Improving timeliness of industrial short-term statistics using time series analysis

Discussion paper 04005

Frank Aelen

Statistics Netherlands

Voorburg/Heerlen, April 2004

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
2003–2004	= 2003 to 2004 inclusive
2003/2004	= average of 2003 up to and including 2004
2003/'04	= crop year, financial year, school year etc. beginning in 2003 and ending in 2004

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: 6008304005

Statistics Netherlands

Contents

IMPROVING TIMELINESS OF INDUSTRIAL SHORT-TERM STATISTICS USING TIME SERIES ANALYSIS

Summary: National statistical institutes of the European Union are under growing pressure to improve the timeliness of short term economic statistics. The aim of this study was to improve the timeliness of the monthly statistics on sales development in the Dutch manufacturing industry. The focus was on producing timely indicators in the presence of missing data. Currently, missing data are imputed using, besides already available data for the month in question, known data from only one previous month, whereas we propose to use whole time series for that purpose. Furthermore, since there is a trade-off between timeliness and accuracy, the relationship between these two quality measures was investigated in order to be able to make a more balanced decision between them. It was concluded that with respect to the currently used method, timeliness can be improved from about 37 days to about 27 days after the end of the month, without sacrificing accuracy. Further improvement of timeliness is possible at the cost of reduced accuracy.

Keywords: short-term statistics, manufacturing industry, timeliness, accuracy, imputation, time series analysis, X-12-ARIMA

1. Introduction

The past few years, national statistical institutes of the European Union are under growing pressure to improve the timeliness of short term economic statistics. Many of the important users of these statistics, such as the European Central Bank (ECB), Eurostat, the European Commission (EC), the International Monetary Fund (IMF), governments and financial market analysts, consider it a key issue that preliminary data are disseminated earlier.

There is, however, a trade-off between timeliness and accuracy. Improving timeliness without improving the survey process, would almost certainly reduce the accuracy. The challenge is to improve the methods so that when data are disseminated earlier the current level of accuracy can be maintained. On the other hand, for some statistics, reduced accuracy may be acceptable if this results in greatly improved timeliness. Therefore, to make a balanced decision in the trade-off between timeliness and accuracy, it is important to know the relationship between them. This relationship varies among different statistics, depending on the heterogeneity of the underlying population of enterprises and the specific survey process used.

The aim of this study was to improve the timeliness of short-term statistics, in particular the monthly statistics on sales development in the Dutch manufacturing

industry. This statistics plays, among other things, an important role in the forecasting of economic development and the setting of monetary policy. Furthermore, for this particular statistics the relationship between timeliness and accuracy was to be investigated.

The outline of this paper is as follows. In section 2 the current survey process at Statistics Netherlands regarding the monthly statistics on sales development in the manufacturing industry is described briefly, including possible points for improvement. This particular survey is based on an integral collection of data. Section 3 deals with one of the possible improvements, i.e. the imputation of missing data. Currently, missing data are imputed using a trend-based method that uses, besides already available data for the month in question, known data from only one previous month. This method is investigated and some improvements are suggested. Furthermore, we propose an alternative method to impute missing data using whole time series of data. Note that already available data for the month in question are also used here. Simulations to determine the accuracy of the current and improved trend-based methods, as well as of the proposed time series method, as a function of timeliness, are described in section 4. Results of the simulations are presented in section 5. The methods and results are discussed in section 6, and some conclusions are drawn in section 7.

2. Overview of current survey process and possible improvements

2.1 Current survey process at Statistics Netherlands[2]

The statistics on sales development in the manufacturing industry is based on the population of enterprises within the SIC divisions 15-37 (Appendix A) that have more than 20 employees. In the period 1998-2002 this concerned on average 6,450 enterprises.

Sales data of these enterprises are collected integrally on a monthly basis, mainly by paper questionnaires. Most of the enterprises respond within a period of a few months after the end of the month in question. Figure 1 shows, as a function of the number of days after the end of a month, the percentage of enterprises that has responded, as well as the percentage of turnover associated with these enterprises. Currently, preliminary values for the sales development in the manufacturing industry are disseminated on average 37 days after the end of the month. At that time 62%±4% (mean ± standard deviation; period 1998-2002) of the enterprises has responded, corresponding to 74%±4% of the turnover. To compare, 14 days after the

[2] During the time this study was done another system (IMPECT) to compile short-term statistics became operational. In section 6 it is shown, however, that results of this new system, regarding the assessment of sales development in the manufacturing industry, are expected to be similar to the method described here.

end of the month only 26%±6% of the enterprises has responded, corresponding to 23%±8% of the turnover. Eventually, 88%±6% of the enterprises responded, corresponding to 97%±2% of the turnover.

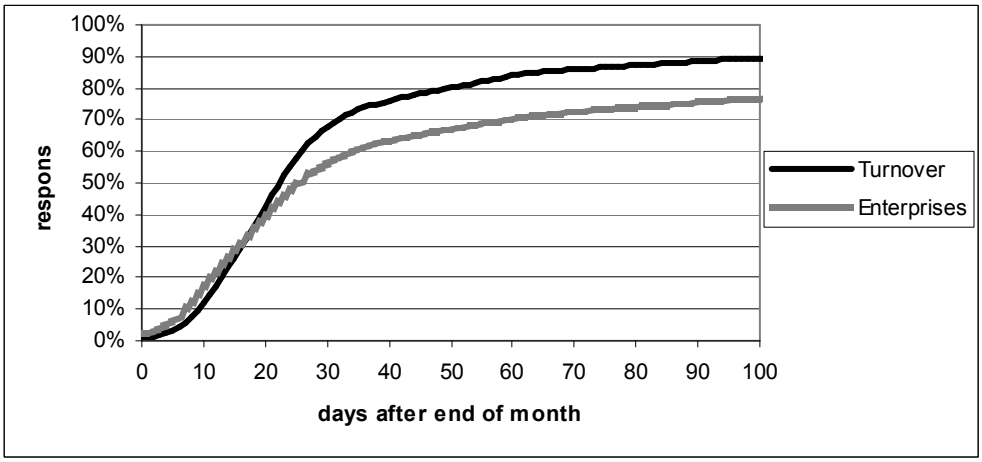

Figure 1. Percentage of enterprises that has responded to the questionnaire and the percentage of turnover associated with these enterprises, both as a function of the number of days after the end of a month.

Completed questionnaires are edited within a day after receipt in order to remove evident errors. This data editing is done in part automatically (e.g. when the turnover is mistakenly not given in thousands of euros) and in part manually (e.g. when the turnover is not plausible with respect to previous values). In some cases the respondent is contacted.

The next step in the process is the imputation of missing data. At any moment in time not all enterprises have responded (yet) to the questionnaire. The missing data are imputed using a trend with respect to the previous month. This trend is determined using an aggregate of more or less similar enterprises that do have responded. Since data come in not only for the latest month but also for earlier months, the still necessary imputations of those earlier months are redone with the most recent information available. More details on the imputation method are given in section 3.

Finally, the sales development with respect to the same month of the previous year is calculated. In doing this, only the enterprises are used that are in the intersection of the populations of those two months. So, enterprises that, for one reason or another, have moved in or out of the population during the year, are left out of the calculation. Then, the sales development of month t with respect to one year before is assessed straightforwardly as:

$$\Delta T_t = \left(T_t - T_{t-12}\right) / T_{t-12}\,, \tag{1}$$

where ΔT is the sales development, T is the total turnover of all enterprises in the intersection and subscripts indicate the month. Since all enterprise data are known, either imputed or not, the sales development can also be assessed for any desired subdivision of the manufacturing industry. Note that this subdivision is independent of the aggregates of enterprises that were used for the imputation of missing data.

Moreover, during a year enterprises can migrate from one subdivision of the population to another. The sales development of a subdivision is based on the enterprises belonging to the intersection of the subdivisions at months t and t-12.

Statistics Netherlands disseminates (preliminary) sales development data on:

- The manufacturing industry as a whole: SIC divisions 15-37 (MI1);
- Manufacture of food, beverages and tobacco: SIC divisions 15/16 (MI6-1);
- Manufacture of textiles, clothing, leather and leather products: SIC divisions 17-19 (MI6-2);
- Manufacture of wood, products of wood and building materials: SIC divisions 20/26 (MI6-3);
- Manufacture of paper and paper products, printing and publishing: SIC divisions 21/22 (MI6-4);
- Manufacture of refined petroleum, chemical, rubber and plastic products: SIC divisions 23-25 (MI6-5);
- Manufacture of metal products, electrical and transport equipment, furniture and other manufacturing: SIC divisions 27-36 (MI6-6).

In parentheses the codes are given that are used in this paper to indicate the manufacturing industry as a whole (MI1), as well as the six subdivisions of the manufacturing industry (MI6).

2.2 Possible improvements

A lot of literature is available on the subject of improving timeliness of short-term statistics. Here, only some of the most important possible improvements are mentioned, in relationship to the current survey process of the statistics on sales development in the Dutch manufacturing industry. Some of these possible improvements have only been referred to in literature, and have not yet been fully investigated. A lot of interesting ideas are discussed in the final report of the Expert group on Sampling for Timely European Indicators (ESTEI - Deville et al, 2002), where also many references can be found.

The use of paper questionnaires is, with respect to timeliness, not very efficient. More advanced data collection methods, such as internet/email, touchtone data entry, fax, or telephone (possibly only for a first best guess), will probably improve timeliness. Also adapting questionnaires more to the accounting practices of enterprises may be helpful.

A further improvement for compiling early preliminary data may be the use of a subsample of enterprises, for which a lot of effort is put into getting early figures (Deville et al, 2002). In that case, the largest enterprises should certainly be part of the subsample.

Completed questionnaires are edited after receipt in order to remove evident errors. Depending on the number and size of questionnaires, editing can be quite time-consuming. Selective editing techniques, in which an algorithm is used to automatically select the most important (possible) errors, may reduce editing time.

Such algorithms are also in use at Statistics Netherlands, but these may be improved by using time series analysis (Revilla, 2002) instead of using data from only one or a few previous months.

After data collection and editing, the still missing data have to be dealt with. There are a number of approaches to do this. The missing data can be estimated for each enterprise separately or on an aggregated level. In case the missing data are estimated on an aggregated level, it is still possible to pass the aggregate estimates back to the individual enterprises. This latter imputation approach is the one currently employed at Statistics Netherlands in the assessment of sales development. The current method, however, uses besides the already received data for the month in question, data from only one previous month. Better estimates can probably be made by using all historical enterprise data available, i.e. using whole time series of enterprise data.

An alternative approach to deal with missing data, is to directly estimate the population total from the already received data. Generally spoken, design-based procedures, like the calibration estimator or the generalized regression estimator, and model-based procedures, which use e.g. econometric or time series models, can be distinguished. This approach facilitates the use of auxiliary information that is strongly correlated with the short-term indicator in question in order to improve the accuracy of the estimates. In that case, a requirement is that the auxiliary information is already available at the desired dissemination time. Auxiliary information that may be used for instance is data from business cycle surveys, in which enterprises are asked to assess, in a qualitative manner, their present situation and their expectations for the near future. Results from this survey are generally available already at the end of the month.

Summarizing, using more advanced data collection and editing techniques can be very important in improving timeliness of short-term statistics. In conjunction, methods can be developed to deal with missing data, since even when data are collected faster, there will still be missing data at any point in time. In this study we focussed on improving the imputation method used at Statistics Netherlands in the compilation of the statistics regarding sales development in the manufacturing industry.

3. Imputation of missing data

3.1 Introduction

Suppose at a certain point in time one wants to know the sales development for month t for a certain population of enterprises (this can be the entire manufacturing industry or any part of it). This development can be derived from the turnover values of the individual enterprises in the population. Then, at that point in time, the problem arises that the data for month t are not yet known for all individual

enterprises. In that case, the population of enterprises can be divided into two groups, i.e. a group for which the data are known (group I) and a group for which the data are not known (group II). We assume a complete enumeration of the population and historical data for individual enterprises belonging to group I as well as group II are known integrally, either imputed or not. These historical data can be used to facilitate the imputation of missing data for group II for month t. In order to do the imputations, first an estimate Y_{II}^{t} of the missing data is made on the aggregate level, and subsequently this aggregate estimate is passed back to the individual enterprises.

The aggregate turnover for group II for month t (Y_{II}^{t}) that is needed for the imputations can be estimated in several ways. Below, first a trend-based method using data from only one earlier month is described (section 3.2). The method currently used at Statistics Netherlands is based on this principle. Subsequently, a new method based on time series analysis is introduced (section 3.3). In section 3.4 the method of imputing missing enterprise data, given the aggregate estimates, is explained. Section 3.5 concludes with some practical aspects concerning the implementation of the methods. First the currently used trend-based method is treated, second the trend-based method with some improvements, and third the method based on time series analysis.

3.2 Trend-based methods using one earlier month

Often employed methods to estimate the missing data use historical data from only one earlier month, e.g. the previous month (Δt=1) or the month one year before (Δt=12). These methods assume that on an aggregate level the trend from the earlier month to month t is the same for groups I and II. The observed trend for group I can then be applied to the value of group II for the earlier month, in order to estimate the aggregate turnover of group II for month t. So,

$$Y_{II}^{t} = Y_{II}^{t-\Delta t} \frac{Y_{I}^{t}}{Y_{I}^{t-\Delta t}}, \qquad (2)$$

where Y is the total turnover, the subscript indicates the group and the superscript indicates the month[3].

Graphically, this method can be interpreted as follows (figure 2). First, a straight line is drawn from the origin through the point ($Y_{I}^{t-\Delta t}, Y_{II}^{t-\Delta t}$). Then Y_{II}^{t} is found as the vertical value on the line at the horizontal position Y_{I}^{t}.

[3] Do not confuse the term trend as it is used here with the term trend as it is sometimes used in seasonal adjustment methods (e.g. in X-12-ARIMA).

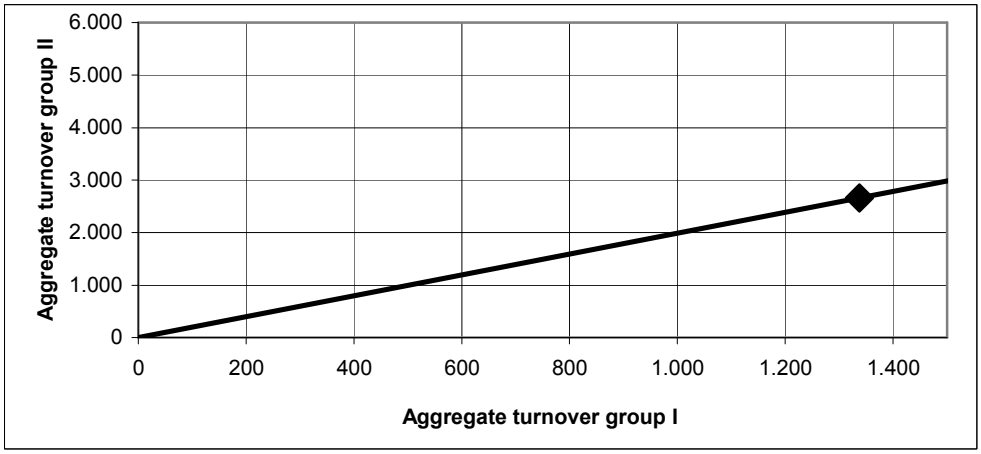

Figure 2. Method using data from only one earlier month. The dot refers to the values for that earlier month. The estimate for group II at month t is found as the vertical value at the straight line at the horizontal position defined by the turnover for group I at month t.

3.3 Proposed method using time series analysis

Instead of using data from only one earlier month in order to estimate Y_{II}^{t}, it is possible to use whole time series of data for groups I and II (figure 2) for that purpose. These time series are constructed by assigning each enterprise in the population at month t to either group I or group II, based on the availability of data for month t. Subsequently, the turnover values of all enterprises are summed, per group, for month t as well as for all previous months.

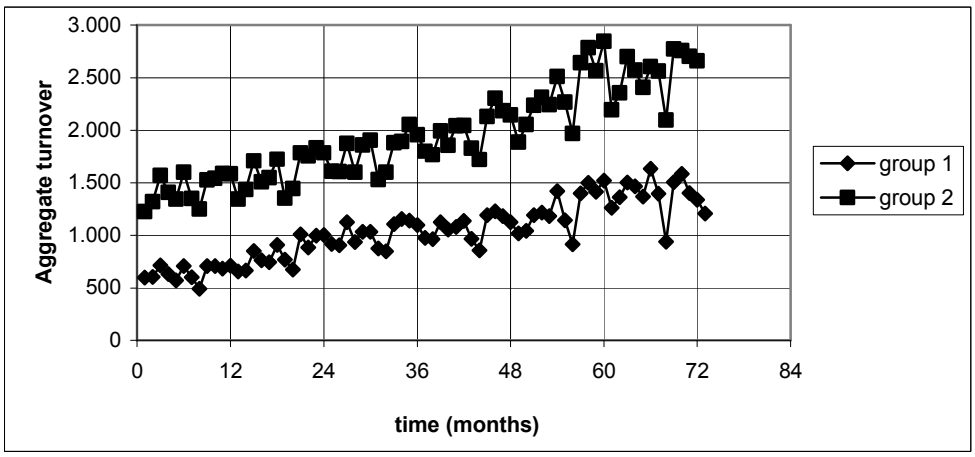

Figure 2. Time series of aggregate turnover for groups I and II. In this example the groups are defined based on the availability of data for month 73. So, group I has data until month 73 and group II has data until month 72. The aim is to estimate the turnover value for group II for month 73.

Similarly as was done in figure 1 for one earlier month, aggregate turnover for group I against group II can now be plotted for all months (excluding month t) in the time series (figure 3). The straight line shown in figure 3 is the regression line of the aggregate turnover of group II on the aggregate turnover of group I, with an intercept, based on ordinary least squares estimation. It can clearly be seen that the

regression line has an intercept that is different from zero. This is generally the case for the time series investigated in this study.

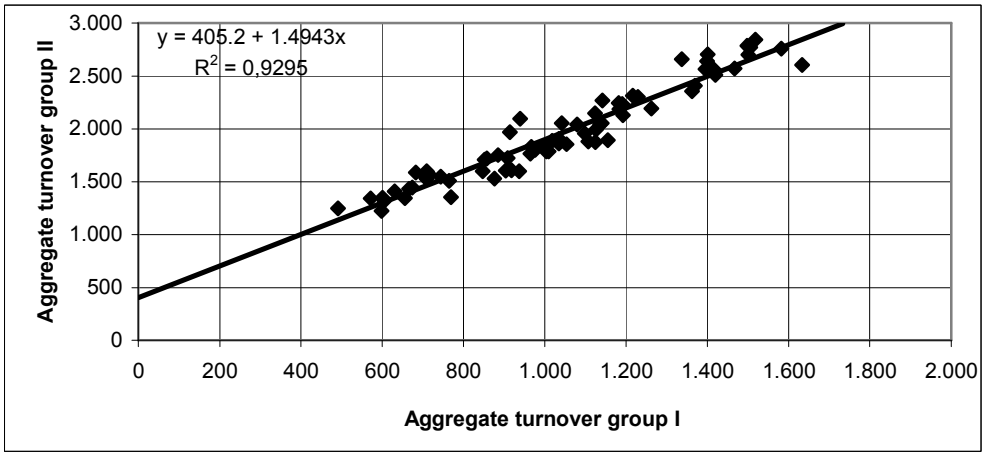

Figure 3. Regression plot of group I against group II for all months in the time series shown in figure 2. Note that the time dimension cannot be seen in this kind of plot.

Because each of the time series of groups I and II is based on the same enterprises at every month, the data points in the regression plot are not independent, i.e. they are correlated in time. Figure 4 shows a time series of residuals with respect to the regression line. It can be seen that the residuals have a distinct temporal behaviour, including seasonal patterns. Seasonal patterns in the residuals occur when groups I and II each have different seasonal patterns so that they do not cancel each other out.

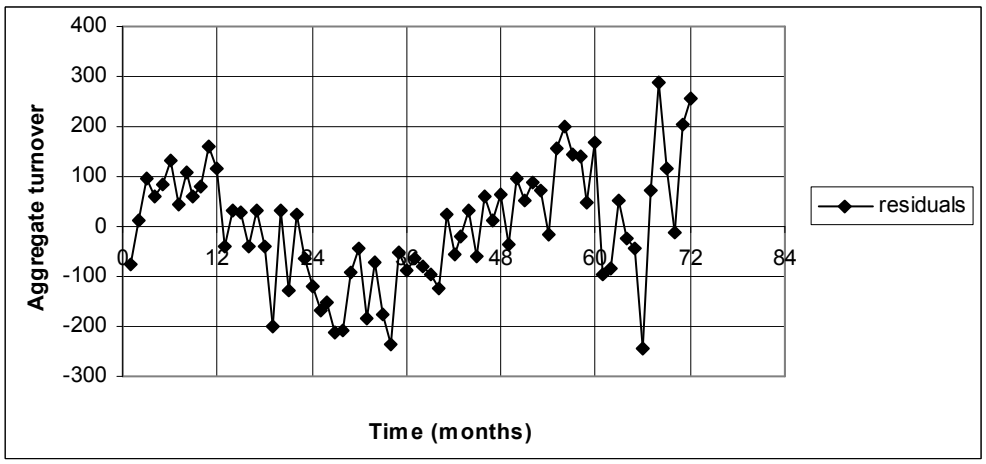

Figure 4. Time series of the residuals with respect to the regression line shown in figure 3. Notice the distinct temporal behaviour.

The method we propose makes use of the time series of groups I and II in order to make a forecast for group II for month t. Besides regression analysis of groups I and II (cf. figure 3), also the temporal behaviour of the residuals is taken into account (cf. figure 4). This is done by modelling the time series according to the regARIMA method:

$$Y_{II}^{t} = \beta_0 + \beta_1 Y_I^t + Z_t, \tag{3}$$

where β_0 and β_1 are the regression coefficients and Z_t are the residuals. The residuals are modelled using so-called ARIMA models (AutoRegressive Integrated Moving Average), which are often used in the analysis of economic time series. ARIMA models describe the time series in terms of seasonal and non-seasonal components by a number of parameters. These parameters, as well as the regression coefficients β_0 and β_1, are estimated in the regARIMA method by Maximum Likelihood Estimation (MLE), using the time series for groups I and II until month t-1. Subsequently, the estimates for β_0, β_1 and the ARIMA parameters, as well as the known value of Y_I^t, are inserted in equation 3 in order to obtain an estimate for Y_{II}^t. Optionally, the presence of outliers in the time series is dealt with (automatically) by adding more regression coefficients. For more details on regARIMA modelling see Bell (1998). A brief overview is given in Appendix B, where also the most relevant terminology is explained. Note that equation 2, concerning the trend-based method, can be derived from equation 3 by letting Z_t=0, β_0=0 and $\beta_1 = Y_{II}^{t-\Delta t} / Y_I^{t-\Delta t}$.

3.4 Imputation of missing enterprise data using aggregate estimates

Given the estimate of the aggregate turnover for group II for month t (Y_{II}^t), as well as the known aggregate turnover Y_{II}^{t-1} for the previous month, the trend Y_{II}^t / Y_{II}^{t-1} for group II from month t-1 to month t is determined. The thus found trend for the aggregate level is then applied to the individual enterprise data y_i^{t-1} for group II for month t-1 in order to determine the imputation values for month t:

$$y_i^t = \frac{Y_{II}^t}{Y_{II}^{t-1}} y_i^{t-1}, \quad i \in II . \tag{4}$$

Note that the trend Y_{II}^t / Y_{II}^{t-1} determined for group II, is generally not the same as the trend Y_I^t / Y_I^{t-1} for group I. Only for the trend-based method (equation 2) with Δt=1 this would necessarily be the case.

3.5 Implementation of the methods

3.5.1 General

A practical complication of the data collection is that turnover values of the enterprises were actually collected by means of observation units. In most cases an enterprise had only one observation unit, but it occurred (on average 0.4%) that an enterprise had more than one. For matters of simplicity in the rest of this paper no explicit distinction is made between the terms enterprise and observation unit. Note, however, that imputations were done on the level of the observation unit and that in case size classes were used to form aggregates, the size classes of the enterprises were passed on to the underlying observation units.

Groups I and II were each time determined based on the availability of enterprise data for month t. A complicating factor here is that some enterprises reported over

periods that were not exactly calendar months, what resulted in situations where data were available for only part of a month. This complication was dealt with by first determining the average turnover per working day for the period such an enterprise reported on, and then multiplying this value with the total number of working days in month t. The thus found value was imputed in order to be used in the assessment of sales development. To distinguish these enterprises from those in groups I and II, they were assigned to a separate group III, which was not used for further forecasting.

3.5.2 Currently used trend-based method

The method currently used at Statistics Netherlands is basically the trend-based method described in section 3.2, with $\Delta t = 1$. The aggregation level used for estimating missing data is based on a subdivision of the manufacturing industry into 124 aggregates, according to 3- and 4-digit SIC divisions.

When for an aggregate not sufficient known data for month t were available to make reasonable estimates of the missing data, known enterprise turnover values of the month before were imputed, taking into account the number of working days in both months. In case no known data were available for an enterprise the month before, the average within the aggregate for that month was used.

Each month, still necessary imputations were redone for all months of the present and previous year, using the at that time most recent information available.

3.5.3 Improved trend-based method

The trend-based method using one earlier month, on which the method currently used at Statistics Netherlands (section 3.5.2) is based, was also implemented with a few differences c.q. improvements with respect to the current method.

Instead of using $\Delta t = 1$ in the imputation of missing data, $\Delta t = 12$ was used. The imputation of missing data for enterprises that do not belong to the intersection (cf. section 2.1) is more problematic in case $\Delta t = 12$ is used, since no data are available for these enterprises at t-12. Therefore, for these enterprises the imputations were done using the trend-based method (section 3.2) with $\Delta t = 1$, in which only enterprises were used that also did not belong to the intersection. In case for a particular enterprise also no data were available for t-1, because the enterprise did not yet exist or did not yet report on that month, a zero turnover value was imputed. Note that enterprises that did not belong to the intersection were not used in the assessment of sales development.

In addition, the aggregates used for imputing could be chosen at will. In the simulations that were performed (section 4), much larger aggregates than the current ones were used. Therefore, in most cases enough data were available to make reasonable estimates of the missing data, and the option of using known or average data for a month before (cf. section 3.5.2) was left out.

Each month, still necessary imputations were redone for the past 12 months, using the at that time most recent information available.

3.5.4 Proposed method using time series analysis

The comments made in section 3.5.3 on the imputation of missing data for enterprises that do no not belong to the intersection and on the use of aggregates, are also valid for the method using time series analysis.

Time series analysis (regARIMA) was done by the X-12-ARIMA program of the U.S. Census Bureau. The X-12-ARIMA specification file that was used is described in Appendix C. A one month forecast was made of the time series of group II. The time series of group I, as well as a time series with a constant value of 1 for each month, were used as regression variables. This last time series was added in order to allow for a nonzero intercept in the linear regression. Additive outliers (AO) and level shifts (LS) were identified. Each time, the best fitting model was chosen automatically out of the 16 ARIMA models (p,d,q)(P,D,Q) with no differencing (d=D=0) and p, q, P and Q zero or one. In case none of the models was accepted by X-12-ARIMA the AUTOMDL spec was left out and modelling was restricted to the regression part and outlier detection.

Also here, the still necessary imputations were redone each month for the past 12 months, using the at that time most recent information available. However, when the turnover of group II was less than 15% of the total turnover, no ARIMA modelling was done because of the then sometimes erratic nature of the time series. So, also in that situation, the AUTOMDL spec was left out and the regARIMA modelling was restricted to the regression part and outlier detection.

4. Simulations

4.1 Assessment of accuracy

In order to determine the accuracy of the developed methods, and to compare them with the method currently used at Statistics Netherlands, a number of simulations was performed. In these simulations accuracy was determined by comparing sales development according to the different methods with sales development according to the final data.

Monthly enterprise data for the manufacturing industry, from January 1993 until July 2003, were available for performing the simulations. Data collected up to 23 September 2003 were used. Approximately 97% of the data, in terms of turnover, was response, whereas the rest was imputed using the current method described in section 3.5.2. These data are referred to as the final data. Importantly, the dates when all the enterprise data were received were known.

The accuracy of the methods was determined for assessment of sales development at a fixed number of days (Δd) after the end of a month. To this purpose, sales development was assessed for the 60 months from January 1998 until December 2002, each time using only the data that had been received until Δd days after the end of the month. Note that for each month the still necessary imputations were redone for the previous months (section 3.5).

The missing (removed) data were imputed using either the method currently in use at Statistics Netherlands (section 3.5.2), the improved trend-based method (section 3.5.3), or the method based on time series analysis (also referred to as the regARIMA method; section 3.5.4). Regarding the improved trend-based method and the method based on regARIMA modelling, simulations were done using the following aggregates for the imputation of missing data:

- The manufacturing industry as a whole (AI1);
- The manufacturing industry subdivided into six aggregates (AI6) defined as the groups of SIC divisions 15/16 (AI6-1), 17-19 (AI6-2), 20/26 (AI6-3), 21/22 (AI6-4), 23-25 (AI6-5) and 27-37 (AI6-6). Note that AI6-6 differs from MI6-6;
- The manufacturing industry subdivided into 30 aggregates (AI30) defined as the combinations of size class (SC) and SIC divisions 15/16 (AI30-1-SC), 17-19 (AI30-2-SC), 20/26 (AI30-3-SC), 21/22 (AI30-4-SC), 23-25 (AI30-5-SC) and 27-37 (AI30-6-SC), with SC 5, 6, 7, 8 or 9 (Appendix A).

Eventually, the sales development (equation 1) according to all methods, at Δd days after the end of a month, was known for 60 months. This sales development was compared on a monthly basis with the sales development according to the final data. Minimum, maximum, average and standard deviation were determined for the differences between the values according to the simulations and according to the final data. This was done for the manufacturing industry as a whole (MI1), as well as for the six subdivisions of the manufacturing industry (MI6).

4.2 Software

The enterprise data were available in a Microsoft SQL-server 2000 database. Stored procedures were written to manipulate data, construct time series, impute data and store intermediate results. Time series analysis (regARIMA), including forecasting, was done by the U.S. Census Bureaus X-12-ARIMA, version 0.2.10 (U.S. Census Bureau, 2002; Findley et al, 1998), which is a DOS program. Microsoft Visual Basic 6.0 was used to run the stored procedures and X-12-ARIMA, as well as to take care of the communication between the database and X-12-ARIMA.

The method currently used by Statistics Netherlands, with which the newly developed methods were compared, was available as a Microsoft Visual Basic 6.0 program using Oracle and Microsoft Access databases.

5. Results

The Dutch manufacturing industry consisted during the years 1998 until 2002 on average of 6,450 enterprises with 20 employees or more. The enterprises were distributed over the aggregates MI6 (section 2.1) as follows: MI6-1 15%, MI6-2 4%, MI6-3 8%, MI6-4 13%, MI6-5 10% and MI6-6 49%. In terms of turnover the distribution is MI6-1 25%, MI6-2 2%, MI6-3 4%, MI6-4 9%, MI6-5 28% and MI6-6 33%. About 10% of the enterprises reported over periods that were not exactly calendar months. Furthermore, on average 91% of the enterprises belonged to the intersection as defined in section 2.1, and these enterprises covered on average 96% of the total turnover of the population. Sales development in the Dutch manufacturing industry varied during the years 1998 until 2002 between -10% and +29%.

The relationship between timeliness and accuracy of the assessment of sales development was determined, in a number of simulations (section 4.1), for the method currently in use at Statistics Netherlands, the improved trend-based method, and the method based on regARIMA modelling. Simulation results are shown in Appendix D. The best aggregation level for the imputation of missing data appeared to be AI30 for the improved trend-based method and AI6 for the method using regARIMA modelling. Results shown below are, if not mentioned explicitly otherwise, based on these aggregation levels.

The standard deviations of the errors are shown in figure 5, as a function of the number of days after the end of the month, for the manufacturing industry as a whole (MI1).

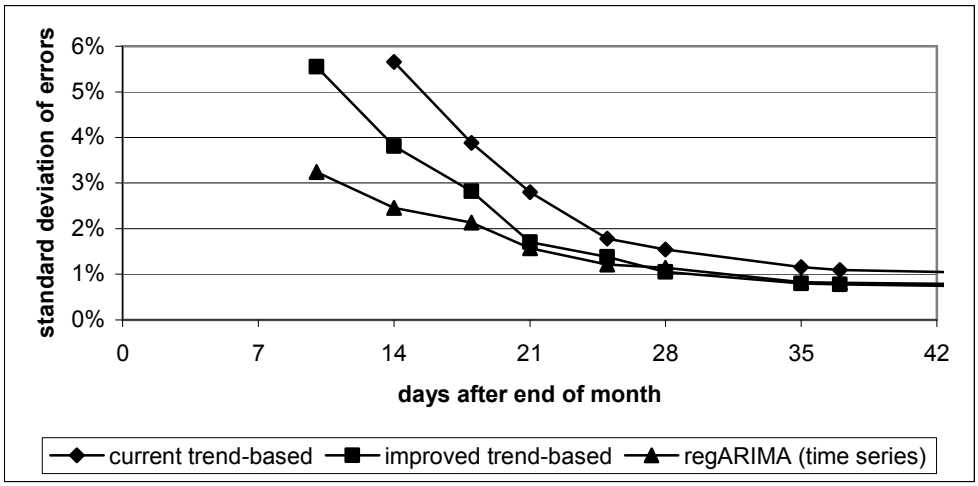

Figure 5. The standard deviation of the errors with respect to the final values is shown for the manufacturing industry as a whole, as a function of the number of days after the end of the month. This is done for the currently used trend-based method, the improved trend-based method and the regARIMA (time series) method. The standard deviation of the errors is a measure of the accuracy of the method to assess sales development.

In table 1 a selection is presented of the most important results shown in appendix D, for the manufacturing industry as a whole (MI1), as well as for the six subdivisions of the manufacturing industry (MI6).

Table 1. The standard deviation of the errors with respect to the final values, for a number of days after the end of the month, regarding the currently used trend-based method, the improved trend-based method and the regARIMA (time series) method. Results are shown for the manufacturing industry as a whole (MI1) as well as for the six subdivisions (MI6) of the manufacturing industry (section 2.1).

		Standard deviation of errors		
Number of days	**Dissemination level**	**Current method**	**Improved trend-based method**	**regARIMA method**
14	MI1	5.67%	3.81%	2.45%
14	MI6-1	3.77%	4.38%	3.12%
14	MI6-2	5.37%	6.48%	3.67%
14	MI6-3	7.15%	4.58%	3.24%
14	MI6-4	5.54%	4.81%	3.73%
14	MI6-5	19.20%	9.90%	5.07%
14	MI6-6	9.55%	5.30%	5.13%
21	MI1	2.80%	1.70%	1.57%
21	MI6-1	3.15%	2.87%	2.37%
21	MI6-2	3.53%	3.24%	2.49%
21	MI6-3	4.02%	2.87%	2.45%
21	MI6-4	4.39%	2.45%	2.43%
21	MI6-5	3.28%	3.67%	2.92%
21	MI6-6	7.19%	3.42%	3.42%
28	MI1	1.54%	1.05%	1.14%
28	MI6-1	2.17%	1.83%	1.69%
28	MI6-2	2.47%	2.51%	1.72%
28	MI6-3	2.74%	2.23%	1.84%
28	MI6-4	2.86%	1.85%	1.90%
28	MI6-5	1.67%	1.94%	1.32%
28	MI6-6	4.17%	2.14%	2.47%
37	MI1	1.09%	0.78%	0.81%
37	MI6-1	1.43%	1.48%	1.24%
37	MI6-2	2.14%	2.19%	1.53%
37	MI6-3	2.36%	1.73%	1.48%
37	MI6-4	1.85%	1.37%	1.46%
37	MI6-5	1.22%	1.23%	1.06%
37	MI6-6	2.97%	1.39%	1.66%
100	MI1	0.57%	0.41%	0.55%
100	MI6-1	0.69%	0.81%	0.66%
100	MI6-2	1.34%	1.41%	1.16%
100	MI6-3	1.05%	0.78%	0.72%
100	MI6-4	0.58%	0.67%	0.60%
100	MI6-5	0.34%	0.51%	0.53%
100	MI6-6	1.62%	0.97%	1.37%

6. Discussion

Regarding the accuracy of the methods, both the improved trend-based method and the regARIMA time series method perform better than the method currently used. For dissemination later than about 27 days after the end of the month, the improved trend-based method performs about equally well as the regARIMA method. However, for earlier dissemination, especially earlier than about 20 days, the loss in accuracy is much less pronounced for the regARIMA method, in comparison with the improved trend-based method (cf. figure 5). Currently data are disseminated at 37±3 days after the end of the month. Using the current method the standard deviation of the errors at 37 days after the end of the month is 1.09%, for the manufacturing industry as a whole. Concerning the six subdivisions (MI6) of the manufacturing industry, standard deviations are larger. Using the regARIMA method data can be disseminated with the same accuracy at around 27 days after the end of the month. Further improving timeliness results in reduced accuracy. At 14 days after the end of the month, for instance, standard deviations are increased to 5.67%, 3.81% and 2.45%, for the current method, the improved trend-based method and the regARIMA method, respectively.

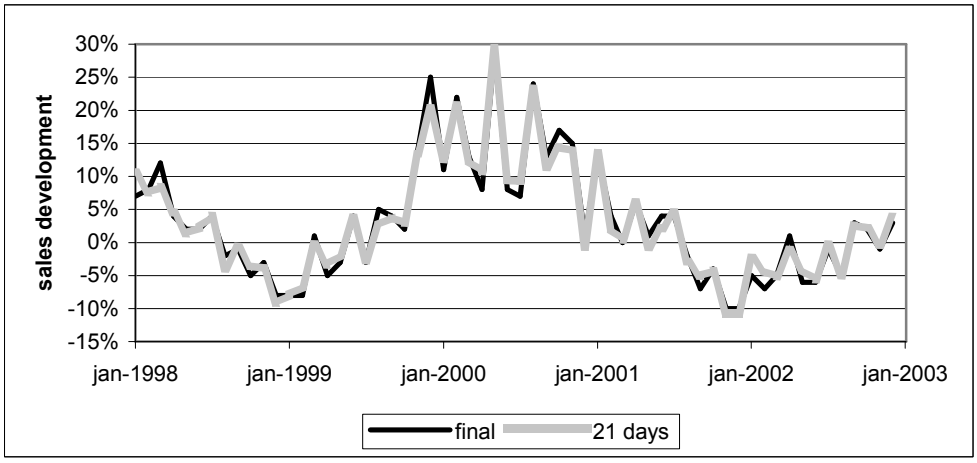

Figure 6. Sales development for the Dutch manufacturing industry as a whole, according to the final data and according to the data available at 21 days after the end of each month. Imputations were done using the regARIMA method.

Besides standard deviations of the errors (cf. figure 5) also average errors and the range of errors (minimum and maximum values) are important in making the decision regarding the trade-off between timeliness and accuracy. To illustrate this, figure 6 shows the sales development in the manufacturing industry as a whole (MI1), according to the final data and according to the data available at 21 days after the end of each month, where imputations were done using the regARIMA method. Minimum, maximum, average and standard deviation of the errors were -4.69% (December 1999), 3.59% (January 1998), -0.08% and 1.57%, respectively. The average error is not significantly different from zero (assuming normally and independently distributed errors). In the figure, it can be seen that the largest error (-4.69%) occurred for a month were the sales development jumped from 14% to 25% (December 1999). At 21 days after the end of the month this jump was

underestimated, but the direction was correct. It may be noted that the response rate regarding turnover at that time was only 31%. The direction of the month-to-month change was in most cases correct at 21 days after the end of the month. The only exceptions occur, expectedly, in situations where the month-to-month changes in sales development are small (July 2000 and February 2002). It should be noted that for interpreting the change from month t-1 to month t, the value for month t-1 at 21 days after the end of month t, is not indicated in the figure. This value is in most cases, however, close to the final value.

In the current survey process automatically imputed data are manually corrected occasionally. When, at a later moment in time, the true values are received the imputed values are overwritten. No history of the original values is kept. This may cause differences between the results of the simulations and the originally disseminated data. The last few years, data were disseminated at 37±3 days after the end of the month by means of press releases. In order to verify the simulation results, the data in the press releases were compared with the final data, and the range, average and standard deviation of the errors were determined. These accuracy measures were compared with simulation results of the current method, 37 days after the end of the month. Differences were generally small and especially the standard deviations were very similar. Summarizing, it can be said that the simulation results for the current method are in agreement with the press releases.

Regarding the imputation of missing values from enterprises in group III (data available for only part of the month) we imputed data for the missing part of the month using average turnover per working day during the part of the month the enterprise did report on. Another way to deal with these enterprises could be to ignore the data received and to treat the enterprises the same as those of group II. Furthermore, it would be possible to impute group III using the average turnover per working day (as is done now) and then to add the enterprises of group III to group I. Both of these options were considered but they did not improve the results.

The improved trend-based method was implemented using $\Delta t = 12$. This gave much better results than $\Delta t = 1$. For instance, 21 days after the end of the month the standard deviation of the errors increases from 1.70% to 2.38% (AI30) when using $\Delta t = 1$ instead of $\Delta t = 12$. This is caused mainly by different seasonal patterns for groups I and II present in the time series, which is not taken into account when using $\Delta t = 1$. Note that the method currently used at Statistics Netherlands uses $\Delta t = 1$. Furthermore, the best aggregation level for the imputation of missing data was found to be AI30 for the trend-based method. These 30 aggregates are composed of 6 SIC divisions and 5 size classes. The method currently used at Statistics Netherlands uses 124 aggregates, based on 3- and 4-digit SIC divisions. Especially for early data dissemination it is better to use larger aggregates than the current method does.

Time series were constructed by summing for each month the turnover values of the enterprises belonging to a certain group (I or II) at month t. Since for months earlier than month t generally not all of these enterprises existed yet, the sums for those months were done over less enterprises. For example, of the enterprises in the

intersection in June 2000, about 30% did not yet exist in January 1993. A way to deal with this problem could be by making use of time series of average enterprise turnover for each month. Using total turnover, however, gave better results than using averages. This is caused by the fact that when enterprises first come into the population they generally do not immediately have the high level of turnover they have for months later in the time series. Using averages would in that case result in more abrupt changes in the time series.

Many different settings were tried in the X-12-ARIMA specification file (Appendix C). Generally the default settings (by the U.S. Census Bureau) gave the best results, but there are a few exceptions to this. First of all, allowing for an intercept in the regression, improved the accuracy of the method and reduced in particular the number of level shift (LS) outliers. In conjunction with this, it can be assumed that the time series of residuals with respect to the regression line are more or less stationary, so that no differencing is needed (d=D=0). The autoregressive and moving average parameters p, q, P and Q were allowed to have all possible combinations of 0 and 1. Sporadically, none of the models was found to be acceptable and ARIMA modelling of the regression residuals was left out. Allowing values of 2 for p, q, P and Q did not improve results. Manual adjustment of the specification file for each specific time series may further improve the accuracy of the method, although in general the specification file presented in Appendix C proved to be quite satisfactory.

When the turnover of group II was below a certain percentage of the total turnover, and the time series were more erratic, only regression and outlier detection were done and ARIMA modelling was left out. This percentage was set to 15% but is not very critical. However, setting the percentage to 0% (always ARIMA modelling) or 100% (never ARIMA modelling) gave worse results. In most simulations a percentage of 15% meant that only for the last two or three months the imputations were (re)done including ARIMA modelling.

During the time this study was done Statistics Netherlands changed the current survey process described in section 3.5.2. The new survey process is part of the system for the so-called Implementation of the Economic Transformation Process (IMPECT). For practical reasons no simulations could be done yet with this system. The main difference, regarding the imputation of missing values, between the IMPECT system and the system described in section 3.5.2 is that when imputation aggregates are too small for accurate estimates, they are automatically enlarged. However, the main conclusions from this study, regarding trend-based methods, still hold. In case of early dissemination it is better to take larger aggregates (AI30) from the start, and it is important to use $\Delta t = 12$ instead of $\Delta t = 1$. In the IMPECT system $\Delta t = 1$ is still used, and results are expected to be similar to the method described in section 3.5.2.

7. Conclusions

A new method, based on time series analysis, was developed to improve the imputation of missing data in the monthly assessment of sales development in the Dutch manufacturing industry. Besides already available data for the month in question, time series of historical enterprise data were used. The timeliness of this statistics can be improved from 37±3 days currently to about 27 days with the newly developed method, at the same level of accuracy. Further improving timeliness results in reduced accuracy. The final decision about the moment in time to disseminate data has to be made in consultation with the most important users of the statistics. To aid making this decision, the relationship between timeliness and accuracy was investigated (cf. figure 5). Furthermore, it has to be considered in case of early dissemination, whether or not to include data on any subdivisions of the manufacturing industry.

In addition, it was shown that the current trend-based method can also be improved considerably by simply using larger aggregates for the imputation of missing data, and determining trends with respect to the same month the year before instead of with respect to the previous month. This improved version of the current method performs, for dissemination later than about 27 days after the end of the month, similar to the method based on time series analysis. Implementation of the time series method, however, will require more substantial modifications of the survey process. Therefore, if it will be decided to disseminate data later than about 27 days after the end of the month, for practical reasons it is probably best to use the improved version of the current method. However, for earlier dissemination the method based on time series analysis is preferred. Especially for dissemination earlier than about 20 days after the end of the month, the loss in accuracy is much less pronounced for the method based on time series analysis, in comparison with the trend-based methods (cf. figure 5).

The time series method developed in this study focussed on the monthly statistics of sales development in the Dutch manufacturing industry. This statistics is based on an integral collection of data from enterprises with more than 20 employees. Applying the same method to other statistics that are based on integral data collection seems straightforward, although differences in population characteristics (e.g. heterogeneity) may have to be taken into consideration. The method will need to be modified for sample-based statistics for which no time series are available for all individual enterprises.

8. References

Bell, W.R. (1998). "An overview of regARIMA modelling". Research report. Statistical Research Division. U.S. Census Bureau.

Deville, J.C. et al (2002). "Final report of the expert group on sampling for timely European indicators". Eurostat. Unit A-4.

Findley, D.F. et al (1998). "New capabilities and methods of the X-12-ARIMA seasonal adjustment program". Journal of Business and Economic Statistics. 16, 127-176.

Revilla, P. (2002). "An E&I method based on time series modelling designed to improve timeliness". Conference of European statisticians. UNECE Work session on statistical data editing (27-29 May 2002, Helsinki, Finland).

U.S. Census Bureau (2002). "X-12-ARIMA Reference Manual version 0.2.10". http://www.census.gov/srd/www/x12a.

Appendix A. Overview of SIC divisions and size classes

Below an overview is given of the SIC divisions within the manufacturing industry, according to the Standard Industrial Classification of all Economic Activities 1993.

SIC	Description
15	Manufacture of food products and beverages
16	Manufacture of tobacco products
17	Manufacture of textiles
18	Manufacture of wearing apparel; dressing and dyeing of fur
19	Tanning and dressing of leather; manufacture of luggage, handbags, saddlery, harness and footwear
20	Manufacture of wood and of products of wood and cork, except furniture; manufacture of articles of straw and plaiting materials
21	Manufacture of pulp, paper and paper products
22	Publishing, printing and reproduction of recorded media
23	Manufacture of coke, refined petroleum products and nuclear fuel
24	Manufacture of chemicals and chemical products
25	Manufacture of rubber and plastic products
26	Manufacture of other non-metallic mineral products
27	Manufacture of basic metals
28	Manufacture of fabricated metal products, except machinery and equipment
29	Manufacture of machinery and equipment n.e.c.
30	Manufacture of office machinery and computers
31	Manufacture of electrical machinery and apparatus n.e.c.
32	Manufacture of radio, television and communication equipment and apparatus
33	Manufacture of medical, precision and optical instruments, watches and clocks
34	Manufacture of motor vehicles, trailers and semi-trailers
35	Manufacture of other transport equipment
36	Manufacture of furniture; manufacturing n.e.c.
37	Preparation for recycling

Furthermore, the enterprises are classified according to the following classes of size.

Size class	Number of employees
5	20-49
6	50-99
7	100-199
8	200-499
9	500 or more

Appendix B. regARIMA modelling in X-12-ARIMA

In this Appendix regARIMA modelling, as it is done in the X-12-ARIMA computer program of the U.S. Census Bureau, is introduced briefly. For more details on regARIMA modelling see Bell (1998) or Findley et al (1998).

The regARIMA method models discrete time series in the following way:

$$Y_t = \sum_{i=1}^{r} \beta_i X_{it} + Z_t, \tag{A1}$$

where Y_t is the time series in question, X_{it} time series of regression variables with coefficients β_t, and Z_t the time series of residuals from the regression. The regression residuals in turn are assumed to follow an ARIMA model.

ARIMA models, which are often used to describe economic time series, consist of an AutoRegressive (AR), an Integrated (I) and a Moving Average (MA) term. The autoregressive term relates Z_t to previous values of itself, while the moving average term smoothes buffeting effects from unpredictable events (ε_t). The integrated term is added in order to deal with non-stationary time series, i.e. time series whose mean and/or autocorrelations are not constant over time. Each of the above-mentioned terms has a seasonal as well as a non-seasonal part. Mathematically, ARIMA models have the following form. Let B denote the backshift operator, $BZ_t = Z_{t-1}$. Then,

$$\phi_p(B)\Phi_P(B^s)(1-B)^d(1-B^s)^D Z_t = \theta_q(B)\Theta_Q(B^s)\varepsilon_t, \tag{A2}$$

where s is the length of the seasonal period (12 in case of monthly series) and ε_t a sequence of independent variables with mean 0 and constant variance. The terms ϕ_p (non-seasonal AR), Φ_P (seasonal AR), θ_q (non-seasonal MA) and Θ_Q (seasonal MA) have the form of polynomials with degrees p, P, q and Q, respectively. For example, if $p \geq 1$, then $\phi_p(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$, and if $P \geq 1$, then $\Phi_P(B^s) = 1 - \Phi_1 B^s - \cdots - \Phi_P B^{sP}$. Furthermore, d and D denote the number of times differencing (the I in ARIMA refers to integrated, the inverse of differencing) is done, non-seasonally and seasonally respectively. A particular ARIMA model is generally referred to as a model with order (p,d,q)(P,D,Q).

Outliers in the time series are described using the regression variables. Three different kinds of outliers are dealt with in X-12-ARIMA: additive outliers (AO), level shifts (LS) and temporary changes (TC). Additive outliers have an atypical value at one period, level shifts have an effect that lasts for the rest of the time series and temporary changes have an effect that lasts a few periods.

X-12-ARIMA estimates regARIMA model parameters for a given set of regression variables and an ARIMA model with a given order (p,d,q)(P,D,Q) by Maximum Likelihood Estimation. The program chooses from a set of user-defined ARIMA models with different orders (p,d,q)(P,D,Q) which one fits the regression residuals best. Outliers can be identified automatically by iteratively adding and removing respective regression variables.

Appendix C. X-12-ARIMA specification file

In this study the X-12-ARIMA program, release version 0.2.10 (U.S. Census Bureau, 2002), was used with the following specification file. Only those settings are shown that are different from default.

```
SERIES{
 File = "X:\timeseries2.dat"
 Format = "DateValue"
}
OUTLIER{
}
REGRESSION{
 User = (slope intercept)
 File = "X:\timeseries1.dat"
 Format = "DateValue"
}
AUTOMDL{
 File = "X:\X12A_16.mdl"
 Method = best
 Identify = all
}
FORECAST{
 Maxlead = 1
}
```

The file timeseries2.dat contains the time series of group II and the file timeseries1.dat contains the time series of group I as well as an extra time series with a constant value of 1 for each month. Furthermore, the file X12A_16.mdl consists of the following models (p,d,q)(P,D,Q).

(p,d,q)	(P,D,Q)
(0,0,0)	(0,0,0)
(0,0,0)	(0,0,1)
(0,0,0)	(1,0,0)
(0,0,0)	(1,0,1)
(0,0,1)	(0,0,0)
(0,0,1)	(0,0,1)
(0,0,1)	(1,0,0)
(0,0,1)	(1,0,1)
(1,0,0)	(0,0,0)
(1,0,0)	(0,0,1)
(1,0,0)	(1,0,0)
(1,0,0)	(1,0,1)
(1,0,1)	(0,0,0)
(1,0,1)	(0,0,1)
(1,0,1)	(1,0,0)
(1,0,1)	(1,0,1)

Appendix D. Results of simulations

Below, simulation results are shown, for a number of days after the end of the month, for the current method, the improved trend-based method and the regARIMA time series method. Different aggregates (AI) were used for the imputation of missing values. Results are shown for the total manufacturing industry (MI1) and for the six subdivisions (MI6), regarding minimum, maximum, average and standard deviation of the errors in assessment of sales development with respect to the final values.

days	method	AI	MI	minimum	maximum	average	st. dev.
10	regARIMA	6	MI1	-11.91%	5.84%	-1.00%	3.24%
10	regARIMA	6	MI6-1	-12.16%	6.41%	-1.01%	3.41%
10	regARIMA	6	MI6-2	-8.95%	7.26%	-0.62%	3.46%
10	regARIMA	6	MI6-3	-16.75%	11.80%	-0.61%	5.08%
10	regARIMA	6	MI6-4	-10.53%	10.46%	-0.66%	4.27%
10	regARIMA	6	MI6-5	-23.04%	19.37%	-1.18%	7.17%
10	regARIMA	6	MI6-6	-23.49%	15.97%	-1.03%	6.28%
10	improved trend	1	MI1	-18.15%	6.64%	-1.96%	5.57%
10	improved trend	1	MI6-1	-15.68%	10.95%	-1.84%	6.20%
10	improved trend	1	MI6-2	-25.14%	19.99%	-0.86%	7.79%
10	improved trend	1	MI6-3	-25.95%	26.55%	-0.91%	10.51%
10	improved trend	1	MI6-4	-16.43%	10.13%	-1.40%	5.85%
10	improved trend	1	MI6-5	-26.27%	13.68%	-2.99%	8.60%
10	improved trend	1	MI6-6	-26.49%	16.65%	-1.26%	8.67%
10	improved trend	30	MI1	-18.78%	7.29%	-2.31%	5.55%
10	improved trend	30	MI6-1	-18.17%	21.74%	-0.90%	6.71%
10	improved trend	30	MI6-2	-25.76%	9.67%	-1.16%	6.42%
10	improved trend	30	MI6-3	-22.85%	17.84%	-0.53%	6.20%
10	improved trend	30	MI6-4	-27.01%	28.14%	-1.92%	8.20%
10	improved trend	30	MI6-5	-45.83%	35.22%	-3.60%	17.25%
10	improved trend	30	MI6-6	-21.75%	8.89%	-3.17%	7.67%
14	current	124	MI1	-7.54%	31.01%	2.04%	5.66%
14	current	124	MI6-1	-11.54%	11.30%	0.33%	3.77%
14	current	124	MI6-2	-8.13%	20.01%	2.08%	5.37%
14	current	124	MI6-3	-17.83%	29.97%	1.55%	7.15%
14	current	124	MI6-4	-12.16%	13.34%	-0.93%	5.54%
14	current	124	MI6-5	-11.33%	140.36%	3.73%	19.20%
14	current	124	MI6-6	-13.69%	27.74%	3.01%	9.55%
14	regARIMA	1	MI1	-9.66%	6.46%	0.04%	2.75%
14	regARIMA	1	MI6-1	-7.34%	7.90%	-0.23%	3.54%
14	regARIMA	1	MI6-2	-13.19%	14.31%	0.67%	5.76%
14	regARIMA	1	MI6-3	-12.47%	27.43%	0.92%	8.30%
14	regARIMA	1	MI6-4	-7.95%	10.41%	0.37%	3.57%
14	regARIMA	1	MI6-5	-16.74%	11.15%	-0.58%	4.90%
14	regARIMA	1	MI6-6	-20.61%	23.78%	0.84%	7.75%
14	regARIMA	6	MI1	-9.65%	4.00%	-0.42%	2.45%
14	regARIMA	6	MI6-1	-7.59%	12.33%	-0.31%	3.12%
14	regARIMA	6	MI6-2	-10.01%	11.59%	0.38%	3.67%
14	regARIMA	6	MI6-3	-8.28%	7.63%	-0.37%	3.24%
14	regARIMA	6	MI6-4	-9.78%	11.62%	-0.12%	3.73%
14	regARIMA	6	MI6-5	-14.00%	15.21%	-0.07%	5.07%
14	regARIMA	6	MI6-6	-18.30%	11.04%	-0.81%	5.13%
14	regARIMA	30	MI1	-11.93%	6.23%	-0.48%	3.32%

days	method	AI	MI	minimum	maximum	average	st. dev.
14	regARIMA	30	MI6-1	-15.37%	6.32%	-0.76%	3.86%
14	regARIMA	30	MI6-2	-12.74%	9.46%	0.15%	3.78%
14	regARIMA	30	MI6-3	-7.14%	9.93%	-0.34%	3.16%
14	regARIMA	30	MI6-4	-11.58%	8.34%	-0.33%	4.22%
14	regARIMA	30	MI6-5	-22.66%	10.95%	-0.68%	5.79%
14	regARIMA	30	MI6-6	-20.99%	13.96%	-0.18%	5.47%
14	improved trend	1	MI1	-12.37%	8.20%	-0.08%	4.07%
14	improved trend	1	MI6-1	-12.89%	9.59%	-0.26%	4.98%
14	improved trend	1	MI6-2	-16.14%	15.06%	0.67%	6.77%
14	improved trend	1	MI6-3	-16.24%	27.76%	0.84%	8.78%
14	improved trend	1	MI6-4	-11.51%	8.73%	0.29%	4.26%
14	improved trend	1	MI6-5	-18.79%	15.81%	-0.77%	6.78%
14	improved trend	1	MI6-6	-18.24%	22.54%	0.69%	7.65%
14	improved trend	6	MI1	-11.70%	7.23%	-0.42%	3.91%
14	improved trend	6	MI6-1	-8.83%	15.03%	0.00%	3.61%
14	improved trend	6	MI6-2	-27.96%	13.89%	-0.41%	6.26%
14	improved trend	6	MI6-3	-9.38%	19.94%	0.12%	4.94%
14	improved trend	6	MI6-4	-15.98%	10.59%	-0.84%	4.14%
14	improved trend	6	MI6-5	-29.92%	17.02%	-0.30%	10.16%
14	improved trend	6	MI6-6	-15.92%	13.17%	-1.04%	5.36%
14	improved trend	30	MI1	-10.29%	7.11%	-0.50%	3.81%
14	improved trend	30	MI6-1	-13.52%	10.64%	-0.38%	4.38%
14	improved trend	30	MI6-2	-25.76%	14.80%	-0.33%	6.48%
14	improved trend	30	MI6-3	-8.79%	14.38%	-0.09%	4.58%
14	improved trend	30	MI6-4	-12.94%	16.52%	-0.54%	4.81%
14	improved trend	30	MI6-5	-27.93%	25.51%	-0.32%	9.90%
14	improved trend	30	MI6-6	-13.97%	13.32%	-1.04%	5.30%
18	current	124	MI1	-5.42%	12.03%	1.70%	3.88%
18	current	124	MI6-1	-4.44%	11.43%	0.79%	3.05%
18	current	124	MI6-2	-5.79%	11.87%	1.70%	3.96%
18	current	124	MI6-3	-18.01%	18.99%	1.18%	5.23%
18	current	124	MI6-4	-13.03%	19.17%	0.10%	5.39%
18	current	124	MI6-5	-8.32%	13.14%	0.74%	4.56%
18	current	124	MI6-6	-12.42%	33.25%	3.81%	9.63%
18	regARIMA	6	MI1	-6.98%	4.07%	-0.44%	2.13%
18	regARIMA	6	MI6-1	-6.26%	8.50%	0.24%	2.45%
18	regARIMA	6	MI6-2	-9.14%	8.28%	0.31%	2.90%
18	regARIMA	6	MI6-3	-8.34%	5.92%	-0.08%	2.73%
18	regARIMA	6	MI6-4	-10.28%	7.85%	0.04%	3.07%
18	regARIMA	6	MI6-5	-11.30%	7.87%	-0.83%	3.76%
18	regARIMA	6	MI6-6	-12.62%	8.96%	-0.84%	4.30%
18	improved trend	30	MI1	-6.61%	7.68%	-0.17%	2.82%
18	improved trend	30	MI6-1	-7.19%	9.51%	0.21%	3.29%
18	improved trend	30	MI6-2	-9.46%	8.89%	-0.01%	3.89%
18	improved trend	30	MI6-3	-10.66%	6.01%	-0.43%	3.30%
18	improved trend	30	MI6-4	-8.95%	7.42%	-0.37%	3.22%
18	improved trend	30	MI6-5	-17.75%	9.41%	-0.47%	5.95%
18	improved trend	30	MI6-6	-8.88%	19.94%	-0.40%	4.91%
21	current	124	MI1	-5.48%	9.28%	1.18%	2.80%
21	current	124	MI6-1	-5.86%	11.65%	0.75%	3.15%
21	current	124	MI6-2	-6.70%	8.86%	1.40%	3.53%
21	current	124	MI6-3	-13.81%	15.30%	0.67%	4.02%
21	current	124	MI6-4	-9.37%	13.71%	0.21%	4.39%
21	current	124	MI6-5	-7.83%	11.78%	0.51%	3.28%

days	method	AI	MI	minimum	maximum	average	st. dev.
21	current	124	MI6-6	-12.16%	26.60%	2.58%	7.19%
21	regARIMA	1	MI1	-5.35%	3.25%	0.17%	1.64%
21	regARIMA	1	MI6-1	-10.47%	5.99%	-0.15%	2.92%
21	regARIMA	1	MI6-2	-11.35%	13.01%	0.56%	4.92%
21	regARIMA	1	MI6-3	-7.41%	21.14%	0.83%	5.91%
21	regARIMA	1	MI6-4	-7.13%	7.37%	0.43%	2.77%
21	regARIMA	1	MI6-5	-6.29%	9.15%	-0.24%	2.87%
21	regARIMA	1	MI6-6	-13.18%	16.72%	0.80%	5.28%
21	regARIMA	6	MI1	-4.69%	3.59%	-0.08%	1.57%
21	regARIMA	6	MI6-1	-8.23%	4.67%	-0.04%	2.37%
21	regARIMA	6	MI6-2	-6.75%	8.32%	0.50%	2.49%
21	regARIMA	6	MI6-3	-7.38%	7.58%	-0.32%	2.45%
21	regARIMA	6	MI6-4	-5.90%	7.57%	0.11%	2.43%
21	regARIMA	6	MI6-5	-10.07%	7.04%	0.30%	2.92%
21	regARIMA	6	MI6-6	-9.92%	9.89%	-0.46%	3.42%
21	improved trend	6	MI1	-6.04%	4.04%	0.18%	1.75%
21	improved trend	6	MI6-1	-6.01%	7.27%	0.31%	2.78%
21	improved trend	6	MI6-2	-7.56%	9.35%	0.67%	3.09%
21	improved trend	6	MI6-3	-9.25%	5.82%	-0.39%	2.73%
21	improved trend	6	MI6-4	-8.01%	4.48%	-0.41%	2.29%
21	improved trend	6	MI6-5	-9.44%	8.93%	0.86%	3.81%
21	improved trend	6	MI6-6	-10.13%	6.56%	-0.39%	3.67%
21	improved trend	30	MI1	-6.43%	3.48%	0.04%	1.70%
21	improved trend	30	MI6-1	-6.15%	6.94%	0.18%	2.87%
21	improved trend	30	MI6-2	-7.39%	9.21%	0.54%	3.24%
21	improved trend	30	MI6-3	-9.11%	6.47%	-0.36%	2.87%
21	improved trend	30	MI6-4	-8.41%	4.54%	-0.47%	2.45%
21	improved trend	30	MI6-5	-10.11%	8.59%	0.55%	3.67%
21	improved trend	30	MI6-6	-8.44%	6.10%	-0.47%	3.42%
25	current	124	MI1	-3.32%	5.43%	0.78%	1.78%
25	current	124	MI6-1	-5.32%	10.08%	0.18%	2.50%
25	current	124	MI6-2	-9.30%	7.31%	1.31%	2.93%
25	current	124	MI6-3	-11.76%	14.68%	0.42%	3.61%
25	current	124	MI6-4	-5.22%	12.17%	0.64%	3.26%
25	current	124	MI6-5	-6.90%	6.60%	0.48%	2.27%
25	current	124	MI6-6	-8.92%	14.58%	1.64%	4.34%
25	regARIMA	1	MI1	-1.90%	5.79%	0.48%	1.42%
25	regARIMA	1	MI6-1	-8.29%	4.87%	0.10%	2.34%
25	regARIMA	1	MI6-2	-7.71%	11.48%	0.85%	4.09%
25	regARIMA	1	MI6-3	-7.50%	24.43%	1.00%	5.64%
25	regARIMA	1	MI6-4	-2.66%	7.70%	0.96%	2.62%
25	regARIMA	1	MI6-5	-6.19%	7.26%	0.24%	2.09%
25	regARIMA	1	MI6-6	-7.23%	16.93%	0.94%	3.96%
25	regARIMA	6	MI1	-2.86%	3.63%	0.12%	1.21%
25	regARIMA	6	MI6-1	-4.14%	5.09%	-0.08%	1.96%
25	regARIMA	6	MI6-2	-6.49%	6.98%	0.61%	2.23%
25	regARIMA	6	MI6-3	-5.56%	7.72%	-0.20%	2.26%
25	regARIMA	6	MI6-4	-7.30%	5.75%	0.45%	2.29%
25	regARIMA	6	MI6-5	-5.91%	7.07%	0.59%	1.96%
25	regARIMA	6	MI6-6	-8.85%	8.63%	-0.18%	2.70%
25	improved trend	6	MI1	-2.42%	4.35%	0.28%	1.40%
25	improved trend	6	MI6-1	-5.88%	6.01%	0.08%	2.40%
25	improved trend	6	MI6-2	-7.45%	8.70%	0.52%	2.92%
25	improved trend	6	MI6-3	-4.68%	7.66%	-0.20%	2.24%

days	method	AI	MI	minimum	maximum	average	st. dev.
25	improved trend	6	MI6-4	-7.52%	6.14%	-0.10%	2.25%
25	improved trend	6	MI6-5	-8.37%	11.94%	0.94%	3.45%
25	improved trend	6	MI6-6	-10.09%	6.93%	0.03%	2.88%
25	improved trend	30	MI1	-3.43%	3.60%	0.18%	1.38%
25	improved trend	30	MI6-1	-6.06%	5.62%	0.01%	2.27%
25	improved trend	30	MI6-2	-7.02%	8.43%	0.46%	2.95%
25	improved trend	30	MI6-3	-4.38%	7.76%	-0.14%	2.38%
25	improved trend	30	MI6-4	-7.78%	6.07%	-0.08%	2.31%
25	improved trend	30	MI6-5	-9.34%	7.36%	0.74%	3.09%
25	improved trend	30	MI6-6	-11.26%	6.88%	-0.13%	2.80%
26	regARIMA	6	MI1	-2.86%	4.11%	0.08%	1.18%
26	regARIMA	6	MI6-1	-4.14%	4.22%	-0.03%	1.81%
26	regARIMA	6	MI6-2	-4.74%	7.06%	0.48%	1.92%
26	regARIMA	6	MI6-3	-5.21%	2.71%	-0.36%	1.93%
26	regARIMA	6	MI6-4	-6.07%	5.75%	0.61%	2.08%
26	regARIMA	6	MI6-5	-5.85%	7.07%	0.35%	1.96%
26	regARIMA	6	MI6-6	-8.40%	7.53%	-0.15%	2.69%
27	regARIMA	6	MI1	-2.45%	3.83%	0.04%	1.10%
27	regARIMA	6	MI6-1	-3.92%	3.84%	-0.10%	1.76%
27	regARIMA	6	MI6-2	-3.26%	4.26%	0.49%	1.72%
27	regARIMA	6	MI6-3	-4.65%	3.05%	-0.30%	1.84%
27	regARIMA	6	MI6-4	-6.61%	5.75%	0.49%	2.01%
27	regARIMA	6	MI6-5	-6.11%	7.07%	0.33%	1.86%
27	regARIMA	6	MI6-6	-5.96%	6.86%	-0.18%	2.48%
28	current	124	MI1	-2.52%	4.74%	0.66%	1.54%
28	current	124	MI6-1	-4.32%	10.36%	0.22%	2.17%
28	current	124	MI6-2	-9.78%	4.98%	1.09%	2.47%
28	current	124	MI6-3	-8.31%	6.28%	0.32%	2.74%
28	current	124	MI6-4	-5.49%	10.30%	0.67%	2.86%
28	current	124	MI6-5	-3.11%	6.70%	0.41%	1.67%
28	current	124	MI6-6	-8.94%	12.97%	1.31%	4.17%
28	regARIMA	1	MI1	-1.95%	4.51%	0.40%	1.34%
28	regARIMA	1	MI6-1	-8.91%	4.78%	0.04%	2.18%
28	regARIMA	1	MI6-2	-6.89%	11.74%	0.65%	3.46%
28	regARIMA	1	MI6-3	-5.46%	19.73%	0.74%	4.46%
28	regARIMA	1	MI6-4	-3.12%	6.29%	0.78%	2.26%
28	regARIMA	1	MI6-5	-4.77%	4.47%	0.17%	1.67%
28	regARIMA	1	MI6-6	-5.93%	13.86%	0.84%	3.62%
28	regARIMA	6	MI1	-3.08%	4.20%	0.11%	1.14%
28	regARIMA	6	MI6-1	-4.66%	3.37%	-0.14%	1.69%
28	regARIMA	6	MI6-2	-3.24%	5.57%	0.43%	1.72%
28	regARIMA	6	MI6-3	-4.52%	4.69%	-0.18%	1.84%
28	regARIMA	6	MI6-4	-5.19%	5.75%	0.60%	1.90%
28	regARIMA	6	MI6-5	-3.14%	4.13%	0.39%	1.32%
28	regARIMA	6	MI6-6	-6.53%	7.12%	-0.04%	2.47%
28	improved trend	6	MI1	-2.08%	5.41%	0.34%	1.28%
28	improved trend	6	MI6-1	-4.25%	4.07%	0.04%	1.93%
28	improved trend	6	MI6-2	-5.66%	8.07%	0.19%	2.42%
28	improved trend	6	MI6-3	-4.59%	5.88%	-0.19%	2.07%
28	improved trend	6	MI6-4	-5.42%	5.89%	0.11%	1.87%
28	improved trend	6	MI6-5	-3.69%	8.45%	1.12%	2.66%
28	improved trend	6	MI6-6	-4.93%	7.56%	0.06%	2.34%
28	improved trend	30	MI1	-2.64%	4.31%	0.21%	1.05%
28	improved trend	30	MI6-1	-4.34%	4.24%	0.00%	1.83%

days	method	AI	MI	minimum	maximum	average	st. dev.
28	improved trend	30	MI6-2	-5.96%	8.67%	0.15%	2.51%
28	improved trend	30	MI6-3	-5.24%	6.81%	-0.12%	2.23%
28	improved trend	30	MI6-4	-5.44%	5.61%	0.14%	1.85%
28	improved trend	30	MI6-5	-4.10%	5.59%	0.76%	1.94%
28	improved trend	30	MI6-6	-5.16%	6.21%	-0.07%	2.14%
35	current	124	MI1	-1.83%	3.82%	0.41%	1.16%
35	current	124	MI6-1	-3.67%	4.74%	-0.07%	1.49%
35	current	124	MI6-2	-8.33%	5.37%	0.75%	2.31%
35	current	124	MI6-3	-6.56%	6.12%	0.57%	2.39%
35	current	124	MI6-4	-8.58%	5.39%	0.45%	2.00%
35	current	124	MI6-5	-2.92%	5.80%	0.33%	1.35%
35	current	124	MI6-6	-5.88%	12.34%	0.81%	3.15%
35	regARIMA	1	MI1	-1.89%	4.35%	0.34%	1.17%
35	regARIMA	1	MI6-1	-3.72%	4.60%	0.01%	1.69%
35	regARIMA	1	MI6-2	-5.21%	7.37%	0.47%	2.40%
35	regARIMA	1	MI6-3	-4.10%	8.51%	0.45%	2.87%
35	regARIMA	1	MI6-4	-2.84%	4.51%	0.67%	1.57%
35	regARIMA	1	MI6-5	-2.71%	3.99%	0.13%	1.34%
35	regARIMA	1	MI6-6	-5.21%	9.18%	0.71%	2.73%
35	regARIMA	6	MI1	-1.31%	3.05%	0.19%	0.83%
35	regARIMA	6	MI6-1	-3.31%	3.06%	0.07%	1.28%
35	regARIMA	6	MI6-2	-2.96%	3.60%	0.35%	1.65%
35	regARIMA	6	MI6-3	-3.74%	2.68%	0.00%	1.43%
35	regARIMA	6	MI6-4	-3.30%	4.64%	0.64%	1.48%
35	regARIMA	6	MI6-5	-2.04%	3.02%	0.22%	1.13%
35	regARIMA	6	MI6-6	-3.04%	4.96%	0.18%	1.68%
35	improved trend	6	MI1	-1.42%	3.75%	0.30%	1.00%
35	improved trend	6	MI6-1	-3.78%	3.67%	0.11%	1.56%
35	improved trend	6	MI6-2	-4.38%	6.22%	0.19%	2.32%
35	improved trend	6	MI6-3	-4.31%	4.53%	0.01%	1.69%
35	improved trend	6	MI6-4	-3.99%	3.90%	0.37%	1.45%
35	improved trend	6	MI6-5	-1.92%	5.54%	0.67%	1.94%
35	improved trend	6	MI6-6	-3.06%	4.97%	0.24%	1.69%
35	improved trend	30	MI1	-1.13%	3.04%	0.25%	0.80%
35	improved trend	30	MI6-1	-3.86%	3.72%	0.10%	1.51%
35	improved trend	30	MI6-2	-4.70%	6.54%	0.21%	2.42%
35	improved trend	30	MI6-3	-5.40%	4.72%	0.07%	1.78%
35	improved trend	30	MI6-4	-3.96%	3.96%	0.39%	1.42%
35	improved trend	30	MI6-5	-1.64%	2.88%	0.47%	1.22%
35	improved trend	30	MI6-6	-3.04%	4.77%	0.16%	1.53%
37	current	124	MI1	-1.67%	3.67%	0.38%	1.09%
37	current	124	MI6-1	-3.76%	4.76%	-0.10%	1.43%
37	current	124	MI6-2	-8.33%	4.72%	0.77%	2.14%
37	current	124	MI6-3	-6.56%	6.08%	0.57%	2.36%
37	current	124	MI6-4	-8.51%	4.92%	0.35%	1.85%
37	current	124	MI6-5	-1.92%	5.75%	0.38%	1.22%
37	current	124	MI6-6	-5.35%	12.02%	0.75%	2.97%
37	regARIMA	1	MI1	-2.11%	4.35%	0.31%	1.18%
37	regARIMA	1	MI6-1	-4.02%	5.08%	-0.02%	1.69%
37	regARIMA	1	MI6-2	-5.02%	7.21%	0.40%	2.34%
37	regARIMA	1	MI6-3	-4.05%	8.34%	0.43%	2.73%
37	regARIMA	1	MI6-4	-2.73%	4.13%	0.57%	1.54%
37	regARIMA	1	MI6-5	-2.41%	3.20%	0.12%	1.26%
37	regARIMA	1	MI6-6	-5.23%	8.59%	0.68%	2.67%

days	method	AI	MI	minimum	maximum	average	st. dev.
37	regARIMA	6	MI1	-1.44%	3.03%	0.12%	0.81%
37	regARIMA	6	MI6-1	-3.05%	2.89%	0.04%	1.24%
37	regARIMA	6	MI6-2	-2.90%	3.39%	0.32%	1.53%
37	regARIMA	6	MI6-3	-4.46%	2.71%	-0.06%	1.48%
37	regARIMA	6	MI6-4	-3.31%	4.24%	0.53%	1.46%
37	regARIMA	6	MI6-5	-1.80%	2.97%	0.23%	1.06%
37	regARIMA	6	MI6-6	-3.01%	4.90%	0.01%	1.66%
37	regARIMA	30	MI1	-2.15%	2.63%	0.16%	0.84%
37	regARIMA	30	MI6-1	-3.43%	2.61%	0.02%	1.17%
37	regARIMA	30	MI6-2	-3.57%	3.61%	0.26%	1.57%
37	regARIMA	30	MI6-3	-4.80%	4.32%	0.00%	1.53%
37	regARIMA	30	MI6-4	-4.76%	4.07%	0.39%	1.46%
37	regARIMA	30	MI6-5	-1.91%	1.99%	0.23%	0.80%
37	regARIMA	30	MI6-6	-5.48%	4.12%	0.19%	1.75%
37	improved trend	1	MI1	-2.03%	5.39%	0.39%	1.43%
37	improved trend	1	MI6-1	-4.27%	5.39%	0.07%	1.99%
37	improved trend	1	MI6-2	-4.64%	8.80%	0.57%	2.76%
37	improved trend	1	MI6-3	-4.55%	10.04%	0.58%	3.19%
37	improved trend	1	MI6-4	-2.94%	5.08%	0.64%	1.63%
37	improved trend	1	MI6-5	-4.08%	4.07%	0.19%	1.58%
37	improved trend	1	MI6-6	-4.53%	8.34%	0.74%	2.64%
37	improved trend	30	MI1	-1.27%	2.80%	0.22%	0.78%
37	improved trend	30	MI6-1	-3.76%	3.72%	0.10%	1.48%
37	improved trend	30	MI6-2	-4.70%	5.42%	0.15%	2.19%
37	improved trend	30	MI6-3	-5.08%	4.72%	0.04%	1.73%
37	improved trend	30	MI6-4	-3.95%	3.86%	0.38%	1.37%
37	improved trend	30	MI6-5	-1.69%	2.88%	0.46%	1.23%
37	improved trend	30	MI6-6	-3.04%	4.18%	0.11%	1.39%
100	current	124	MI1	-0.53%	1.77%	0.27%	0.57%
100	current	124	MI6-1	-1.94%	1.75%	-0.04%	0.69%
100	current	124	MI6-2	-2.94%	3.49%	0.60%	1.34%
100	current	124	MI6-3	-2.73%	2.89%	0.06%	1.05%
100	current	124	MI6-4	-1.49%	1.49%	0.27%	0.58%
100	current	124	MI6-5	-0.82%	0.98%	0.14%	0.34%
100	current	124	MI6-6	-2.10%	4.81%	0.63%	1.62%
100	regARIMA	1	MI1	-0.97%	2.08%	0.29%	0.62%
100	regARIMA	1	MI6-1	-2.24%	1.40%	0.04%	0.78%
100	regARIMA	1	MI6-2	-3.37%	3.33%	0.32%	1.32%
100	regARIMA	1	MI6-3	-2.06%	2.98%	-0.17%	1.09%
100	regARIMA	1	MI6-4	-1.34%	1.48%	0.35%	0.63%
100	regARIMA	1	MI6-5	-1.09%	1.10%	0.08%	0.48%
100	regARIMA	1	MI6-6	-1.83%	5.49%	0.71%	1.67%
100	regARIMA	6	MI1	-0.91%	1.70%	0.25%	0.55%
100	regARIMA	6	MI6-1	-2.03%	1.49%	0.16%	0.66%
100	regARIMA	6	MI6-2	-2.40%	2.28%	0.34%	1.16%
100	regARIMA	6	MI6-3	-2.04%	1.20%	-0.25%	0.72%
100	regARIMA	6	MI6-4	-1.41%	1.89%	0.44%	0.60%
100	regARIMA	6	MI6-5	-1.18%	1.44%	0.17%	0.53%
100	regARIMA	6	MI6-6	-2.41%	3.89%	0.40%	1.37%
100	improved trend	6	MI1	-0.83%	1.86%	0.25%	0.49%
100	improved trend	6	MI6-1	-1.49%	2.30%	0.29%	0.83%
100	improved trend	6	MI6-2	-1.90%	3.12%	0.39%	1.29%
100	improved trend	6	MI6-3	-1.41%	2.08%	-0.23%	0.73%
100	improved trend	6	MI6-4	-1.08%	2.49%	0.52%	0.68%

days	method	AI	MI	minimum	maximum	average	st. dev.
100	improved trend	6	MI6-5	-1.30%	2.22%	0.27%	0.76%
100	improved trend	6	MI6-6	-2.55%	3.05%	0.21%	0.95%
100	improved trend	30	MI1	-0.85%	1.49%	0.26%	0.41%
100	improved trend	30	MI6-1	-1.53%	2.30%	0.29%	0.81%
100	improved trend	30	MI6-2	-2.32%	3.42%	0.36%	1.41%
100	improved trend	30	MI6-3	-1.42%	1.97%	-0.19%	0.78%
100	improved trend	30	MI6-4	-1.14%	2.89%	0.54%	0.67%
100	improved trend	30	MI6-5	-1.05%	1.54%	0.23%	0.51%
100	improved trend	30	MI6-6	-2.59%	2.40%	0.22%	0.97%