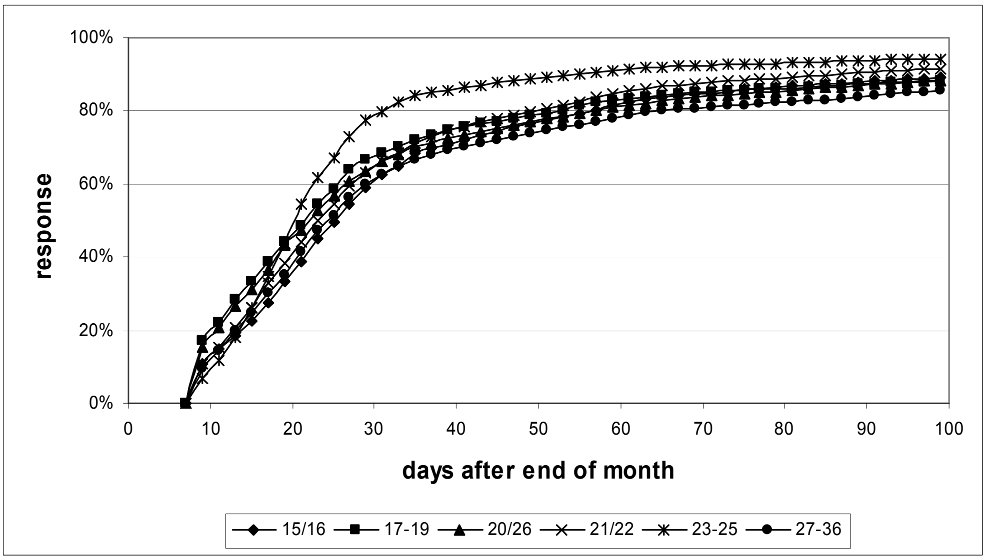

Figure 2: Response as percentage of turnover for all branches

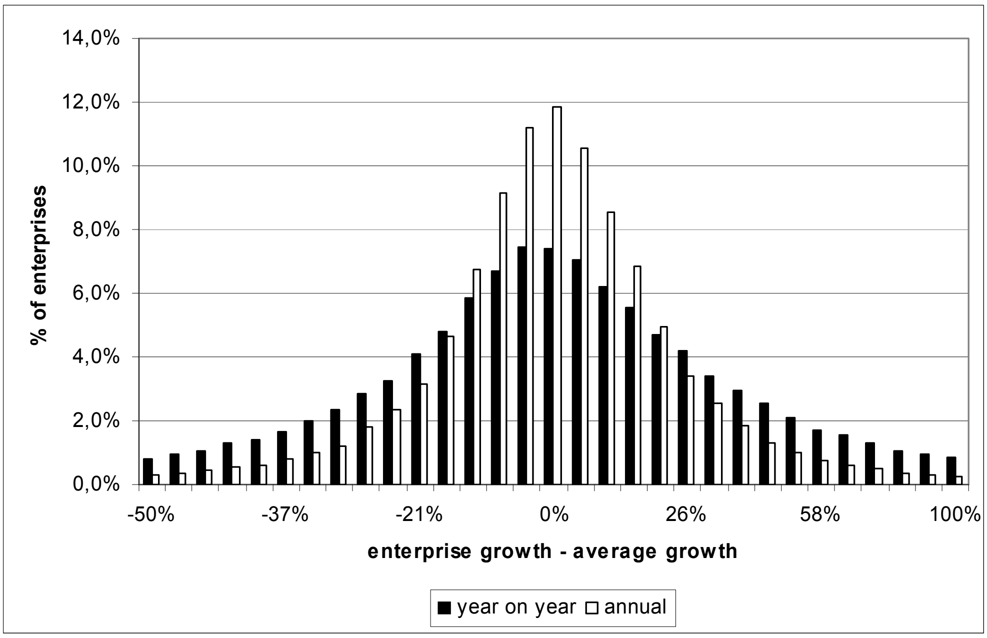

Figure 3: Distribution of enterprise turnover growth

The turnover growth of the individual enterprises is studied in Figure 3. It presents a histogram of the percentage of enterprises as a function of the difference between the average growth in turnover and the enterprise growth on a year on year basis and on an annual basis[3].

From this figure it is concluded that the distribution of the turnover growth per enterprise is extremely wide. This is explained by the assumption that enterprises don't have a continuous turnover but the work is divided in projects. Every month, a different number of projects are finished and the projects are of different value. As a result of this, the turnover differs a lot on a monthly basis. The most extreme example is a ship yard which finishes a ship once every x months and has the other month no turnover. Therefore it is very difficult to predict accurately the growth of an individual enterprise, using time series analysis with or without related indicators. This study will focus on meso- or macro level of the turnover.

2.3 Compilation method

Development of industrial turnover growth is commonly expressed as the percentage change over a twelve month period. This is calculated as the change in total turnover of those enterprises (n) that are in the sample of the reporting period (t) as well as in the sample twelve months earlier (t-12).

$$growth^t = \frac{\sum_{i=1}^{n} turnover_i^t - \sum_{i=1}^{n} turnover_i^{t-12}}{\sum_{i=1}^{n} turnover_i^{t-12}}$$

Growth is calculated for total, domestic and foreign turnover, with a division in the six branches of Table 2 (on page 4). Some enterprises change SIC code during a year. These are used in the aggregate of SIC 15-37, but not in the calculation of the turnover per branch. In case of missing observations (i.e. enterprises that have not yet reported), imputations are made, based on the assumption that the turnover would have moved in the same proportion as the sales of enterprises in the same 3-digit SIC-category (124 groups) which were recorded. Currently, the reference for imputation is the turnover of last month. Imputations are adjusted continuously as more information becomes available.

2.4 Simulations

A simulation model was developed to simulate the current compilation process and the alternative processes. The response data of January 1998 till December 2002 were used, e.g. 60 months. For each month, the final growth of the turnover is calculated. This is compared with the growth x days after the end of the reporting month, resulting in a function describing the error as a function of time. These 60 functions are aggregated to three time dependent functions, namely

---

[3] Year on year compares the turnover in month t compared to month t-12 and annual compares the turnover in year y with the turnover in year y-1.

- a root mean square value (RMSE), the root of average squared error.
- a maximum overestimation and a maximum underestimation. The extremes of the overestimation are the maximum positive errors and the extremes of the underestimation are the maximum negative errors. All error functions are between these two lines.

Figure 4 presents these error functions for the current compilation method for the aggregate turnover growth (RMSE in left figure and the extremes in right figure), the results for the branches of Table 2 (on page 4) in the appendix C.

All errors converge to zero as the number of days after the reporting month increases. Therefore month t-12 in the above equation has a negligible error. The first publication of the turnover growth after 37 days uses the data received until day 30 after the reporting month. The error in the turnover growth is determined by the error in the turnover of the reporting month. Currently, the turnover growth of the reporting month has a RMSE of 1.43 %-point (Aelen(2003)).

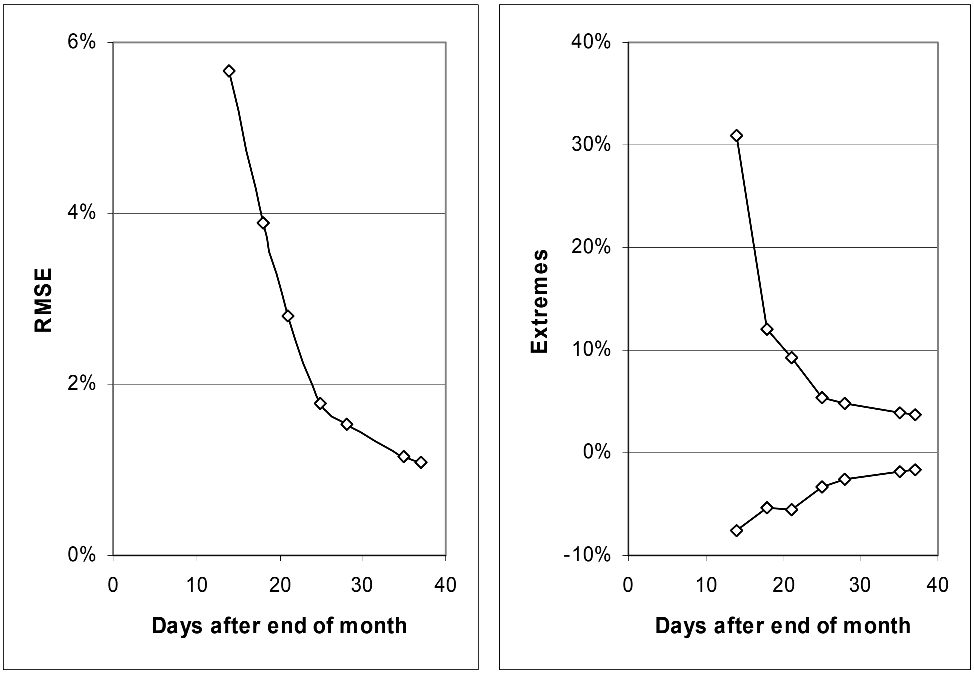

Figure 4: Performance of the current compilation method for the aggregate manufacturing industry turnover growth of the reporting month (Aelen (2003)).

3. Improved methods

3.1 Improved traditional method

Simulations demonstrate that the current imputation method can be improved with some simple adjustments[4]:

1. The best reference month for the imputation is t-12 instead of t-1.
2. Imputation groups based on the five size categories and the six branches (in total 30 groups) perform better than 124 SIC-groups.

The improved performance is shown in Figure 5. The RMSE is always lower for the improved method than for the current method. After day 20, the most extreme underestimation and overestimation are smaller for the improved method than for the current method. The timeliness of the aggregate turnover growth improves seven days based on the RMSE or the accuracy of the turnover of the reporting month improves from 1.43 %-points to 0.96 %-points on day 30. The table with the performance of the improved method for six branches is presented in the appendix C.

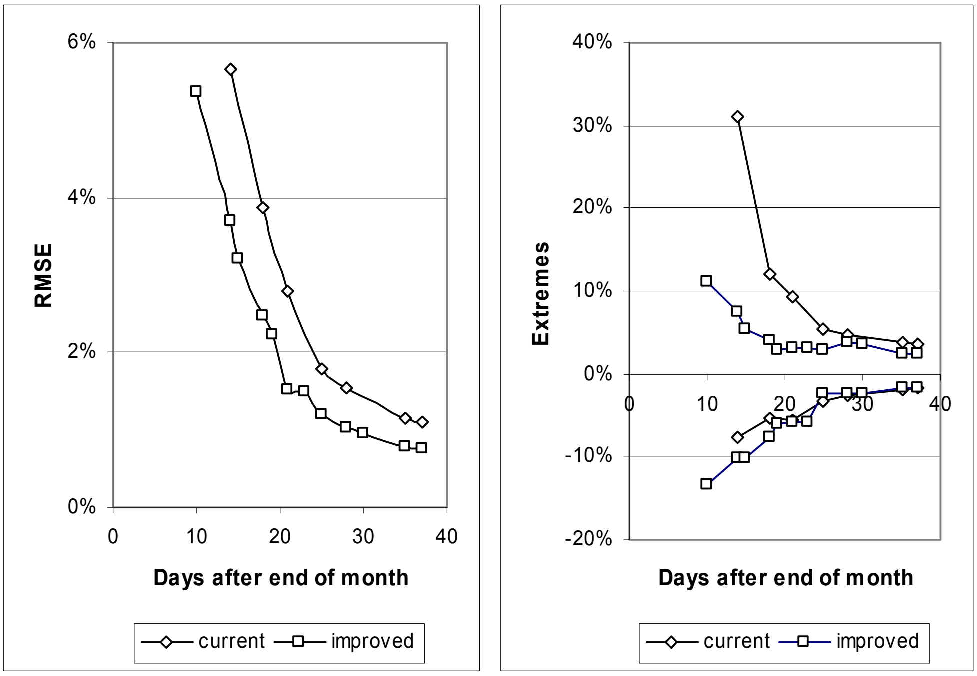

Figure 5: Comparison between the performance of the current and the improved traditional method for the aggregate turnover growth (in %-points) of the reporting month.

Frank Aelen (Aelen (2003)) has developed an additional time series model which corrects the imputation factor for the non-responding enterprises. The method is based on a division of the enterprises in two groups, one with the ones that have

[4] Frank Aelen (2003) arrives at similar conclusions.

responded in that month and the other with the ones that did not respond. The development of the turnover of both groups is also calculated for the preceding months and used as input for a time series model, called the regarima-method. It predicts the turnover of the non-respondents in this month. The regarima method improves the accuracy of the results more than the traditional method until day 21 after the reporting month. However after 21 days, the accuracy of the regarima method and the improved traditional method are almost similar. Note that the accuracy at day 21 is worse than the current data download for publication at day 30.

3.2 Selective response chasing

Better contact and cooperation with enterprises and trying to get their response as early as possible may significantly speed up the response. However, it may be too costly to do this for the entire sample.

Concentrating on the most important enterprises may be an interesting option. About 3% of the enterprises (with 500 employees or more) are responsible for more than 40% of industrial turnover, see Table 3. Assuming that chasing these enterprises would result in a response of 100% within 14 or 21 days, the response measured as percentage of the total turnover is increased significantly, see Table 4. Thereby the accuracy of the estimate is increased. The response of enterprises of sizes 5 till 8 is used as it is currently received.

Days after end of month	Current response	Day 14 response chasing	Day 21 response chasing
13	19,6%	19,6%	19,6%
15	25,1%	56,5%	25,1%
21	45,0%	65,8%	45,0%
23	51,3%	68,6%	68,6%
29	65,1%	75,4%	75,4%
35	72,4%	79,9%	79,9%
43	76,8%	82,8%	82,8%

Table 4: Response of enterprises with and without response chasing

Figure 6 presents the comparison between the improved method with and without selective response chasing for the aggregate turnover. The errors (RMSE and the extremes) drop significantly on day 15 and on day 22 for respectively response chasing on day 14 or on day 21. The current method has a RMSE error of 1.43 %-point after 30 days for the aggregate turnover growth. With selective response chasing, the RMSE is 1.28 %-point after 15 days and 0.90 %-point after 22 days for respectively 14 or 21 days response chasing.

The performance for the different branches is presented in appendix C. In case of 14 days response chasing, some branches (SIC: 17-19, 20/26 and 23-25) are on day 15 not as accurate as the current method (without improvements) on day 30. The accuracy of the branches on day 22 is in all cases improved for response chasing on day 21.

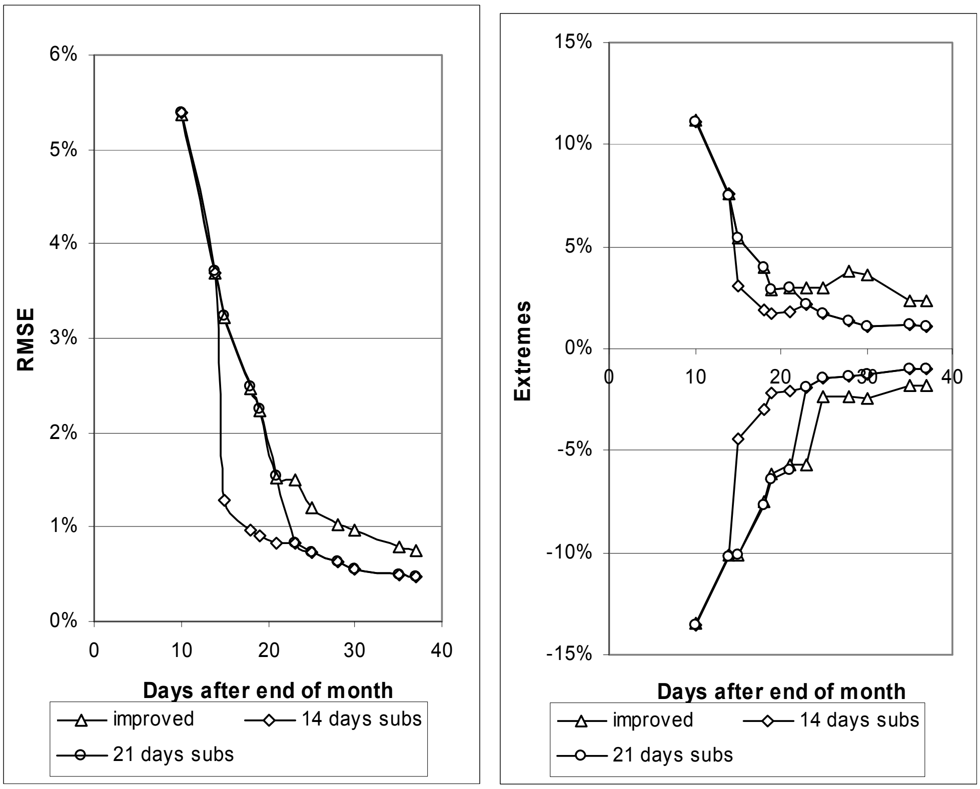

Figure 6: Performance of the selective response chasing method for the aggregate turnover growth in %-points

3.3 Expected value

At first glance, the expected value approach seems to offer an attractive combination of nowcasting and measuring. Instead of starting the statistical process with "observed values", the process begins with "expected values". This will be done on a macro level and in a later stage on a meso level for an aggregate of enterprises. These expected values are then replaced by statistical data from enterprises as soon as they become available. The estimates for the other missing values are also improved with this data. The processing system could be structured in such a way that an intermediate result can be produced at any given moment. In the first days following the end of the reporting period this provisional result is based primarily on models and approximations, but in the course of time the provisional result will gradually turn into a statistical result. At a certain point in time the aggregates will be reliable enough to be published, see Algera (2001).

The expected value method consists of the following steps. In the first step a turnover index is composed. Then in the second step, a forecasting method for the index is developed, i.e. the actual expected value method. Finally the expected values are combined with the response to determine when the results are accurate enough to publish.

3.3.1 *Index*

The analysis of the index is divided in the following topics. In the first paragraph, the index is derived from the microdata. Then an analysis is made of the components of the index. In the final part the index is deflated with the producer price index.

3.3.1.1 *Derivation of the index*

The derived index compares the turnover of the reporting month t with t-12 for the enterprises which are recorded in both months. Turnover data are available starting from January 1993. The first twelve months cannot be used for calculating a turnover growth and have therefore the starting value 100. January 1994 is the first real value. Indices are compiled for the aggregate turnover and the turnover of six branches (see Table 2 on page 4).

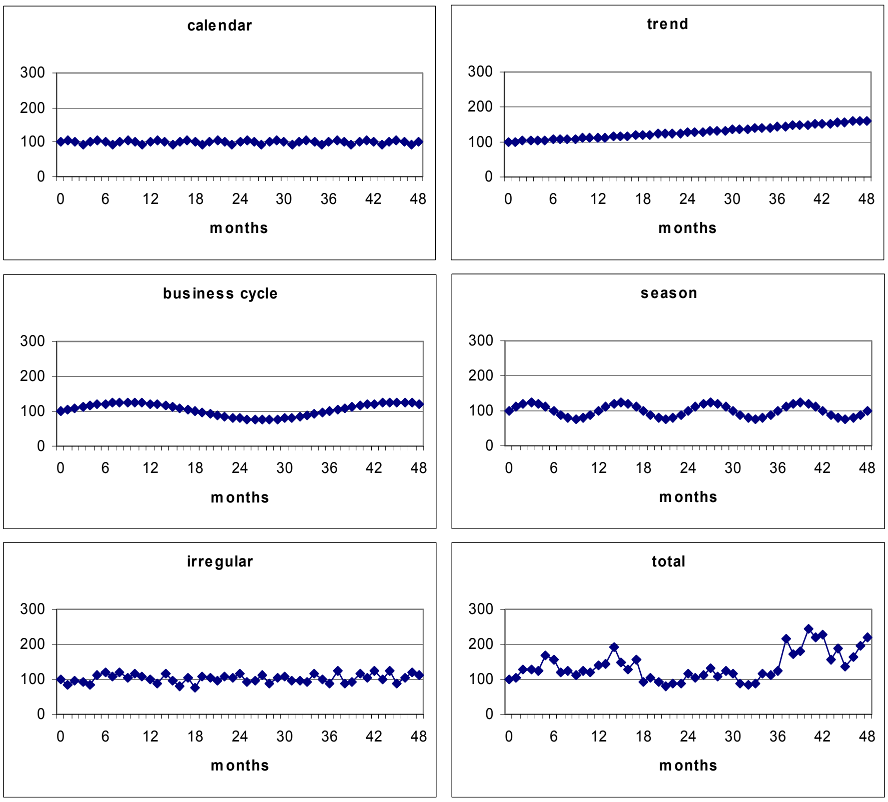

Figure 7 : Decomposition of an index in components

3.3.1.2 *Analysis of the components*

It is not possible to obtain an expected value of the turnover without splitting the index into its components. The standard approach is to make a distinction between the following five components. These components must be extrapolated into the future for an expected value:

1. calendar effects (C): for example January in a particular year has more weekend days than other years; this can result in less turnover[5]. These effects are called calendar effects. Because the number of working days is known in advance, calendar effects can be extrapolated easily.
2. the trend (T) is a linear function of the long term growth of the turnover. A linear function can be extrapolated without a problem.
3. the business cycle (B) is the short term fluctuation in the long term trend. The simplest method for extrapolation is using the last value of the business cycle. Other methods incorporate a series of values or information from related sources such as business tendency surveys.
4. the seasonal component (S) captures effects that are typical for the month or quarter. Several mathematical methods exist for predicting seasonal effects. However this component can be minimized by creating an index which compares month t with month t-12. This results in a very small seasonal effect which is influenced to a large extent by the irregular component. Therefore the seasonal component is unknown.
5. effects that are not incorporated in the other components are called irregular effects (I). These effects are very difficult to predict. One possible method could be using irregular effects in related information, such as tendency surveys. However in most cases, it is not possible to make a model based prediction.

The decomposition in components is carried out with X-12 ARIMA. Effects of the Dutch feast days are included in the calendar effects.

3.3.1.3 Deflation of the index

The next step is to deflate the turnover index to a volume index. The advantages of a volume index become clear after seasonal decomposition with X-12 ARIMA. The first advantage is that a volume index has smaller business cycle fluctuations, see Figure 8. The second and most important advantage is that the irregular fluctuations are reduced because they are partially caused by price fluctuations. Because both components are unknown, the deflated index is easier to predict.

---

[5] In some cases, more weekend days will result in more turnover. This is often caused by consumer spending in branches such as retail or catering industry

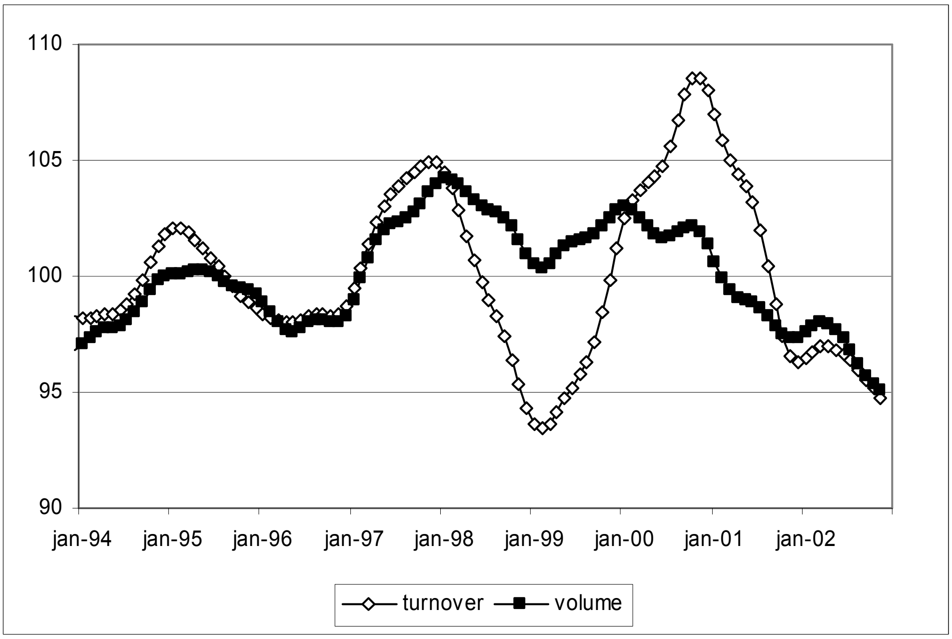

Figure 8: Difference between business cycle in volume and turnover index

3.3.2 Prediction methods for expected value

Two prediction methods have been tested, the first method uses the forecast module in X12-ARIMA and the second method is a combination of a seasonal decomposition with X12-ARIMA and an analysis with ARIMA models.

3.3.2.1 Forecasting with X12-ARIMA

A forecasting module is built in X12-ARIMA. It determines simultaneously the regression coefficients for the calendar effects (trading day, leap year and Dutch feast days) and the coefficients for the ARIMA prediction model (components 2 till 5 in the previous section) in a so-called Regarima model. It is also possible to incorporate outliers in the analysis; however this did not result in better predictions. The setup-file for X12-ARIMA is presented in appendix A. The advantage of this method is that it is easy to use. On the other hand the forecasting method selects the optimal model from a list of ARIMA-models in a pre-defined file (appendix B). The selection is on basis of two criteria. The first is the “in-sample” prediction error. The second criterion is an analysis of the correlation between the residuals (the Box-Ljung-test) and it is not possible to supply a model for each correlation. Therefore the results are expected to be not optimal.

	1 month ahead	**2 months ahead**	**3 months ahead**
RMSE	2.56%	2.58%	2.72%
Min	-7.20%	-7.42%	-7.80%
Max	8.21%	8.80%	9.62%

Table 5: Forecasting results with Regarima model.

Table 5 presents the results of the forecast module of X12-ARIMA. The root mean square (RMSE), minimum (Min) and the maximum (Max) overestimation are presented as function of the number of months ahead estimation. The difference between the errors for the different months is very small, which indicates to the fact that the errors are dominated by the irregular component.

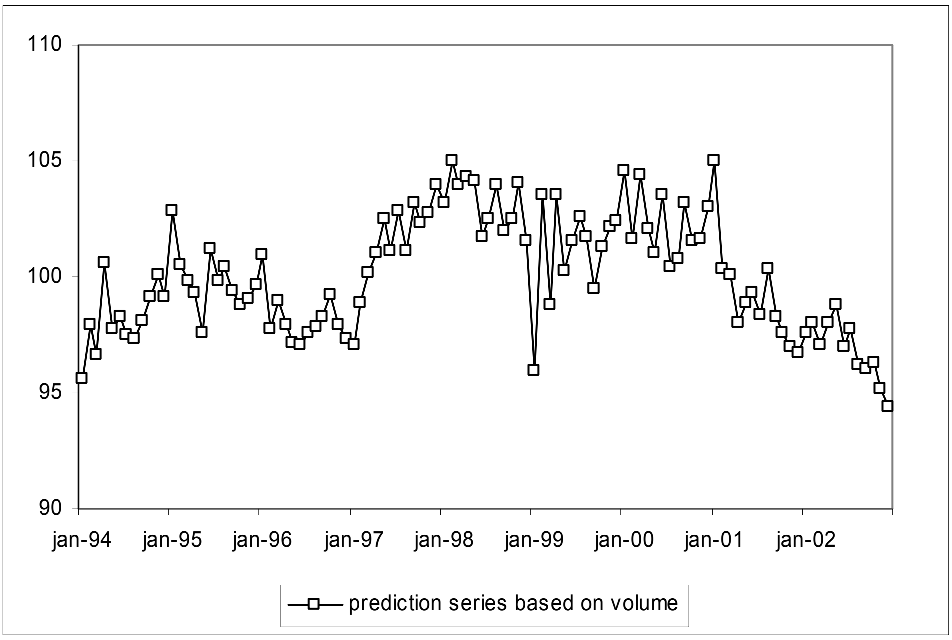

Figure 9: Time series to be predicted

3.3.2.2 Combination of seasonal decomposition with ARIMA models

The results of the Regarima module are the starting point for the analysis in this paragraph. This module presents a forecasting solution in which the calendar effects are separated from the index to be forecasted. The linear trend is removed from this index. The result is presented in Figure 9. The objective is to predict this remainder with a seasonal ARIMA model that is specified on basis of a Box-Jenkins derivation[6].

	1 month ahead	**2 months ahead**	**3 months ahead**
RMSE	2.39%	2.45%	2.96%
Min	-6.59%	-8.38%	-7.61%
Max	5.10%	4.13%	7.72%

Table 6: Forecasting results with an ARIMA analysis

[6] Several mathematical models were tested such as regression models, ARMA and ARIMA models. They were tested on time series based on the business cycle with and without irregular and seasonal effects. The best results were obtained with ARIMA models predicting a time series based on the business cycle and seasonal and irregular effects.

It was not possible to derive one ARIMA-model that fits each time series in the simulation period of 60 months. The correlation between the residuals varies over the time series. Besides a time dependence, there is also a branch-dependence of the derived ARIMA-models.

The results of the forecast are presented in Table 6. There is a slight improvement compared to the results presented in Table 5. In this case also, the accuracy of the results for 1 and 2 months ahead are very similar. This leads to the idea that the irregular component dominates the error. The results are not accurate enough for immediate publication. The expected value must be combined with early response to obtain sufficient accuracy (this will be described in the next section). The analysis is also carried out for separate branches, the expected value per branch has a larger RMSE.

A lot of effort was put into using related indicators to improve the accuracy of the prediction of the index, such as variables from business tendency surveys and the orders from the turnover survey. The correlation was studied between the business cycle component of index and of the related indicator. There is a correlation between the business cycle of the volume/value of the turnover and some variables from the business tendency surveys. However the improvement in the prediction of the business cycle with related indicators was negligible. The same was done for the irregular component. No irregular components of related indicators were found that have a stable significant correlation with the irregular component and therefore the use of related indicators did not improve the prediction.

Finally, a calculation is carried out which transforms the expected value into an expected growth of the turnover from month t-12 till month t. For this purpose, the expected value is combined with calendar effects, the linear trend and the measured prices.

days	6 branches	aggregate
7	90%	90%
9	85%	85%
11	90%	90%
13	95%	90%
15	85%	80%
17	70%	70%
19	55%	60%
21	20%	35%
23	20%	30%

Days	6 branches	aggregate
25	30%	30%
27	30%	15%
29	10%	0%
31	5%	0%
33	10%	0%
35	0%	0%
37	0%	0%
39	0%	0%
41	0%	0%

Table 7: Percentage of expected value used for imputation factor.

3.3.3 Combination of expected value and response

The final part is finding the optimal imputation factor based on a linear combination of expected value and early response as a function of the number of days after the reporting period. The optimal use of the expected value as a function of days after the reporting month is presented in Table 7. The response is used for the remaining part, i.e. 100 % - percentage expected value.

The expected values have been calculated for the aggregate turnover growth and for the 6 SIC branches of Table 2 (on page 4). Using expected values for six branches results in almost the same RMSE error and extremes for the aggregate turnover growth as using an aggregate expected value. However the accuracy of the estimation of the individual branches is better with an expected value for each branch. Therefore the results of the method using expected values for each branch are presented in Figure 10.

The expected value method improves the accuracy of the turnover growth during the first three weeks significantly, both the RMSE and the maximal overestimation and underestimation. If the current accuracy for publication is required, an expected value does not improve the timeliness. The expected value is considered not accurate enough.

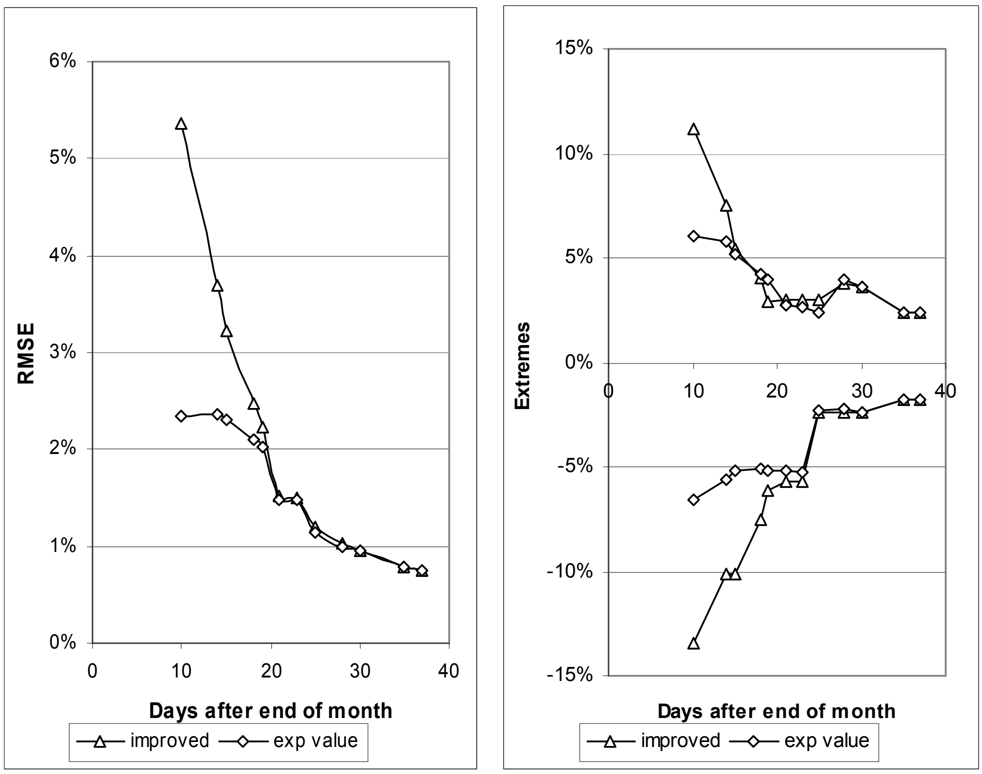

Figure 10: Performance of the expected value method for aggregate turnover growth in %-points

4. Conclusion

The following conclusions can be drawn:

- The current imputation method can be improved by replacing the 124 SIC-aggregates for 30 SIC-size aggregates for imputation of missing values and by using t-12 as the reference month instead of t-1. The method enables us to publish one week earlier, while maintaining the current level of accuracy.

- Response chasing of large enterprises results in an increase of the accuracy or/and the timeliness. The current response of day 14 combined with all the largest (size 9) enterprises can be used for a publication with the same accuracy as the current publication using data received until day 30. If all enterprises of size 9 have responded after 21 days, the publication is more accurate and timelier. With response chasing until day 14, the results for the 6 SIC branch are mixed, some are accurate enough to be published, others are less accurate than the current publication at day 30. Response chasing until day 21 will result in a more accurate result for the aggregate turnover growth and for the six branches.
- The use of related indicators from business tendency surveys to predict the business cycle does not improve the expected value because month on month variations are dominated by irregular movements. No indicators were found that predict this component.
- The expected value method improves the preliminary estimation; however the accuracy of the improved preliminary estimation is considered not enough to improve the timeliness of the publication. Series of expected values for six SIC branches are preferred above one series for the aggregate turnover because the estimation for the branches is better.

References

- Aelen, Frank (2003), Improving timelines of industrial short-term statistics using time series analysis, 27 November 2003, Statistics Netherlands
- Algera, S.B. (2001), Process redesign of short-term indicators and the role of Eurostat and NSI's, paper presented at the 3rd Meeting of the Task Force on Benchmarking in infra-annual economic statistics, Stockholm, 11 May 2001
- Vollebregt, M(2002), Syllabus bij de cursus seizoenscorrectie met X-12-ARIMA en Vivaldi ,26 November 2002, Statistics Netherlands (in Dutch)

Appendix A: X12-ARIMA setup file

```
SERIES{
        title = "arima2"
        Format = "DateValue"
        Period = 12
        name = "SINGLE"
        precision = 2
        decimals = 3
}
TRANSFORM{
        Function = auto
        adjust = none
}
REGRESSION{
        Variables =(
        td
        )
        user = (VAR1 )
        file = "H:\Vivaldiw\FEESTDAG.DAT"
        format = "datevalue"
}
AUTOMDL{
        file = "H:\Vivaldiw\x12\arima.mdl"
        mode = fcst
        method = best
        Identify = first
}
X11{
        seasonalma = msr
}
```

Appendix B: List of ARIMA models evaluated by X12-ARIMA

(0 1 1)(0 1 1) *

(1 1 0)(0 1 1) X

(1 1 1)(0 1 1) X

(0 1 2)(0 1 1) X

(2 1 0)(0 1 1) X

(1 1 2)(0 1 1) X

(2 1 1)(0 1 1) X

(2 1 2)(0 1 1) X

([2] 1 2)(0 1 1) X

(2 1 [2])(0 1 1) X

([2] 1 1)(0 1 1) X

(1 1 [2])(0 1 1) X

(2 1 [3])(0 1 1) X

(2 1 [4])(0 1 1) X

(2 1 [5])(0 1 1) X

(1 1 [5])(0 1 1) X

([2] 1 [5])(0 1 1) X

(2 1 [1 2 3 5])(0 1 1) X

(2 1 [1 2 4 5])(0 1 1) X

(2 1 [1 2 5])(0 1 1) X

(2 1 [5 6])(0 1 1) X

(2 1 [6])(0 1 1) X

(2 1 6)(0 1 1) X

(2 1 [1 5])(0 1 1)

Appendix C: results of the methods per SIC-category

The tables present the results of the discussed methods for the aggregate turnover growth (SIC: 15-37) (this page) and a division into six branches (the following pages). Each table presents the results of the current method, the improved (traditional) method, the selective response chasing method with response chasing after 14 and 21 days and finally the expected value method based on an expected value for the aggregate turnover growth and expected values for each branch. In all cases, the RMSE and the minimum and maximum estimation error are presented as a function of the number of days after the reporting period. All errors are given in %-points.

Errors in %-points for the aggregate turnover growth for the manufacturing industry (SIC 15-37)

days	Current method			Improved traditional method			14 days response chasing			21 days response chasing			Expected value aggregate			expected value 6 branches		
	RMSE(%)	min(%)	max(%)	RMSE(%)	min(%)	max(%)	RMSE(%)	min(%)	Max(%)	RMSE(%)	min(%)	max(%)	RMSE(%)	min(%)	max(%)	RMSE(%)	min(%)	Max(%)
10				5,37	-13,47	11,18	5,38	-13,51	11,13	5,38	-13,51	11,13	2,30	-6,52	6,06	2,34	-9,68	6,10
14	5,66	-7,54	31,01	3,70	-10,12	7,55	3,70	-10,12	7,55	3,71	-10,18	7,52	2,27	-5,64	5,82	2,36	-9,74	6,02
15				3,22	-10,09	5,44	1,28	-4,44	3,06	3,23	-10,14	5,43	2,19	-5,15	5,13	2,31	-9,29	5,46
21	2,80	-5,48	9,28	1,52	-5,70	3,02	0,83	-2,06	1,79	1,55	-5,99	3,02	1,40	-5,19	2,71	1,49	-6,43	2,66
23				1,50	-5,70	3,02	0,84	-1,89	2,13	0,84	-1,89	2,13	1,42	-5,26	2,64	1,48	-6,43	2,66
25	1,78	-3,32	5,43	1,20	-2,39	2,99	0,74	-1,42	1,68	0,74	-1,42	1,68	1,13	-2,32	2,42	1,14	-2,93	2,54
28	1,54	-2,52	4,74	1,02	-2,37	3,77	0,64	-1,33	1,38	0,64	-1,33	1,38	1,02	-2,22	3,92	1,00	-2,12	4,02
30	1,43			0,96	-2,43	3,64	0,56	-1,26	1,11	0,56	-1,26	1,11	0,96	-2,43	3,64	0,96	-2,30	3,73
35	1,16	-1,83	3,82	0,79	-1,79	2,39	0,49	-0,98	1,13	0,49	-0,98	1,13	0,79	-1,79	2,39	0,79	-1,79	2,39

Errors in %-points for the turnover growth of the food beverages and tobacco industry (SIC:15/16)

days	Current method			Improved traditional method			14 days response chasing			21 days response chasing			Expected value aggregate			expected value 6 branches		
	RMSE(%)	min(%)	max(%)	RMSE(%)	min(%)	max(%)	RMSE(%)	min(%)	Max(%)	RMSE(%)	min(%)	Max(%)	RMSE(%)	min(%)	max(%)	RMSE(%)	min(%)	max(%)
10				4,72	-10,53	10,26	4,74	-10,53	10,13	4,74	-10,53	10,13	5,93	-11,46	11,87	2,51	-5,66	6,84
14	3,77	-11,54	11,30	3,47	-10,50	6,27	3,47	-10,50	6,27	3,44	-10,50	5,82	5,12	-8,97	11,08	2,29	-4,87	5,72
15				3,28	-10,50	6,27	1,65	-3,22	4,00	3,24	-10,50	6,02	4,73	-8,50	10,18	2,24	-4,50	5,15
21	3,15	-5,86	11,65	2,58	-5,99	6,19	1,34	-3,30	2,45	2,57	-5,98	6,13	2,60	-5,83	5,11	2,23	-5,24	4,97
23				2,33	-5,58	5,64	1,33	-3,70	2,74	1,33	-3,70	2,74	2,33	-5,03	5,19	2,10	-5,49	4,87
25	2,50	-5,32	10,08	2,05	-5,11	3,61	1,23	-3,24	2,74	1,23	-3,24	2,74	2,14	-4,86	4,02	1,84	-5,05	3,05
28	2,17	-4,32	10,36	1,74	-4,27	3,32	1,08	-2,34	2,47	1,08	-2,34	2,47	1,76	-4,03	3,55	1,65	-4,17	3,34
30	1,98			1,57	-3,96	3,23	0,97	-2,23	2,15	0,97	-2,23	2,15	1,57	-3,96	3,23	1,54	-3,97	3,24
35	1,49	-3,67	4,74	1,33	-3,96	2,87	0,82	-2,23	1,78	0,82	-2,23	1,78	1,33	-3,96	2,87	1,33	-3,96	2,87

Errors in %-points for the turnover growth of the textiles, clothing and leather products industry (SIC:17-19)

days	Current method			Improved traditional method			14 days response chasing			21 days response chasing			Expected value aggregate			expected value 6 branches		
	RMSE(%)	min(%)	max(%)	RMSE(%)	Min(%)	max(%)	RMSE(%)	min(%)	max(%)	RMSE(%)	min(%)	max(%)	RMSE(%)	min(%)	max(%)	RMSE(%)	min(%)	max(%)
10				5,97	-16,56	14,04	5,97	-16,58	14,04	5,97	-16,58	14,04	4,98	-9,12	11,77	3,76	-10,92	9,17
14	5,37	-8,13	20,01	5,53	-15,69	14,48	5,53	-15,69	14,48	5,53	-15,69	14,48	4,47	-8,05	10,36	3,55	-11,21	6,99
15				5,22	-15,69	14,45	3,86	-19,26	6,36	5,22	-15,69	14,45	4,15	-7,51	10,03	3,34	-10,59	5,68
21	3,53	-6,70	8,86	3,26	-5,38	9,48	2,24	-6,66	6,64	3,26	-5,38	9,48	3,16	-4,66	9,75	3,00	-6,13	8,26
23				2,93	-5,06	8,91	2,06	-4,06	6,80	2,06	-4,06	6,80	2,83	-4,55	9,15	2,67	-4,60	8,06
25	2,93	-9,30	7,31	2,79	-5,36	7,72	1,98	-4,61	5,60	1,98	-4,61	5,60	2,66	-4,91	8,17	2,54	-4,39	6,69
28	2,47	-9,78	4,98	2,32	-5,00	7,46	1,80	-4,50	5,32	1,80	-4,50	5,32	2,28	-4,64	7,53	2,16	-4,60	6,84
30	2,42			2,13	-5,00	7,38	1,70	-4,41	5,26	1,70	-4,41	5,26	2,13	-5,00	7,38	2,07	-4,85	7,18
35	2,31	-8,33	5,37	2,04	-4,49	6,40	1,62	-3,95	4,30	1,62	-3,95	4,30	2,04	-4,49	6,40	2,04	-4,49	6,40

Errors in %-points for the turnover growth of the products of wood and building materials industry (SIC:20/26)

days	Current method			Improved traditional method			14 days response chasing			21 days response chasing			Expected value aggregate			expected value 6 branches		
	RMSE(%)	min(%)	max(%)	RMSE(%)	Min(%)	max(%)	RMSE(%)	min(%)	max(%)	RMSE(%)	min(%)	Max(%)	RMSE(%)	min(%)	max(%)	RMSE(%)	min(%)	max(%)
10				7,94	-26,06	23,40	7,94	-26,06	23,40	7,94	-26,06	23,40	6,72	-21,07	16,56	4,29	-9,02	10,59
14	7,15	-17,83	29,97	5,98	-11,08	21,41	5,98	-11,08	21,41	5,97	-11,08	21,41	5,48	-11,33	14,92	3,93	-10,25	10,40
15				5,79	-11,61	21,41	5,56	-10,92	22,77	5,79	-11,61	21,41	5,04	-9,70	13,89	3,68	-10,49	9,86
21	4,02	-13,81	15,30	2,89	-10,10	6,75	2,83	-8,36	10,79	2,89	-10,10	6,75	2,60	-7,44	7,04	2,59	-7,80	5,64
23				2,30	-6,11	6,62	2,43	-5,54	10,79	2,43	-5,54	10,79	2,17	-6,36	6,18	2,15	-6,59	5,59
25	3,61	-11,76	14,68	2,10	-4,20	7,41	1,93	-4,71	5,33	1,93	-4,71	5,33	2,05	-4,35	5,52	1,96	-4,45	5,99
28	2,74	-8,31	6,28	2,10	-5,95	6,72	1,90	-6,27	4,63	1,90	-6,27	4,63	2,04	-6,05	5,80	1,97	-6,36	5,86
30	2,64			1,95	-7,07	5,26	1,79	-7,33	3,95	1,79	-7,33	3,95	1,95	-7,07	5,26	1,89	-7,13	4,98
35	2,39	-6,56	6,12	1,64	-5,85	4,24	1,56	-6,11	4,34	1,56	-6,11	4,34	1,64	-5,85	4,24	1,64	-5,85	4,24

Errors in %-points for the turnover growth of the paper products, publishing, printing and reproduction industry (SIC:21/22)

days	Current method			Improved traditional method			14 days response chasing			21 days response chasing			Expected value aggregate			expected value 6 branches		
	RMSE(%)	min(%)	max(%)	RMSE(%)	Min(%)	max(%)	RMSE(%)	min(%)	Max(%)	RMSE(%)	min(%)	max(%)	RMSE(%)	min(%)	max(%)	RMSE(%)	min(%)	max(%)
10				7,97	-16,92	24,55	7,98	-16,92	24,55	7,98	-16,92	24,55	4,52	-8,71	10,55	2,73	-10,30	5,41
14	5,54	-12,16	13,34	4,69	-12,44	15,61	4,69	-12,44	15,61	4,71	-12,44	15,61	4,13	-8,89	8,16	2,45	-9,29	5,28
15				4,49	-11,59	15,61	2,16	-4,36	7,44	4,51	-11,59	15,61	3,89	-8,84	7,70	2,35	-8,89	5,14
21	4,39	-9,37	13,71	2,36	-6,83	3,54	1,62	-6,30	3,15	2,40	-6,83	3,54	2,31	-6,46	4,12	2,14	-7,46	3,11
23				2,31	-7,20	4,26	1,59	-6,30	3,15	1,59	-6,30	3,15	2,27	-7,61	5,14	2,15	-7,46	4,51
25	3,26	-5,22	12,17	2,20	-6,59	4,91	1,50	-5,78	2,98	1,50	-5,78	2,98	2,20	-6,63	5,21	2,02	-7,50	4,46
28	2,86	-5,49	10,30	1,70	-4,40	4,54	1,22	-3,62	2,61	1,22	-3,62	2,61	1,72	-4,56	4,62	1,63	-4,69	4,31
30	2,61			1,56	-4,17	3,91	1,17	-3,69	2,59	1,17	-3,69	2,59	1,56	-4,17	3,91	1,53	-4,41	3,78
35	2,00	-8,58	5,39	1,35	-2,85	2,63	0,99	-2,73	1,84	0,99	-2,73	1,84	1,35	-2,85	2,63	1,35	-2,85	2,63

Errors in %-points for the turnover growth of the petroleum/ chemical/ rubber/ plastics industry (SIC:23-25)

days	Current method			Improved traditional method			14 days response chasing			21 days response chasing			Expected value aggregate			expected value 6 branches		
	RMSE(%)	min(%)	max(%)	RMSE(%)	Min(%)	max(%)	RMSE(%)	min(%)	max(%)	RMSE(%)	min(%)	Max(%)	RMSE(%)	min(%)	max(%)	RMSE(%)	min(%)	max(%)
10				15,86	-38,96	37,55	15,90	-38,92	37,52	15,90	-38,92	37,52	11,68	-26,53	14,35	3,87	-11,06	6,90
14	19,20	-11,33	140,36	9,53	-27,91	19,48	9,53	-27,91	19,48	9,54	-27,89	19,46	9,72	-24,15	13,89	3,42	-10,58	7,99
15				7,91	-25,36	11,53	2,68	-8,60	9,40	7,90	-25,32	11,53	8,98	-23,62	13,45	3,55	-12,64	7,93
21	3,28	-7,83	11,78	3,67	-10,48	8,68	1,83	-5,56	4,90	3,67	-10,48	8,68	4,10	-12,00	7,97	3,38	-9,75	7,73
23				3,36	-10,48	6,07	1,70	-4,82	4,69	1,70	-4,82	4,69	3,62	-11,78	5,91	3,15	-9,75	5,66
25	2,27	-6,90	6,60	3,18	-9,68	7,67	1,53	-5,12	3,24	1,53	-5,12	3,24	3,20	-10,97	7,18	2,90	-8,77	7,46
28	1,67	-3,11	6,70	2,22	-5,23	5,73	1,27	-4,36	2,35	1,27	-4,36	2,35	2,21	-5,54	5,62	2,13	-4,55	5,51
30	1,58			1,87	-4,50	5,26	1,04	-3,09	2,14	1,04	-3,09	2,14	1,87	-4,50	5,26	1,87	-4,29	5,27
35	1,35	-2,92	5,80	1,28	-1,64	4,04	0,86	-2,58	2,13	0,86	-2,58	2,13	1,28	-1,64	4,04	1,28	-1,64	4,04

Errors in %-points for the turnover growth of the metal products and electrical equipment industry (SIC:27-36)

days	Current method			Improved traditional method			14 days response chasing			21 days response chasing			Expected value aggregate			expected value 6 branches		
	RMSE(%)	min(%)	max(%)	RMSE(%)	Min(%)	max(%)	RMSE(%)	min(%)	max(%)	RMSE(%)	min(%)	max(%)	RMSE(%)	min(%)	max(%)	RMSE(%)	min(%)	max(%)
10				6,25	-17,06	18,01	6,25	-17,91	18,00	6,25	-17,91	18,00	4,57	-12,24	9,11	4,66	-19,77	9,10
14	9,55	-13,69	27,74	4,59	-15,60	9,26	4,59	-15,60	9,26	4,59	-16,46	9,25	4,21	-11,71	8,51	4,58	-19,41	10,38
15				4,37	-14,58	8,88	2,35	-7,45	2,94	4,36	-15,45	8,86	4,02	-11,31	8,06	4,45	-18,99	9,91
21	7,19	-12,16	26,60	2,92	-9,04	5,10	1,85	-4,95	3,04	2,91	-9,87	5,15	2,95	-8,21	5,96	2,93	-10,64	5,85
23				2,72	-9,04	6,01	1,59	-4,52	2,24	1,59	-4,52	2,24	2,78	-8,33	6,83	2,77	-10,64	7,00
25	4,34	-8,92	14,58	2,30	-7,99	6,16	1,45	-3,01	2,44	1,45	-3,01	2,44	2,38	-7,97	5,90	2,46	-10,57	5,79
28	4,17	-8,94	12,97	1,87	-5,26	4,48	1,43	-3,16	2,76	1,43	-3,16	2,76	1,88	-5,19	4,80	1,90	-5,20	5,18
30	3,88			1,77	-4,50	4,27	1,34	-3,12	2,85	1,34	-3,12	2,85	1,77	-4,50	4,27	1,77	-4,76	4,53
35	3,15	-5,88	12,34	1,45	-3,17	3,16	1,07	-2,27	2,29	1,07	-2,27	2,29	1,45	-3,17	3,16	1,45	-3,17	3,16