Notater
Documents
71/2012

Statistics Norway
Statistisk sentralbyrå

Øyvind Langsrud

Automatic outlier handling and model selection in seasonal adjustment

– A history analysis study involving three suggested outlier algorithms

Documents 71/2012

Øyvind Langsrud

Automatic outlier handling and model selection in seasonal adjustment

- A history analysis study involving three suggested outlier algorithms

Statistisk sentralbyrå • Statistics Norway
Oslo–Kongsvinger

Documents In this series, documentation, method descriptions, model descriptions and standards are published.

Published December 2012

ISBN 978-82-537-8559-2 (printed)
ISBN 978-82-537-8560-8 (electronic)
ISSN 1891-5906
Subject: 00.90

Print: Statistics Norway

Symbols in tables	**Symbol**
Category not applicable	.
Data not available	..
Data not yet available	…
Not for publication	:
Nil	-
Less than 0.5 of unit employed	0
Less than 0.05 of unit employed	0.0
Provisional or preliminary figure	*
Break in the homogeneity of a vertical series	—
Break in the homogeneity of a horizontal series	|
Decimal punctuation mark	.

Preface

When using X-12-ARIMA for seasonal adjustment, several modeling decisions have to be taken. This can be viewed as a question of balancing the requirement of optimal seasonal adjustment at each time point against the requirement of minimal revisions. This problem is also addressed in Eurostat's ESS Guidelines on Seasonal Adjustment. The work in this report is a part of Statistics Norway's activity related to the implementation of these guidelines. The author would like to thank Ane Seierstad, Aslaug Hurlen Foss and Dinh Quang Pham for valuable comments and discussions.

Statistics Norway, 7 Desember 2012

Hans Henrik Scheel

Abstract

The first part of this paper presents the most important results from a history analysis of 52 Norwegian economic time series (Langsrud, 2011). It is illustrated how revisions is affected by two automatic ARIMA model selection methods (automdl and pickmdl). Furthermore, it is shown that straightforward re-identification of outliers (the concurrent method) leads to big revisions. From this knowledge the second part of the paper considers the problem of automatically dealing with outliers (Langsrud, 2012). How should potential outliers be handled before the final decision is made? Three algorithms are suggested which can be named as “jump in and out”, “jump in” and “jump out”. It is demonstrated how revisions and out-of-sample forecasts (quality of model) are affected by using the algorithms. The results are compared to the concurrent method. The results indicate, however, that the best improvements are obtained by increasing the outlier detection limit. The analyses were made by running X-12-ARIMA via the R programming language.

Contents

1 Introduction

Times series analysis by using regARIMA models (linear regression models with ARIMA time series errors) is an important part of seasonal adjustment methodology. Alternative modeling variants can be compared fairly by looking at out-of-sample forecasts and revisions. How well the regARIMA models describe the data can be evaluated by out-of-sample forecast diagnostics. Since the model changes over time, the level of revisions is another important quality aspect.

In X-12-ARIMA revision diagnostics are based on absolute values of relative differences. On the other hand, the history analysis of X-12-ARIMA produce sums of squared out-of-sample forecast errors based on log-transformed data (log model). However, in this paper, revision differences and forecast errors will be treated in a similar way. Details of this approach are described in Section 2.

Section 3 will illustrate how ARIMA model selection procedures and outlier detection limits affect revisions and forecasts.

At Statistics Norway a common practice is that outliers are re-identified by the automatic procedure every time new data become available (the concurrent method). According to the *ESS Guidelines on Seasonal Adjustment* (Eurostat, 2009) this method is *"To be avoided"*. Furthermore these guidelines say: *"The concurrent adjustment strategy generates the most accurate seasonally adjusted data at any given time point but will lead to more revisions, many of which will be small and perhaps in opposing directions"*. Our aim is to implement improvements in order to reduce revision—but how should this be done? Manual decisions are not always practical. The concurrent method is automatic. An interesting question is whether it is still possible to use an automatic procedure.

Section 4 suggests three algorithms and the results from history analysis are presented in Section 5.

In the history analysis we consider 52 Norwegian economic time series where 32 series are related to the production index (years 1989-2009). The other 20 series are related to the index of household consumption of goods (years 1979-2009). The time intervals for the history analyses are set to the last 14 and 20 years, respectively.

2 A unified approach for forecast errors and revisions

Assume a time series Y_t where $t = 1, 2, \ldots, N$. By using time series modeling, a h-step ahead forecast, $Y_{t+h|t}$, can be produced. The relative forecasting error on the original scale is approximately equal to the absolute error on the log scale (natural logarithm). More precisely we have the inequality

$$\frac{Y_{t+h|t} - Y_{t+h}}{Y_{t+h|t}} \leq \Big(\log(Y_{t+h|t}) - \log(Y_{t+h})\Big) \leq \frac{Y_{t+h|t} - Y_{t+h}}{Y_{t+h}} \qquad (1)$$

Thus, the log-scale difference can be interpreted as a relative difference. Furthermore, it can be viewed as a compromise between $Y_{t+h|t}$ and Y_{t+h} as

Table 1: *Difference measures used to calculate root mean square error (RMSE).*

	Difference measure
One-month revisions of seasonally adjusted data	$\log(A_{t\|t+1}) - \log(A_{t\|t})$
One-year revisions of seasonally adjusted data	$\log(A_{t\|t+12}) - \log(A_{t\|t})$
First-year average absolute month-to-month revisions of seasonally adjusted data	$\frac{1}{12}\sum_{h=1}^{12}\left\|\log(A_{t\|t+h}) - \log(A_{t\|t+h-1})\right\|$
One-month out-of-sample forecasts	$\log(Y_{t+1\|t}) - \log(Y_{t+1})$

the denominator when calculating the relative difference.

The forecasting errors on a time interval can be summarized as root mean square error (RMSE):

$$\text{RMSE} = \sqrt{\frac{1}{N_1 - N_0 + 1}\sum_{t=N_0}^{N_1}\Big(\log(Y_{t+h|t}) - \log(Y_{t+h})\Big)^2} \tag{2}$$

The relation between relative difference and log-scale difference can be expressed similarly for seasonal adjusted data. That is

$$\frac{A_{t|t+h} - A_{t|t}}{A_{t|t}} \approx \Big(\log(A_{t|t+h}) - \log(A_{t|t})\Big) \tag{3}$$

where $A_{t|t+h}$ is the seasonal adjustment of Y_t calculated from the series where Y_{t+h} is the last observation. Commonly, percentage revisions (multiply by 100%) are based on the left side of this expression. However, in this paper, the log-scale difference is used (right side) and three types of revision measures will be considered. In addition we will look at one-month out-of-sample forecasts. All four measures are described in Table 1. RMSE in equation (2) corresponds to the last line in Table 1 ($h = 1$).

3 History analysis: ARIMA model selection

Two ARIMA model selection procedures were compared; automdl (default in program) and pickmdl (default at Statistics Norway). When using pickmdl only five model candidates were allowed. The automdl procedure selects the model from a broader range of candidates. To ensure that the ARIMA model were the only modeling difference, fixed outliers were used. These outliers were found from a single analysis of the whole series by using the ARIMA model according to automdl. Effects of trading days and moving holidays were included in the regression specification.

The results are given in Figures 1. Each point in each scatter plot represent the results from one series and RMSE were calculated as described in the Section 2. The axis values are percentage values. One series with extreme behaviour was omitted from the plotting (because of axis limits).

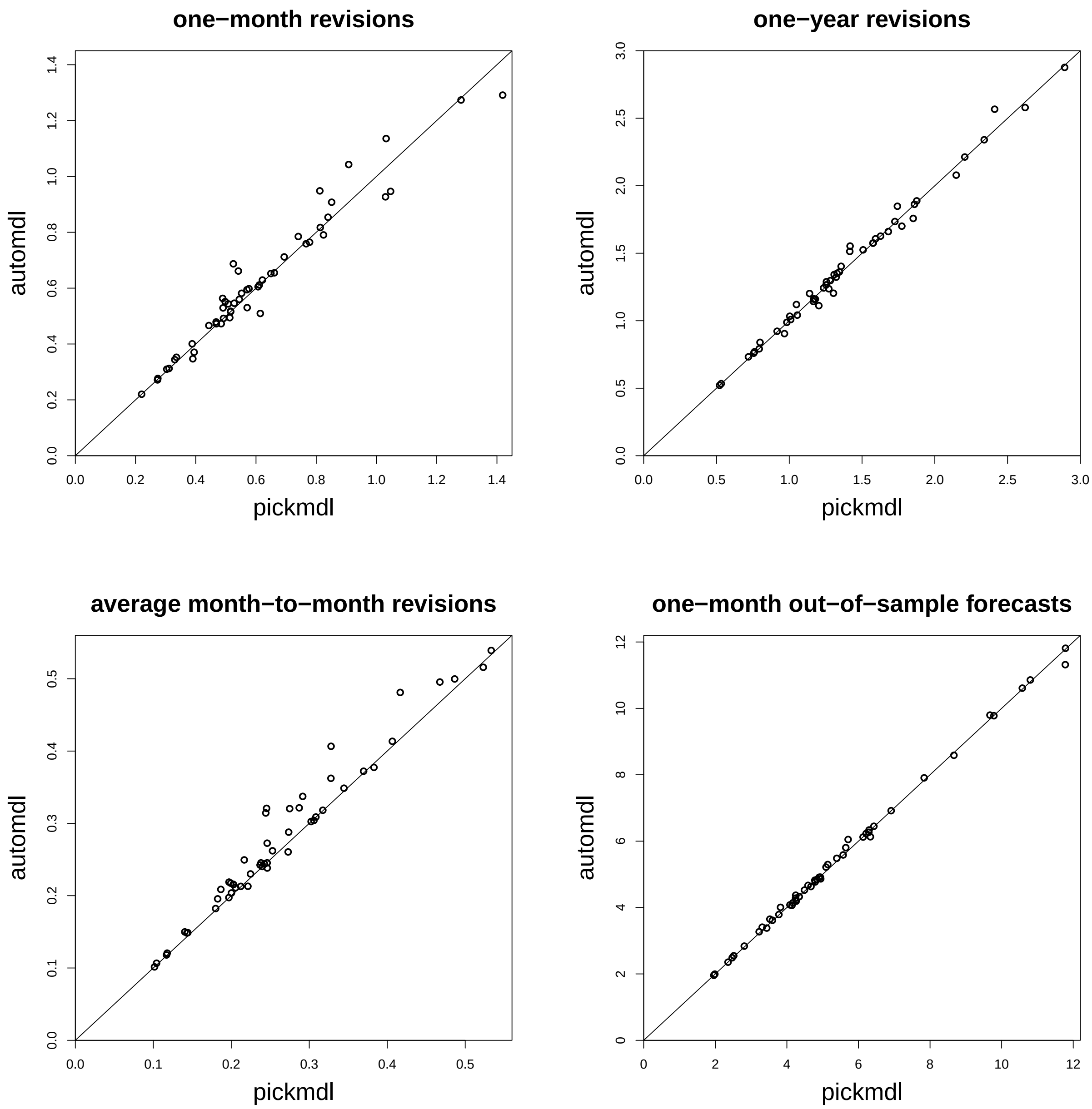

Figure 1: *RMSE of revisions and out-of-sample forecasts. Results from history analysis of 51 series are plotted. Two ARIMA model selection procedures are compared. The diagonal line represents equal values.*

The amount of average month-to-month revisions can be read from the two axes of the lower left panel. In this plot most points are above the diagonal line. That is, pickmdl resulted in reduced revisions compared to automdl. This result coincides with the results obtained by Stuckey and Campbell (2012). The other tree panels in Figure 1 do not show visible differences.

Two outlier detection limits ($t = 4$ and $t = 8$) were compared in a similar way. Both additive outliers and level shifts were allowed. The results can be found in the lower right panel of Figures 2-5. It is clear that increasing the oulier detection limit reduce revision. The methodology behind the other panels in these figures are described below.

4 Three suggested outlier algorithms

A final decision whether an observation is an outlier or not cannot be made (automatically) when the observation is new. Below we assume that this decision is made after one year ("month number 13"). The three algorithms are possible suggestions of how to deal with outliers the first year. Below AO and LS denote, respectively, additive outlier and level shift.

Algorithm 1: Outliers are allowed to **jump in and out** of the model. The outlier span is 13 months (critical t-value = 4, AO and LS). The one year old observation can become a fixed outlier.

Algorithm 2: Outliers are allowed to **jump in**. The outlier span is 13 months (critical t-value = 4, AO and LS). All outliers found become fixed outliers.

Algorithm 3: Outliers are allowed to **jump out**. This algorithm aims to include outlier candidates as preliminary outliers the first year. When a preliminary outlier is one year old, it is decided whether it should remain as a fixed outlier. Preliminary outliers are chosen the first month based on a small critical t-value (=2). Algorithm 1 is run to determine whether the one year old observation should become a fixed outlier. This means that "jump in" may happen rarely. In practise three runs are conducted:

- The first run is exactly as Algorithm 1. The one year old observation can become a fixed outlier.
- In the second run the outlier span is one month and critical t-value = 2 (only AO). Whether the newest observation becomes a preliminary outlier is decided. This run do not use preliminary outliers found earlier. Instead the outliers found in the first run are used ("month 2 – month 12").
- The last run use all the fixed and preliminary outliers. The seasonal adjustment is based on this run.

Instead of one year one may choose two or three years. It is also possible to modify these algorithms so that final decisions are made once a year at annual revisions. Note that Algorithm 3 is related to "current adjustment" which uses forecasted seasonal and calendar factors. If the critical t-value for preliminary outliers is set to zero, all observations are identified as preliminary outliers. This means that the seasonal and calendar factors are estimated from the older data, which is similar to using forecasts.

5 History analysis: Automatic outlier treatment

The concurrent method with an extreme outlier detection limit were tested in addition to the three algorithms. This method applied to the most recent data for all the 52 series gave a total number of nine outliers. The results from the history analysis are given in Figures 2-5. Again, one series with extreme behaviour was omitted from the plotting (because of axis limits). Figure 2 and Figure 3 look at the revisions after one and twelve months. Figure 4 looks at first-year average month-to-month revisions and Figure 5 looks at one-month out-of-sample forecasts. Details about the four measures can be found in Table 1.

The lower left scatter plot of Figure 2 compares Algorithm 3 (horizontal axis) to the ordinary concurrent method (vertical axis). In this plot all points are above the diagonal line. That is, Algorithm 3 resulted in reduced one-month revisions for all the series. In fact, it is no doubt that Algorithm 3 is the best method in Figure 2. On the other hand, in Figure 3 (one-year revision), Algorithm 3 is the worst method. After one year most preliminary outliers become non-outliers with extra revision as a consequence. When looking at first year average month-to-month revisions (Figure 4) Algorithm 3 is still very good. However, the out-of sample forecasts (Figure 5) of Algorithm 3 is not good.

From Figure 5 we can see that the Algorithm 2 may fail. The most extreme series has been investigated closely. The seasonal pattern has changed, but Algorithm 2 will never learn. December month always become an outlier. A reformulation of this algorithm may solve this problem. Anyway, it seems that the revision performance of Algorithm 1 is nearly as good as Algorithm 2.

The most convincing result from the history analysis is that the best way of reducing revisions is by increasing the outlier detection limit. The out-of sample forecast performance do not decrease.

6 Concluding remarks

The results illustrate that purely automatic use of X-12-ARIMA leads to revision problems. These problems can be reduced by using pickmdl instead of automdl and by increasing the outlier detection limit.

It is also possible to reduce revision by using an automatic outlier algorithm. The price to pay is that the model quality is worsened (out-of-sample forecasts).

A better solution is, however, to handle these problems in a non-automatic way.

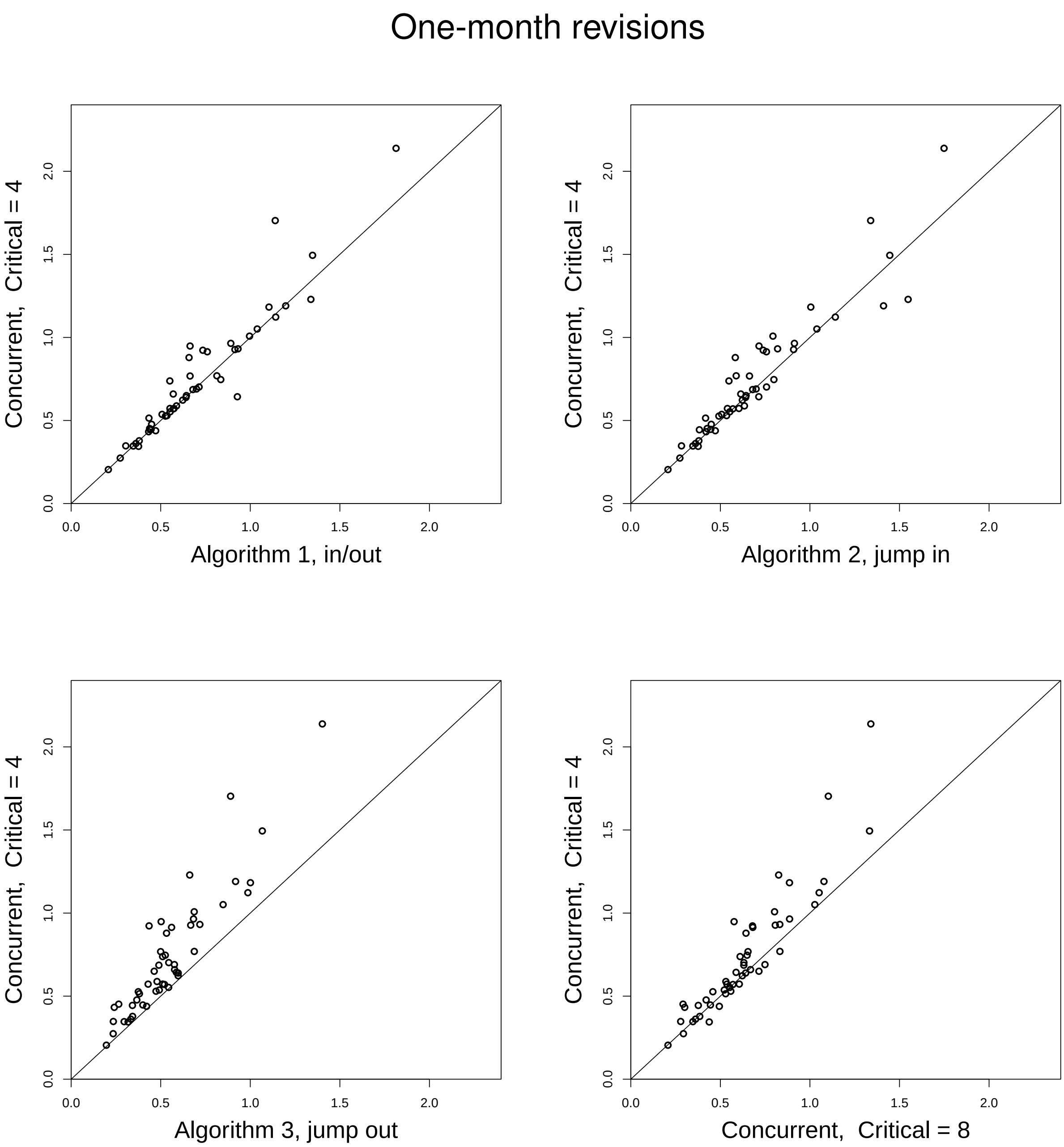

Figure 2: *RMSE of **one-month revisions** of seasonally adjusted data. Results from history analysis of 51 series are plotted. Four alternatives (horizontal axis) are compared to the ordinary concurrent method (vertical axis). The diagonal line represents equal values.*

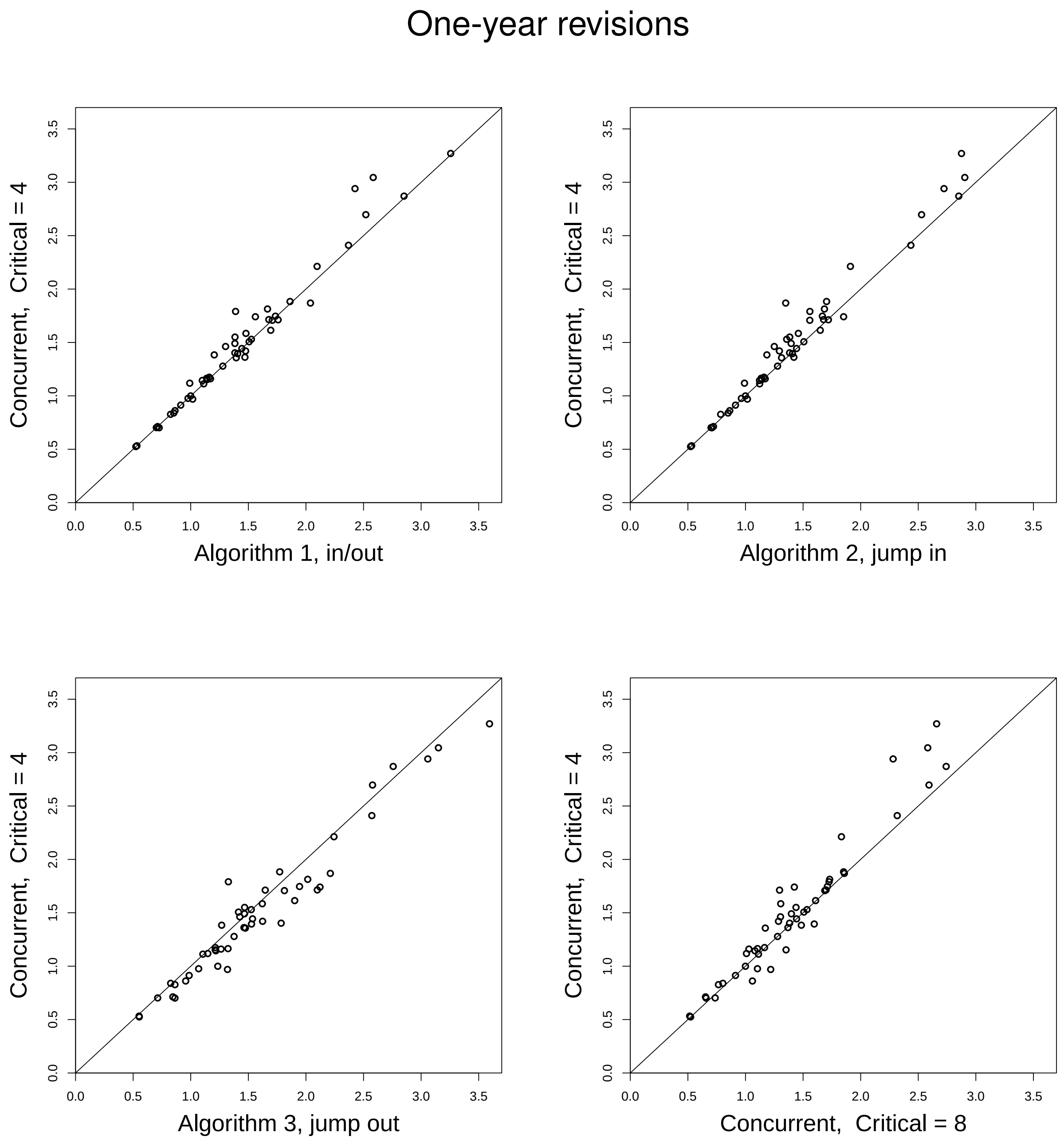

Figure 3: *RMSE of **one-year revisions** of seasonally adjusted data. Results from history analysis of 51 series are plotted. Four alternatives (horizontal axis) are compared to the ordinary concurrent method (vertical axis). The diagonal line represents equal values.*

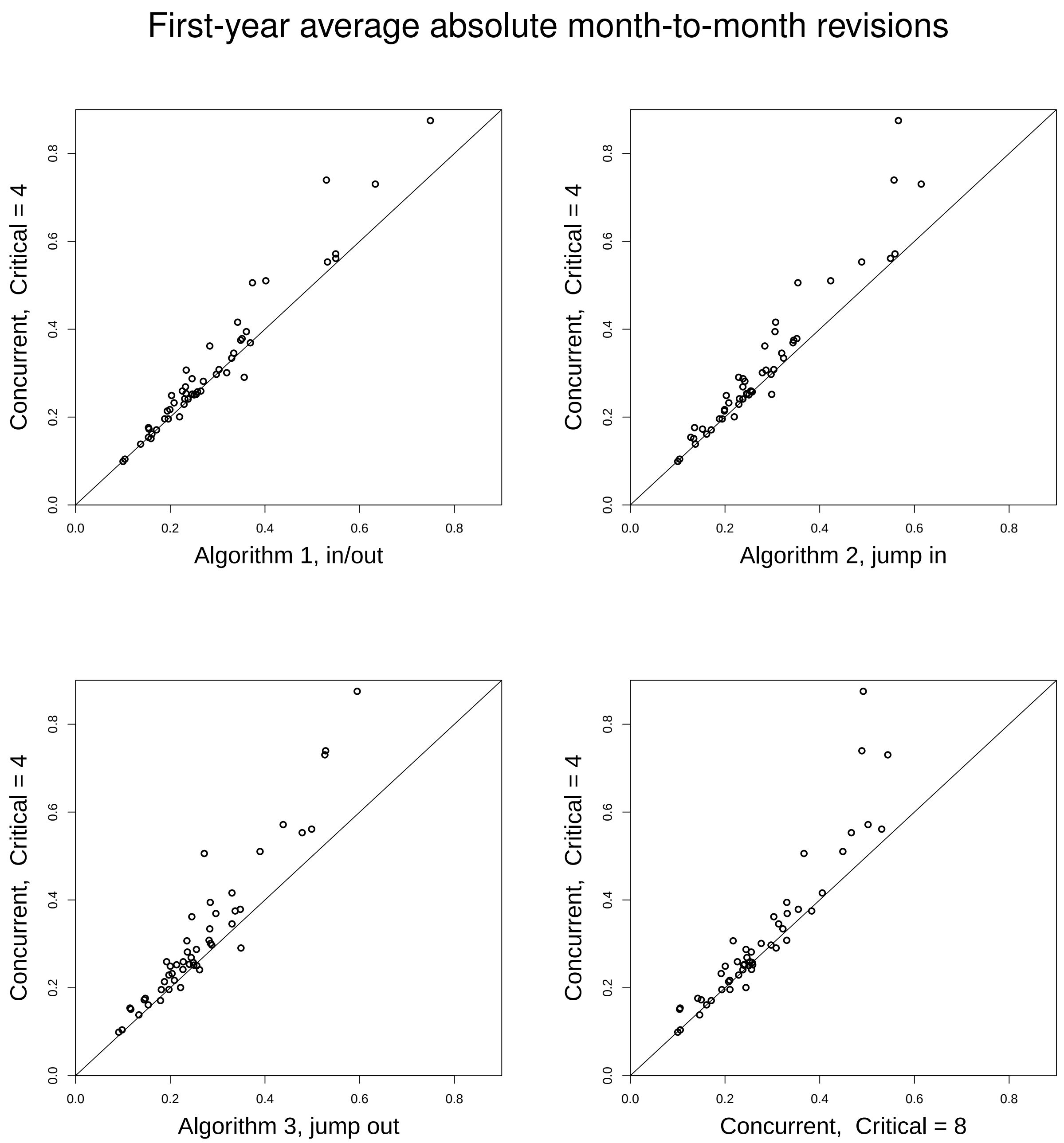

Figure 4: *RMSE of **first-year average absolute month-to-month revisions** of seasonally adjusted data. Results from history analysis of 51 series are plotted. Four alternatives (horizontal axis) are compared to the ordinary concurrent method (vertical axis). The diagonal line represents equal values.*

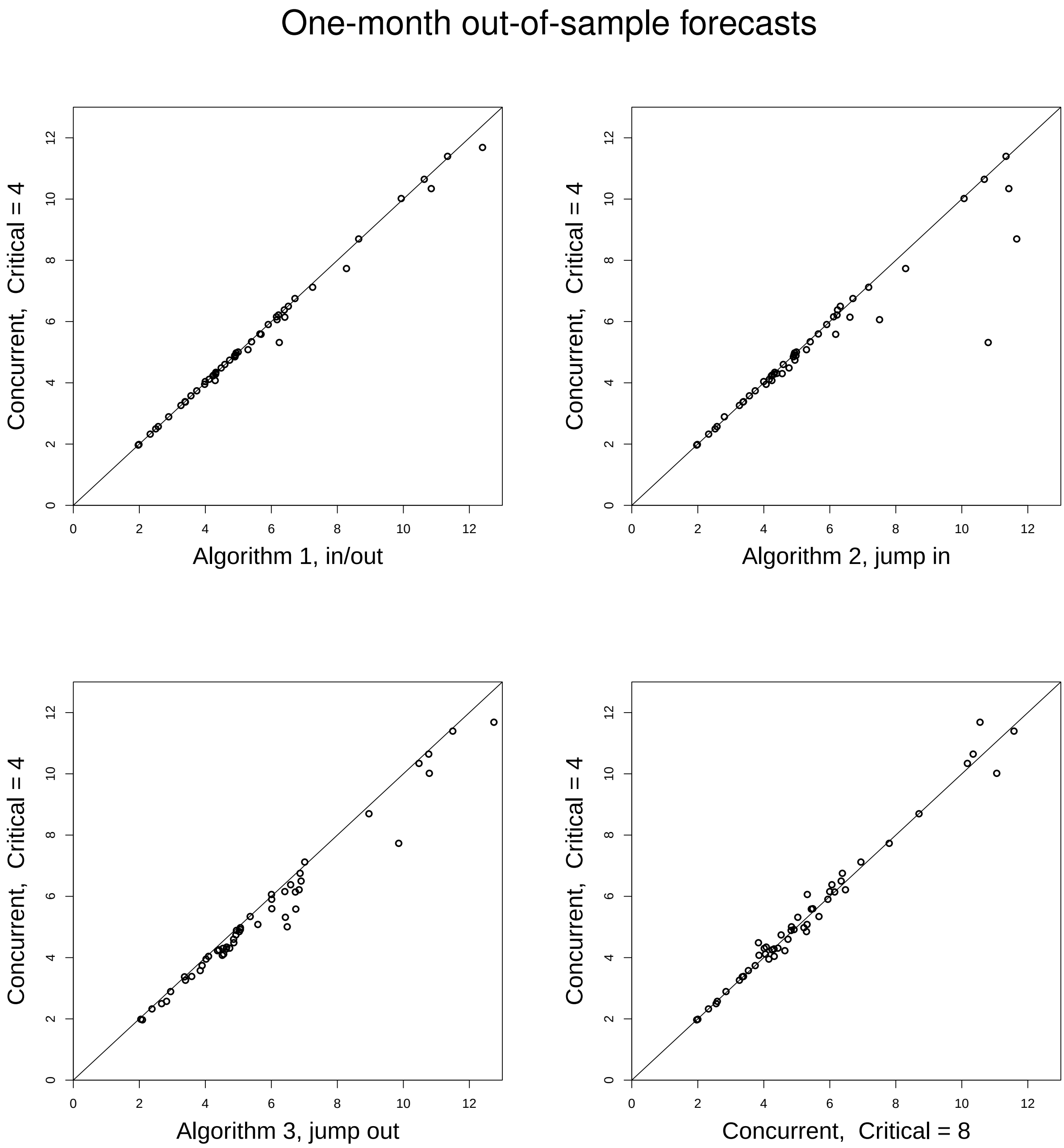

Figure 5: *RMSE of **one-month out-of-sample forecasts**. Results from history analysis of 51 series are plotted. Four alternatives (horizontal axis) are compared to the ordinary concurrent method (vertical axis). The diagonal line represents equal values.*

References

Eurostat (2009), "ESS Guidelines on Seasonal Adjustment", *Office for Official Publications of the European Communities*, Luxembourg.

Langsrud, Ø. (2011), "Seasonal Adjustments: Causes of Revisions", *2011 European Establishment Statistics Workshop*, 12 –14 September 2011, Neuchâtel, Switzerland.

Langsrud, Ø. (2012), "Fixing Outliers Automatically in Seasonal Adjustment: Testing three algorithms by history analysis", *2012 Workshop on methodological issues in Seasonal Adjustment*, 6 March 2012, Eurostat, Luxembourg.

Stuckey, A. and Campbell, J. (2012), "A comparison of automated ARIMA model selection methods to benefit seasonal adjustment using X12-ARIMA", *The 32nd International Symposium On Forecasting*, 24–27 June 2012, Boston, USA.

U.S. Census Bureau (2009), "X-12-ARIMA Reference Manual, Version 0.3", Washington, D.C.

B

Retun to:
Statistisk sentralbyrå
NO-2225 Kongsvinger

Statistics Norway

71/2012

Automatic outlier handling and model selection in seasonal adjustment

From:
Statistics Norway

Postal address:
PO Box 8131 Dept
NO-0033 Oslo

Office address:
Kongens gate 6, Oslo
Oterveien 23, Kongsvinger

E-mail: ssb@ssb.no
Internet: www.ssb.no
Telephone: + 47 62 88 50 00

ISBN 978-82-537-8559-2 (printed)
ISBN 978-82-537-8560-8 (electronic)
ISSN 1891-5906

ISBN 978-82-537-8559-2

9 788253 785592

Design: Siri Boquist

Statistisk sentralbyrå
Statistics Norway