Use of ARIMA Models for Improving the Revisions of X-11 Seasonal Adjustment

Time Series Analysis Section
Australian Bureau of Statistics

Section 1: Introduction

1. The Australian Bureau of Statistics (ABS) uses an enhanced version of the X-11 Variant of the Census Method II Seasonal Adjustment Program (Shiskin et. al, 1967). The X-11 method applies moving average techniques to decompose the time series into estimates for the trend, seasonal and irregular components.

2. A linear filter with symmetric weights (ie. a symmetric filter) has many desirable properties for seasonal adjustment. However, symmetric filters cannot be applied at the ends of a time series. The standard X-11 uses asymmetric filters to solve this problem. Asymmetric filters are often designed with assumptions about certain properties of the missing observations. ie. forecasts are implicitly applied. As a result of the forecasting, revisions will occur between the first few estimates of the seasonal factors and trend at a particular time point and the final trend estimate which is calculated using a symmetric moving average. The problem of the revisions derived from the application of asymmetric filters is often called the "end-weight" problem.

3. Revisions resulting from the applications of asymmetric filters are necessary to improve seasonal adjustment estimation as more data becomes available. However, revisions are undesirable and methods to minimise them are an ongoing research pursuit for the ABS. A better forecast of the missing observations can lead to a reduction in the revisions of the seasonal adjustment estimates.

4. Assuming the symmetric filters used in X-11 are satisfactory, the purpose of this paper is to investigate the possibility of reducing the "end-weight" problem by using X-11 with ARIMA extensions. In this paper, we

1. evaluate the fitting performance of a set of the most often applied ARIMA models;
2. confirm that a reduction in revisions is achieved when using ARIMA forecasting instead of the asymmetric filters in the standard X-11;
3. identify the conditions under which an ARIMA performs better than the standard X-11.

5. To achieve these goals a range of ABS time series (820 series) have been used. Our research showed that:

1. In comparison with the standard X-11, on average, an ARIMA model achieves 6-7% reduction in revision to the first seasonally adjusted estimate;

2. The average absolute percentage change in irregulars (denoted by STAR) is the most powerful statistical measure to predict the possible revision size;

3. The ratio of the average absolute percentage change in the irregular component and the average absolute percentage change in the X-11 trend (denoted by I/C) is the most powerful statistical measure to indicate when an ARIMA model is likely to be better than the standard X-11. The larger the I/C, the better an ARIMA model performs over the standard X-11.

6. This report is organised as follows. Section 2 gives a short literature review of the "end-weight" problem. Section 3 presents a comparison of the revisions from the standard X-11 versus X-11 with ARIMA extension. Section 4 presents statistical analyses aimed at identifying how the revision is related to a set of common statistical measures, and which measures are most likely to predict if X-11 with ARIMA extension can produce less revision than the standard X-11. Our conclusions, recommendation and future research directions are discussed in Section 5.

Section 2: Background

7. X-11 is a seasonal adjustment package developed by the US Bureau of the Census in the 1950's and 1960's (Shiskin et. al 1967, Ladiray and Quenneville 2001). X-11 uses moving averages to decompose a time series into estimates of the trend, seasonal and irregular components. The trend and seasonal components use their own set of symmetric and asymmetric moving averages (filters) to calculate estimates in the middle and at the ends of the data. The details are outlined in Table 1.

Table 1: Moving averages (filters) used in the X-11 seasonal adjustment package.

	Trend	Seasonal
symmetric	The symmetric moving averages used to estimate the trend component are calculated using the Henderson moving average (Henderson, 1916) and the length of the filter is determined by a relative variation measure of the irregular and trend.	The symmetric moving averages used for the seasonal are based on convoluted simple moving averages eg. 3x5, 3 term of a 5 term simple moving average (Dagum, 1996). The "optimal" length of the seasonal filter is determined by a relative variation measure of the irregular and seasonal.
asymmetric	The asymmetric moving averages used for the ends of the trend are calculated using a simple linear model developed by Musgrave (Doherty, 1992).	The asymmetric moving averages for the seasonal moving averages are based on an unknown and undocumented methodology (see Appendix B for more details).

8. There are two broad categories of methods to improve the revision ("end-weight") problem: (1) improvement of asymmetric filters (Gray and Thomson 1996a 1996b 1996c, McLaren and Steel 1998, Quenneville and Ladiray 2000), and (2) forecast the missing data using more advanced time series dynamic models. The latter approach is used for this report.

9. The use of asymmetric filters results in the revision of initial estimates of the seasonally adjusted and trend as subsequent data points are added to the time series. The class of asymmetric filters in X-11 use implicit static forecast formulae. The optimal choice of an asymmetric filters is determined by the statistical characteristics of the time series. These characteristics are typically the average absolute percentage changes in irregular/seasonal (denoted by I/S) and the average absolute percentage changes irregular/trend (denoted by I/S). Therefore, the asymmetric filter approach used in X-11 is a semi-dynamic model in that the implicit forecast formulae are not completely dependent on the nature of the time series. For

a further discussion of the asymmetric filter see Appendix B.

10. Statistics Canada developed X-11-ARIMA in the 1970's and 1980's in an effort to reduce revisions. X-11-ARIMA uses Box and Jenkins ARIMA modelling (Box and Jenkins 1970, Dagum 1980) to forecast the original missing data points. Quenneville and Ladiray (2000) use models to forecast missing seasonally adjusted estimates and then apply the symmetric Henderson filter to produce trend estimates. The ABS evaluated X-11-ARIMA in the 1980's but found too many series were not automatically modelled by the package. It is concluded that X-11-ARIMA was not suitable for mass production of ABS seasonal adjustment. The US Bureau of the Census has recently released X-12-ARIMA (Findley et. al 1998), and also adopted ARIMA modelling to predict missing data points. X-12-ARIMA includes a regression-ARIMA component to estimate a range of data contamination effects.

11. This analysis uses the ABS seasonal adjustment package SEASABS (ABS, 1999) to clean the time series of any abrupt changes in the trend and seasonal components, trading day, large extremes and numerous other possible effects. X-12-ARIMA was then used to fit and evaluate the ARIMA extension method to improve the "end-weight" problem. We evaluated the revision properties of the first seasonally adjusted estimates.

12. All series evaluated were seasonally adjusted directly. In practice, many ABS series are seasonally adjusted then aggregated to form an indirect seasonally adjusted estimate at a higher level.

Section 3: Comparison of revisions between X-11 and different ARIMA options

13. The performance of the following four ARIMA modelling options were compared using X-12-ARIMA:

1. the standard X-11 (using asymmetric filters)
2. the "airline" ARIMA model (dynamic model)
3. "super model" (static ARIMA model, see Appendix C for details) with prefixed parameters
4. automatic selection of ARIMA model (see Appendix D).

14. These options let us to evaluate the performance of

- a common dynamic ARIMA model (ie. the "airline" model), an empirical static ARIMA (ie. the "super" model) in comparison with the standard X-11, and
- an "optimal" ARIMA model, which is automatically selected by the default model selection procedure in X-12-ARIMA (see Appendix D for more details), in comparison with the common "airline" model.

The result of such comparisons will provide a clear and insightful picture of the strengths and weaknesses of the different "end-weight" methodologies.

Additional investigations were performed using the "outlier" modification of regression-ARIMA in X-12-ARIMA.

15. The desired properties of revisions are:

1. The first seasonally adjusted and trend estimates (ie. the estimates produced when the time period of interest is the last in the series) are as close as possible to the final estimates when more observations are added to the series (this may be some years after the initial data point was first added to the series).
2. The seasonally adjusted and trend estimates subsequent to the first estimate converge as quickly and smoothly as possible to the final estimates.

16. Revisions are calculated by simulating monthly (or quarterly) updates of historical data over a specific time span. The revision measure used in this paper is the mean absolute percentage revision in the level of seasonally adjusted estimates from the first to the final estimate. This is denoted by R_0 and defined as

$$R_0 = \frac{100}{T_0 - t_0 + 1}\sum_{t=t_0}^{T_0}\frac{\left|A_{t|T} - A_{t|t}\right|}{A_{t|T}}$$

where $A_{t|_\tau}$ is the seasonal adjusted estimate of date t given date τ ($t \leq \tau$); T is the end date of the original series; t_0 is the start date of the original series minus seven years data (t_0 = SD - 84 for monthly series or SD - 28 for quarterly series; T_0 is less than or equal to T minus three years data ($T_0 \leq$ T - 36 for monthly series or T - 12 for quarterly series); The generic form of this measure can be found in Appendix E.

17. To compare the performance of ARIMA options against the standard X-11, the percentage of revision ratio, $RR = 100xR_0(ARIMA)/R_0(X11)$, between the ARIMA option to the standard X-11 has been found most effective. For example, 100 would mean that the ARIMA option and standard X-11 performed equally with regard to revisions, while 90 would mean the ARIMA option was 10% better than the standard X-11.

18. Table 2 summarises the results from a total of 820 time series from a variety of different sources. In general the "airline" model consistently outperforms the standard X-11 with an overall reduction in revisions about 6-7%. This is in spite of the fact that the "airline" model may not pass the rigorous model selection criteria. The static "super model" also outperforms X-11 overall. This implies that the forecast models implicitly embedded in the asymmetric filters of the standard X-11 generally lead to larger revisions than the "airline" and "super" model.

19. For the three X-11 with ARIMA extension (airline, super, automatic) seasonal adjustments, the use of the regression-ARIMA outlier removing option resulted in small gains averaging between 1.5% and 2.8%. However, the trade off was increased variability as measured by standard deviations of R_0.

20. In calculating the statistics for the automatic ARIMA model selection approach (with and without outlier options), all series which failed to meet the model fitting criteria were excluded. The automatic procedure fitted an ARIMA model for 576 (ie. 70.2%) of the series. Of these, 101 series (ie 17.5%) performed worse than the standard X-11. The general performance of the automatic model selection option using X-12-ARIMA (see Appendix E for more details) is almost the same as the "airline" model. A general conclusion is that for Australian data, the sophistication of choosing a "best" ARIMA model is generally no better than just choosing the simple "airline" model (see Appendix F for more detailed comparisons between the performances of the "airline" and automatic models).

Table 2: Means, medians and standard deviations for ARIMA/X11 revision ratio (%)

Group	number of series analysed in group	ARIMA/X11 revision ratios- group mean (medians in brackets) and standard deviations below						fitted by auto model: RR < 100 RR > 100 auto fit (as%)
		airline	airline with outlier option	super model	super model with outlier option	automatic model	automatic model with outlier option	
Selected sub-sample[1]	97	90.5 (90.2) 8.5	88.2 (88.6) 9.0	94.5 (92.6) 12.1	91.4 (90.4) 11.9	91.7 (93.0) 7.6	92.1 (93.0) 10.3	43 3 47.4%
Labour force	258	94.0 (93.6) 7.3	92.5 (92.1) 7.3	97.6 (96.3) 8.4	95.3 (94.3) 8.4	93.5 (94.1) 7.0	92.7 (93.6) 7.0	180 32 82.2%
Retail	202	93.1 (93.4) 6.3	92.3 (91.1) 7.5	93.4 (97.9) 6.1	91.5 (93.3) 8.1	93.7 (93.6) 6.4	93.3 (93.7) 7.5	137 28 81.7%
Building activity	174	94.0 (93.4) 8.4	90.9 (91.4) 11.3	104.6 (102.7) 13.7	101.9 (99.6) 16.2	94.9 (94.2) 7.7	92.3 (92.5) 10.2	70 20 51.7%
Survey of employment & earnings	67	93.8 (92.5) 8.2	82.8 (83.7) 14.7	95.8 (94.8) 8.1	84.6 (86.8) 17.2	94.7 (93.6) 7.0	83.3 (83.4) 16.6	44 11 82.1%
Meat	42	89.3 (88.7) 5.0	89.3 (87.8) 5.3	89.6 (90.3) 5.5	88.8 (89.0) 5.2	88.5 (88.9) 5.7	97.0 (96.7) 9.5	15 0 35.7%
New motor vehicle reg.	27	93.6 (92.8) 4.2	93.5 (92.6) 3.9	94.2 (94.2) 4.1	93.6 (94.1) 4.6	93.4 (91.9) 4.0	97.4 (99.1) 5.0	8 1 33.3%
Profits	20	103.2 (104.6) 10.9	96.5 (99.9) 17.8	101.5 (99.3) 13.5	100.7 (100.7) 23.5	97.1 (97.5) 12.1	88.2 (88.1) 15.8	4 4 40%
Stocks and Sales	15	99.4 (99.4) 12.1	94.0 (99.4) 16.9	104.9 (106.8) 12.4	103.4 (105.4) 18.8	97.3 (95.2) 8.5	95.0 (94.8) 15.8	7 4 73.3%
Job adds	8	96.7 (100.8) 8.4	94.4 (97.8) 8.1	100.9 (104.5) 9.2	99.1 (103.1) 8.3	91.4 (89.7) 9.0	85.4 (83.5) 4.1	5 1 75%
Wine	7	89.7 (90.6) 4.4	89.8 (90.5) 4.5	91.5 (91.2) 5.1	91.3 (91.2) 5.5	88.6 (87.4) 4.9	90.4 (89.4) 6.3	5 0 71.4%
Total[2]	**820[3]**	**93.8 (93.2) 7.7**	**91.3 (91.7) 9.9**	**97.6 (95.3) 10.3**	**94.8 (93.5) 12.9**	**93.8 (93.6) 7.1**	**92.3 (93.2) 9.7**	**475 101 70.2%**

Notes:

1. The selected sub-sample was formed by selecting series with a broad range of characteristics from the total set of analysed series.
2. Total excludes the selected sub-sample.
3. Percentages refer only to the number of series that did not fail.

Section 4: Which statistical measures can predict if X-11 with ARIMA extension can produce less revision than the standard X-11.

21. This section addresses three questions:

1. Is the revision related to a set of common statistical measures?
2. Does the X-11 with ARIMA extension (eg. the "airline" model) always perform better than the standard X-11 for any time series?
3. Under what circumstances does the standard X-11 perform better than ARIMA models? What are the characteristics of a time series that can predict this?

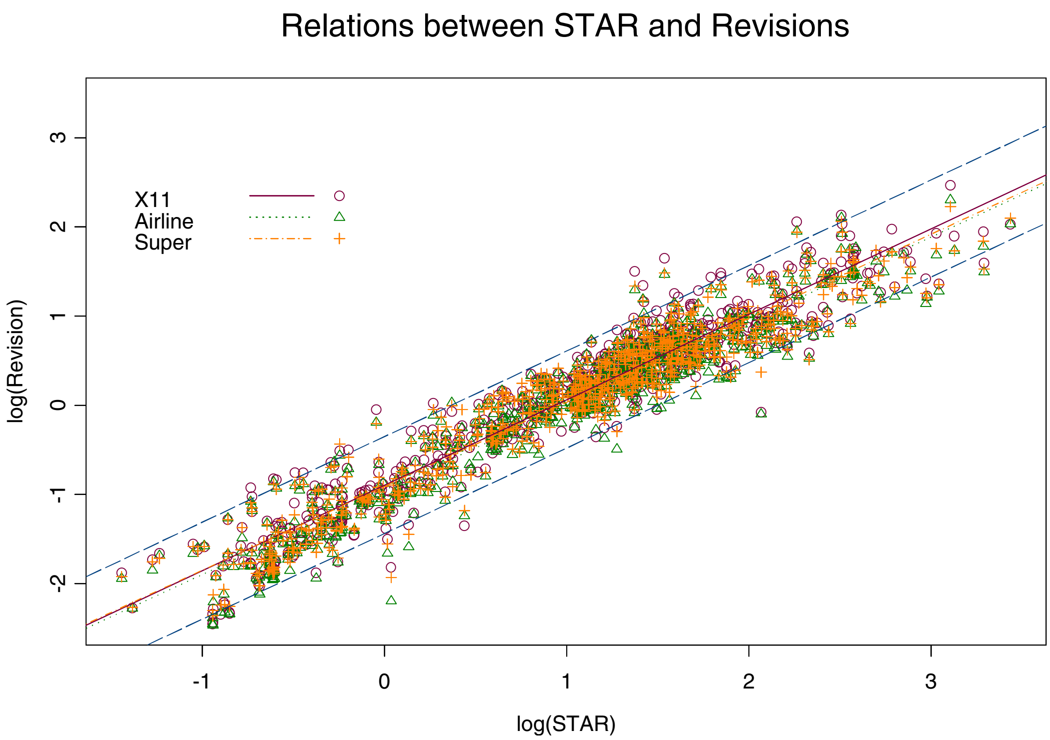
22. A selection of five measures is given in Table 3. All of the measures are quantitative measures of the volatility of a series. Different measures provide information on the volatility of a series from different perspectives. The STAR value measures the volatility of series in the irregular component of a series; I/C and I/S ratio considers the trend and seasonality of a series; TP measures the frequency of cycles of a series rather than its smoothness; and Pouts is a particular measure of the number of outliers in a series which may affect the ARIMA performance.

Table 3: Different measures used within the ABS.

Measure	Description	Source
STAR	A value represents the average absolute percentage change in the irregular (residual) component of a series.	SEASABS
I/C	A ratio of the average absolute percentage change in the irregular component and the average absolute percentage change in the X-11 trend. For an additive series it is the ratio of the average absolute change rather than percentage change in both cases.	SEASABS
I/S	A ratio of the average absolute percentage change in the irregular component and the average absolute percentage change in the X-11 seasonal factors. For an additive series it is the ratio of the average absolute change rather than percentage change in both cases	SEASABS
TP	An indicator measuring the number of turning points in a series. In calculating this measure, we used the trend component of the series and calculated the 10 year average number of turning points for comparability among all series.	Created
Pouts	A ratio measuring the proportion of total number of outliers over total number of observations in a series. The outliers in a series are defined by X-11 SEASABS.	Created

23. We used binary tree recursive partition regression and simple regression (with non-linear transformation) to explore the relationships between the revision measure R_0 (generated by the standard X-11, and X-11 with ARIMA extensions) and the five statistical measures. All of the analyses consistently suggest that the STAR measure is the most significant predictor (with explanatory power of $R^2 > 0.9+$) for the revision size of the three seasonal adjustment methods (the automatic model method is excluded because for 30% the series, X-12-ARIMA could not find a suitable ARIMA model) although all five measures have statistically significant effects (see Appendix G for more details). The result is intuitive in that the higher the measure of volatility, the larger the revisions. This is shown in Figure 1.

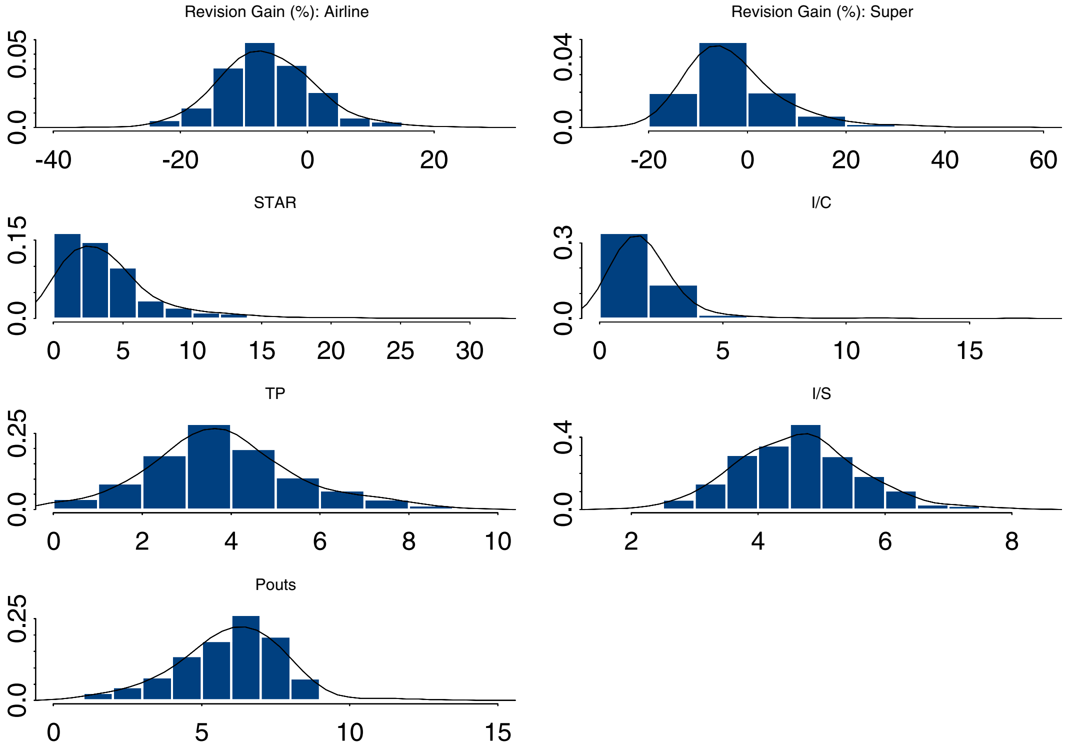
Figure 1: Relation between mean absolute percentage revision in the seasonally adjusted estimates from the first to the final estimate (R_0) and STAR



24. Figure 1 shows that the revisions from the "airline" model tend to be smaller than the standard X-11 when the STAR value increases.

25. As the focus of the paper is on the relative performance of ARIMA methods to X11, we further investigated the possible statistical relationships between the percentage revision gain (defined as RR - 100) and the five measures. Figure 2 shows the distributions of the revision gains of the "airline" and "super" model and the five measures.

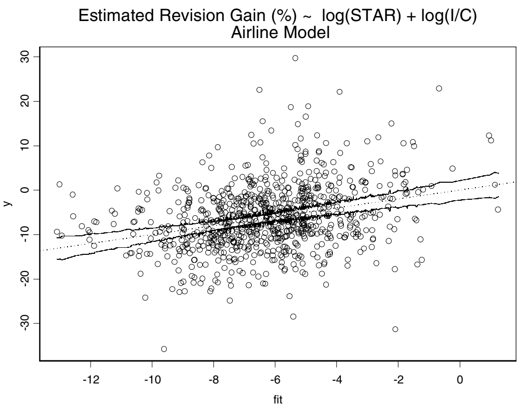
Figure 2: Empirical distributions

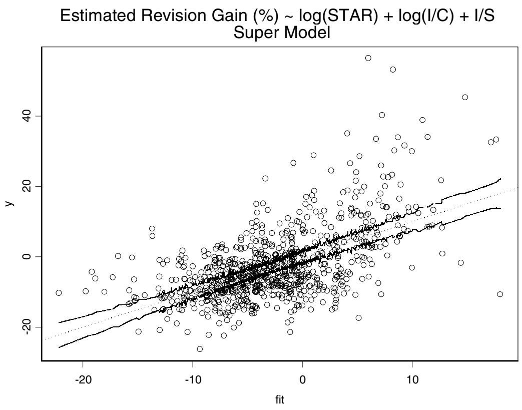


26. The skewed distributions of I/C and STAR suggest a logarithmic transformation is required when a linear regression analysis is performed. The three analyses (see details in Appendix G) of binary tree recursive partition of regression, transformed linear regression and logistic regression consistently show that the I/C is the most significant measure to predict the revision gains, followed by the STAR measure. The rest of the measures are not significant or inconclusive to the revision gains from the "airline" and "super" models. Table 4 shows the results of the linear regression fitting RR-100 to log(STAR), log(I/C) and I/S.

Table 4: Estimated model: RR - 100 = a + b log(STAR) + c log(I/C) + d (I/S)

Airline Model	Super Model
a: -4.264 (-10.715*)	a: 1.228 (2.531*)
b: -0.935 (-3.342*)	b: -0.955 (-2.802*)
c: -2.714 (-7.212*)	c: -7.101 (-15.483*)
d: not significant for this model	d: 2.928 (9.240*)
R^2: 0.0834	R^2: 0.254
Correlation of coefficients: a b b -0.6553 c -0.2776 -0.1270	Correlation of coefficients: a b c b -0.1900 c 0.0327 -0.1250 d -0.9531 -0.0087 -0.1218





27. The estimated coefficients in Table 4 demonstrate that X-11 with ARIMA performs better than the standard X-11 when the I/C and STAR values increase. The estimated coefficients of STAR from the two ARIMA models are very close. This indicates that STAR measure has a relative consistent effect on the revision gain.

28. More importantly, the I/C is the most significant measure. The negative coefficients of I/C in Table 4 suggest that the larger I/C, the better ARIMA are models than the standard X-11. The "airline" model is likely to have less revision when I/C is greater than 0.1675 (ie. log(I/C) is greater than -1.787 (see Graph A13 in Appendix G)). The "super" model is likely to have less revision when I/C is greater than 1.047 (ie. when log(I/C) is greater than 0.0464 (see Graph A16 in Appendix G)). For the total 820 series, there are 816 (99.9%) and 616 (75%) series having I/C values greater than 0.1675 and 1.047 respectively. This implies that the asymmetric trend filter, which is derived from the I/C value, used in the X-11 performs better than the two ARIMA models only when the I/C is extremely small (ie. the movements of the seasonally adjusted series are dominated by the trend movements).

29. The positive coefficient of I/S in Table 4 show that this measure has a positive effect on RR-100 (although it is not significant for the "airline" model). This suggests that the implied seasonal component model embedded in the ARIMA model may not perform as well as the X-11's seasonal asymmetric filter when a series is volatile and has less seasonal pattern changes.

30. As discussed above, given a fixed noise level (ie. a fixed STAR) of a series, a suitable ARIMA model is likely to perform better than the standard X-11 when a series has less trend movement (ie. large I/C) and large seasonal pattern change (ie. small I/S). In other words, the noisier the series the better ARIMA model performs than the standard X-11 if a suitable ARIMA model can be found.

31. A dynamic ARIMA model is a global model using a reasonable length of series to predict missing observations. The Musgrave asymmetric trend filter used in the standard X-11 is a local linear model using only the last few observations with a global parameter I/C to predict missing observations. Under a large I/C, a suitable ARIMA model is more likely to provide a better general indication of the trend movement while the Musgrave asymmetric filter tends to bend the trend movement towards zero. In addition a suitable ARIMA model is also likely to be more adaptive to seasonal pattern changes than the implied forecast formulae of asymmetric seasonal filters (see details in Appendix B).

32. Based on our results, and earlier results of Statistics New Zealand (Statistics New Zealand, 2000), it is possible that ARIMA modelling is better than the X-11 asymmetric filters when a time series is more volatile (large I/C and STAR) and strong in seasonal movement (small I/S), or vice versa.

Section 5: Conclusions and future directions

33. We used the dynamic "airline" and static "super" ARIMA models to show that the implied forecast model of the asymmetric filter in the standard X-11 does not perform as well as the ARIMA models in terms of revisions to the seasonally adjusted estimates. In other words, an imperfect dynamic ARIMA model still produces less revision than the standard X-11 does.

34. We have shown that the revision size in seasonal adjustment is dependent on a range of volatility measures (STAR, I/C, I/S, TP, Pout) of a series. STAR is the most dominant measure for predicting the revision size. An empirical non-linear relationship between the average percentage revision and STAR can be established.

35. We have identified the condition under which the X-11 with ARIMA extension is most likely to have less revision. This condition is that a time series is relatively volatile (ie. with relative large values of I/C and STAR) and strong in seasonality (ie. small I/S). In addition, we also have a better understanding of when the standard X-11 is likely to perform better when a suitable ARIMA cannot be found.

36. Our research suggests that future research for reducing the "end-weight" problem should be focused on:

1. Tuning or relaxing the current ARIMA model selection criteria or constructing appropriate criteria to identify a suitable model that still ensure a reduction in revisions.

2. Changing the standard X-11 default option if a suitable ARIMA model cannot be found. eg. using the standard X-11 when I/C and STAR is relatively small, otherwise, using the "airline" model.

3. Choosing an optimal length of data to balance the "global" nature of an ARIMA model with local dynamics at the end of series.

REFERENCES

Australian Bureau of Statistics (1999). SEASABS v2.0. Australian Bureau of Statistics: Knowledge based version of X-11.

Breiman L., Friedman J.H., Olshen R.A., and Stone, C.J., (1984). Classification and Regression Trees. Wadsworth International Group, Belmont CA. Chambers, J.M., and Hastie, T.J. (1991). Statistical Models in S, pg. 414.

Collett. D. (1991). Modelling Binary Data, London: Chapman and Hall.

Dagum, E.B. (1980). The X-11ARIMA Seasonal Adjustment Method. Ottawa: Statistics Canada, cat. 12-564E.

Dagum, E.B., Chhab, N. and Chiu, K. (1996). Derivation and Properties of the X-11ARIMA and Census X-11 Linear Filters. *Journal of Official Statistics*, 12, No. 4, 329-347.

Doherty, M. (1992). The Surrogate Henderson Filters in X-11, *Statistics New Zealand*, Working Paper.

Findley, D.F., Monsell, B.C., Bell, W.R., Otto, M.C. and Chen B. (1998). New Capabilities and Methods of the X-12-ARIMA Seasonal-Adjustment Program. *Journal of Business and Economic Statistics*, Vol. 16, No. 2., 127-177.

Gray, A. and Thomson, P. (1996a). Design of moving-average trend filters using fidelity, smoothness and minimum revisions criteria, Bureau of the Census, RR96/01

Gray, A. and Thomson, P. (1996b). Design of moving-average trend filters using fidelity and smoothness criteria in Vol 2: Time Series Analysis in Memory of E.J. Hannan. ed. P. Robinson and M. Rosenblatt. Springer Lecture Notes in Statistics 115, 205-219.

Gray, A. and Thomson, P. (1996c). On a family of moving-average trend filters for the ends of series. Proceedings of the American Statistical Association, Section on Survey Research Methods, 1996.

Henderson, R. (1916). Note on Graduation by Adjusted Average. Transactions of the American Society of Actuaries, 17, 43-48.

Ladiray, D. and Quenneville, B. (2001). Seasonal Adjustment with the X-11 method, New York: Springer Verlan, Lecture notes in statistics, 158.

McLaren, C.H. and Steel, D.G. (1999), Designing trend filters and rotation patterns for repeated surveys, School of Mathematics and Applied Statistics, University of Wollongong, Preprint No. 19/98.

Quenneville, B. and Ladiray, D. (2000) Locally Adaptive Trend-Cycle Estimation for X-11. ICES-II Proceeding, Buffalo June 2000.

Shiskin, J., Young, A. H. And Musgrave, J.C. (1967). The X-11 Variant of the Census Method II Seasonal Adjustment Program. Technical Paper 15, Bureau of the Census, U.S. Department of Commerce, Washington, D.C.

Statistics New Zealand, (2000). The Effect of ARIMA Forecasting on Revisions to Seasonally Adjusted Time Series, Statistics New Zealand, Research Report No. 9, July 2000, New Zealand.

Appendix A: The sources of time series and their characteristics

The table below summaries characteristics of the 820 series used in the analysis.

Table A1: The sources of time series and their characteristics

group	number of series analysed in group	period	data span	stock/flow
selected sub-sample[1]	97	mixed	various	various
Labour Force	258	monthly (229) quarterly (17)	23 yrs 15 yrs	stock
Retail	202	monthly	19 yrs (1 series 39 yrs)	flow
Building Activity	174	quarterly	17 - 46 yrs	flow
Employment and Earnings	67	monthly	18 yrs (1 series 12 yrs)	stock
Meat	42	monthly	22 yrs	flow
New Motor Vehicle Registration	27	monthly	34 yrs	flow
Profits	20	quarterly	16 yrs	flow
Stocks and Sales	15	quarterly	12 - 17 yrs	stock/flow
Job Advertisements	8	monthly	10 - 26 yrs	flow
Wine	7	monthly	24 - 26 yrs	flow
Total[2]	**820**			

Notes:

1. The selected sub-sample was formed by selecting series with a broad range of characteristics from the total set of analysed series.
2. Total excludes the selected sub-sample.

Appendix B: Asymmetric weights versus forecasts

Simple example for SeasonalxIrregular (SI) values of a particular month or quarter

Let SI_t for t=1,..,n (year) be the data.

Assume a symmetric 3X5 moving average used in the standard X-11 to produce the seasonal factors

$W_{-3,0}=1/15$, $W_{-2,0}=2/15$, $W_{-1,0}=3/15$, $W_{0,0}=3/15$, $W_{1,0}=3/15$, $W_{2,0}=2/15$, $W_{3,0}=1/15$

To get seasonal factors for S_{n-2}, S_{n-1}, S_n, X-11 uses asymmetrical weights that only involve the recent past but not the future. These are given by

n-2 (third last point)

$W_{-3,1}=4/60$, $W_{-2,1}=8/60$, $W_{-1,1}=13/60$, $W_{0,1}=13/60$, $W_{1,1}=13/60$, $W_{2,1}=9/60$

$S_{n-2}=W_{-3,1}xSI_{n-5} +W_{-2,1}xSI_{n-4} +W_{-1,1}xSI_{n-3} +W_{0,1}xSI_{n-2} +W_{1,1}xSI_{n-1} +W_{-2,1}xSI_n$

n-1(second last point)

$W_{-3,2}=4/60$, $W_{-2,2}=11/60$, $W_{-1,2}=15/60$, $W_{0,2}=15/60$, $W_{1,2}=15/60$

$S_{n-1}=W_{-3,2}xSI_{n-4} +W_{-2,2}xSI_{n-3} +W_{-1,2}xSI_{n-2} +W_{0,2}xSI_{n-1} +W_{1,2}xSI_n$

n (last point)

$W_{-3,3}=9/60$, $W_{-2,3}=17/60$, $W_{-1,3}=17/60$, $W_{0,3}=17/60$

$S_n=W_{-3,3}xSI_{n-3} +W_{-2,3}xSI_{n-2} +W_{-1,3}xSI_{n-1} +W_{0,3}xSI_n$

These asymmetric weights can also be looked at as forecasts of the SI's and then applying the symmetric moving average. For the case above we have

$SI_{n+1}=0.25xSI_n+0.25xSI_{n-1}+0.25xSI_{n-2}+0.25xSI_{n-3}$
$SI_{n+2}=0.25xSI_n+0.25xSI_{n-1}+0.25xSI_{n-2}+0.25xSI_{n-3}$
$SI_{n+3}=SI_{n-2}$

It can be noted that different weighting patterns may be used to different forecast horizons, and in the case of the standard X-11 the forecast is a little bit bizarre.

ARIMA modelling

An alternative to the asymmetric weights is to fit an ARIMA model to the SI_t's. Consider a simple ARIMA model given by

$$(1-B)SI_t=(1-\theta B)E_t$$

where B is the backward shift operator, and θ is an unknown parameter and $E_t \sim N(0, \sigma^2)$. In this case the SI's are forecasted using a linear combination of the historical SI's determined by the differencing and moving average parameter.

The symmetric moving average is then applied to the forecasted SI's. Because of the control over the differencing and moving average parameter a much larger range of filters are available. The ARIMA modelling case could be converted back to asymmetric filters. Generally the asymmetric ARIMA filters would go back a long way into the past (with small weights), but this could be controlled by using a "local" ARIMA model or using a purely autoregressive model.

Appendix C: ARIMA "Super Model"

The "super model" is an empirical ARIMA model with fixed parameters. It has been included in this paper to see how a static ARIMA model performs against the standard X-11. To estimate such a model a possible solution is to combine many series to find a more suitable general model that can capture collection wide characteristics that could not be detected in individual time series.

The form of the ARIMA "super model" is given by

monthly (0,1,13)(0,1,0) $(1-B)(1-B^{12})\log(O_t)=(1-\theta_1 B-\theta_2 B^2 \ldots -\theta_{12} B^{12}-\theta_{13} B^{13})E_t \quad E_t \sim N(0,\sigma^2_O)$

quarterly (0,1,5)(0,1,0) $(1-B)(1-B^4)\log(O_t)=(1-\theta_1 B-\theta_2 B^2-\theta_3 B^3-\theta_4 B^4-\theta_5 B^5)E_t \quad E_t \sim N(0,\sigma^2_O)$

The parameters were estimated by maximising the sum of log likelihoods over a few hundred time series simultaneously. They are given in the table below

Table A2: Super model coefficients

quarterly	lag 1	lag 2	lag 3	lag 4	lag 5		
	0.39627	0.14481	-0.0079364	0.66054	-0.19403		
monthly	lag 1	lag 2	lag 3	lag 4	lag 5	lag 6	lag 7
	0.32696	0.0043625	-0.019155	-0.011883	-0.017619	-0.015959	-0.021955
	lag 8	lag 9	lag 10	lag 11	lag 12	lag 13	
	-0.0010558	-0.023753	-0.0019553	-0.0080748	0.72490	-0.23656	

Appendix D: Automatic modelling X-12-ARIMA

Five default models

Models chosen by empirical research on a large selection of data. See (Dagum 1988) for further details.

Model 1 (0 1 1) (0 1 1) X $(1-B)(1-B^{12})\log(O_t)=u+(1-\theta_1 B)(1-\theta_2 B^{12})E_t$ $E_t \sim N(0,\sigma^2_o)$
Model 2 (0 1 2) (0 1 1) X $(1-B)(1-B^{12})\log(O_t)=u+(1-\theta_1 B-\theta_2 B^2)(1-\theta_3 B^{12})E_t$ $E_t \sim N(0,\sigma^2_o)$
Model 3 (2 1 0) (0 1 1) X $(1-\phi_1 B-\phi_2 B^2)(1-B)^2(1-B^{12})\log(O_t)=u+(1-\theta_1 B^{12})E_t$ $E_t \sim N(0,\sigma^2_o)$
Model 4 (0 2 2) (0 1 1) X $(1-B)^2(1-B^{12})\log(O_t)=u+(1-\theta_1 B-\theta_2 B^2)(1-\theta_3 B^{12})E_t$ $E_t \sim N(0,\sigma^2_o)$
Model 5 (2 1 2) (0 1 1) $(1-\phi_1 B-\phi_2 B^2)(1-B)(1-B^{12})\log(O_t)=(1-\theta_1 B-\theta_2 B^2)(1-\theta_3 B^{12})E_t$ $E_t \sim N(0,\sigma^2_o)$

[Note: X signifies mean to be estimated]

Estimation - exact maximum likelihood

Set of selection criteria

The following set of criteria are used based on forecasting performance, white noise properties of the residuals and cancellation of parameters on the moving average side with the differencing in the ARIMA model.

(A) within sample forecast error (<15% default)

(B) Box-Ljung Q statistic (>5% default)

(C) over differencing on MA parameters (default <0.9)

Options for "best" model

(A) first model, in order, that passes selection criteria

(B) best model in terms of within sample forecast error that passes selection criteria.

Appendix E: Revision measure used in paper

The generic formula for a lag k (k = 0, 1, 2, ..., m) average absolute percentage revision from start point t_0 (t_0 is greater than the start date of series) to the end point T_0 (T_0 is less than the end date of series) is defined as:

The mean absolute percentage revision R_k on specific lags is calculated as

$$R_k = \frac{100}{T_0 - t_0 + 1} \sum_{t=t_0}^{T_0} \frac{\left| A_{t|T} - A_{t|t+k} \right|}{A_{t|T}}$$

k ($0 \leq k \leq 36$ for monthly series or 12 for quarterly series) is the lagged period, where k=0 represents the 1st estimate, k=1 represents the 2nd estimate etc;

T is the end date of the original series;

t_0 is the start date of the original series minus seven years data (t_0 = SD - 84 for monthly series or SD - 28 for quarterly series;

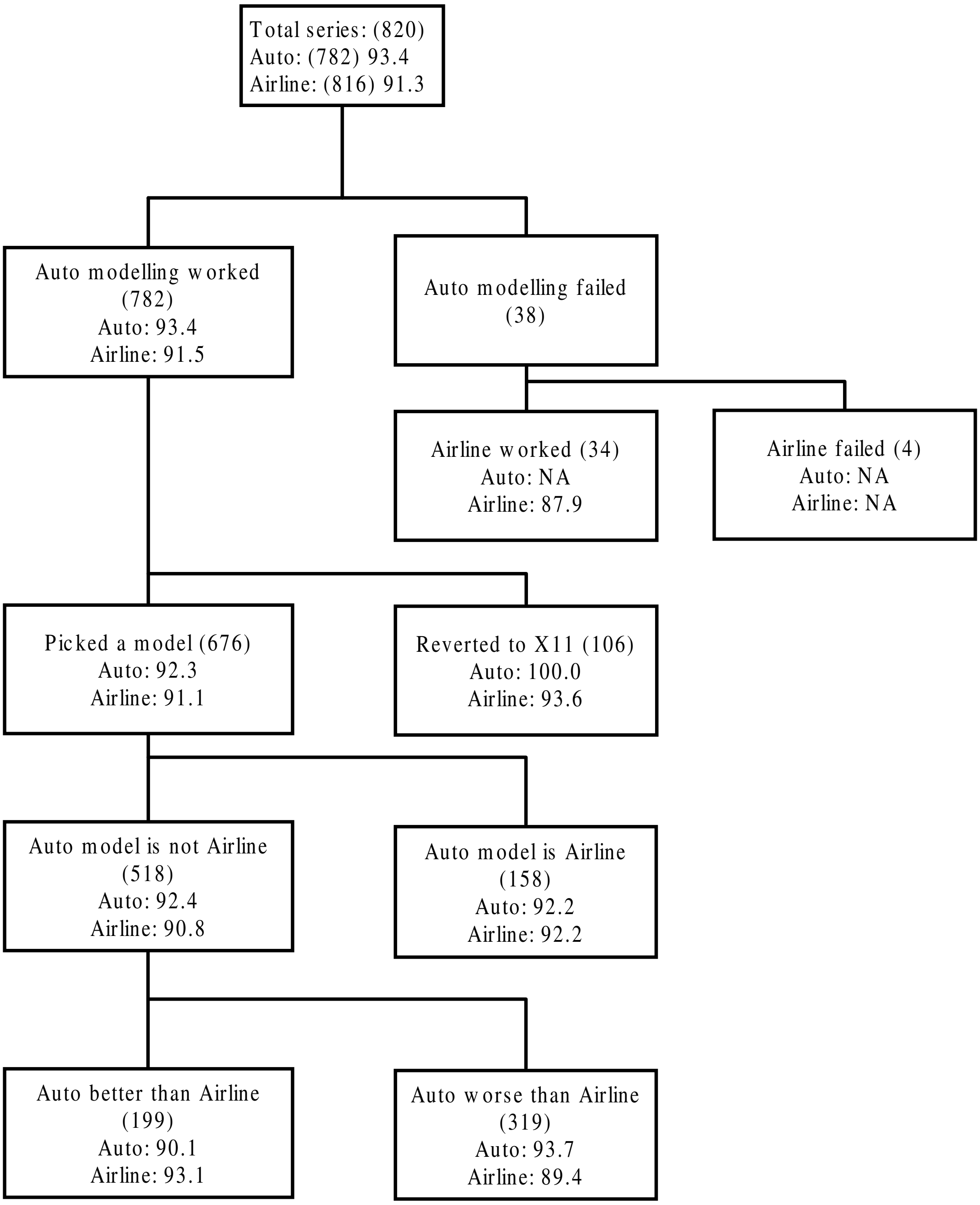
T_0 is less than or equal to T minus three years data ($T_0 \leq T - 36$ for monthly series or T - 12 for quarterly series);

$A_{t|t+k}$ is the lag k estimate of concurrent seasonally adjusted series at date t (eg. if k=0 and t=Aug 98 it represents the 1st estimate of Aug 98), the value of the seasonally adjusted series at time t calculated from the start date of the series up to T_0 + k (when t = T_0).

$A_{t|T}$ is the "final" value of the seasonally adjusted series at time t, the value of the seasonally adjusted series at time t calculated from the start date of the original series to the end date T;

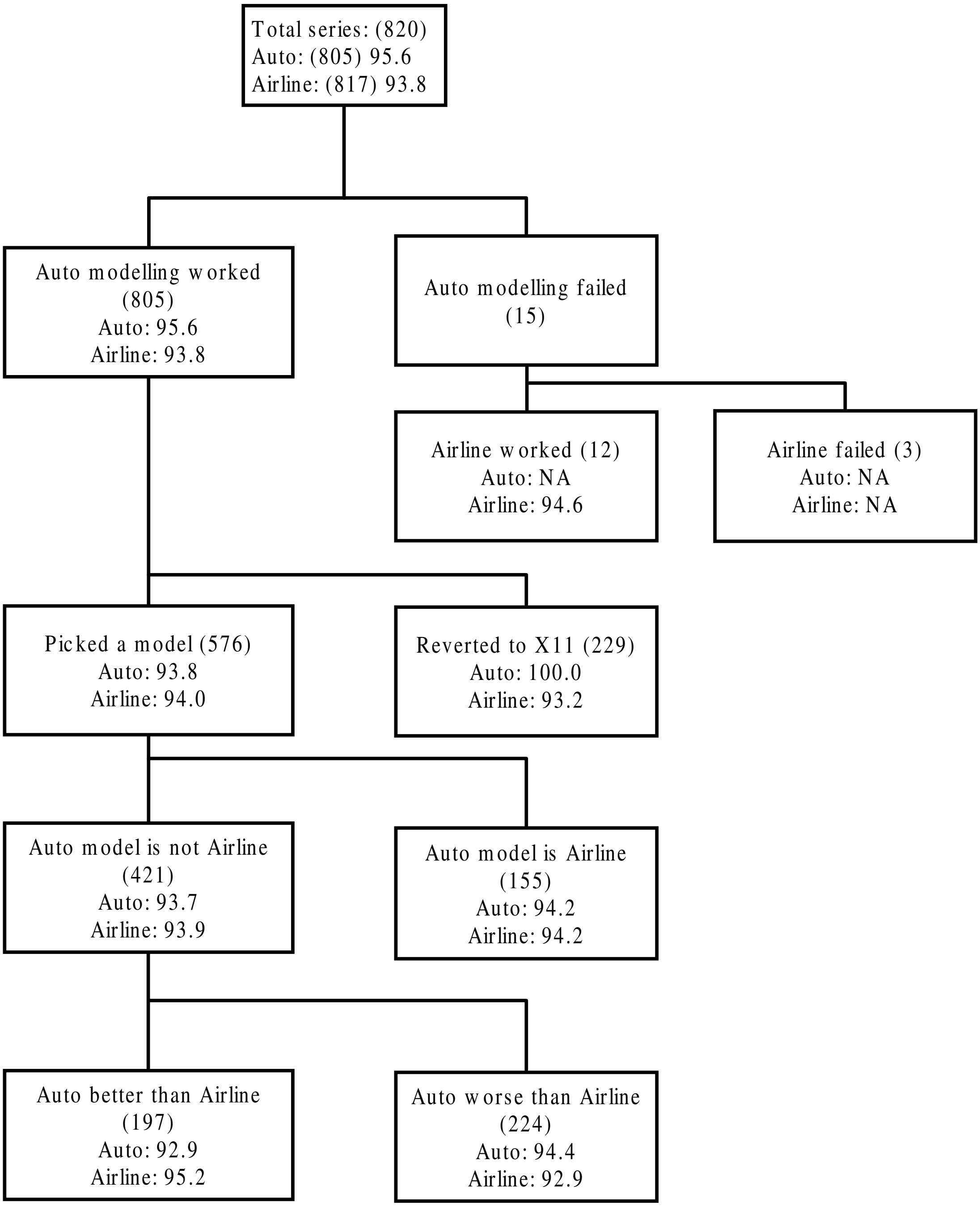
Appendix F: Diagram showing the performance of the automatic model selection within X-12-ARIMA

Figure A1: Automatic ARIMA modelling with outliers turned off.



Note: the number in a pair of brackets is the number of series. The figure against a particular model is the RR of the model against the standard X-11.

Figure A2: Automatic ARIMA modelling with outliers turned on.



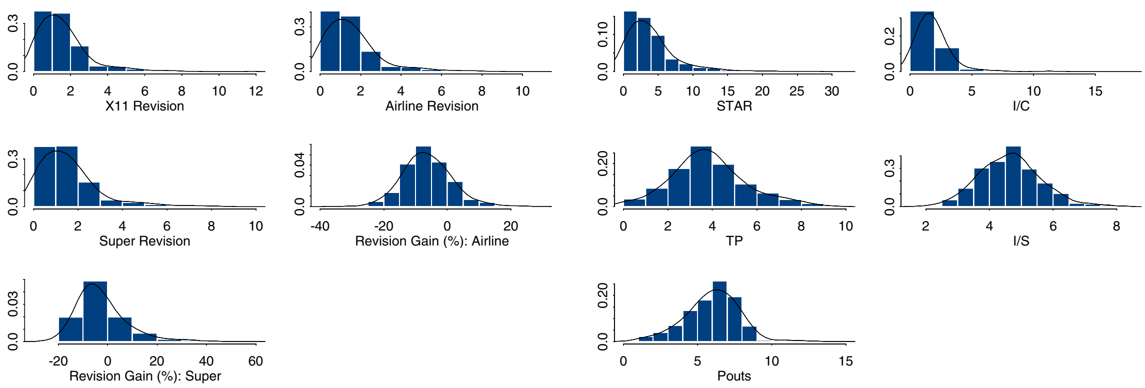
Note: the number in a pair of brackets is the number of series. The figure against a particular model is the RR of the model against the standard X-11.

Appendix G: Statistical Analysis

To explore the possible relationship in predicting revision size with the statistical measures, we have analysed five selected measures (STAR, I/C, I/S, TP and Pouts) against two revision measures, including R_0 (mean absolute percentage revisions), and RR (revision ratios between the R_0 from X-11 with ARIMA extension against the R_0 from the standard X-11).

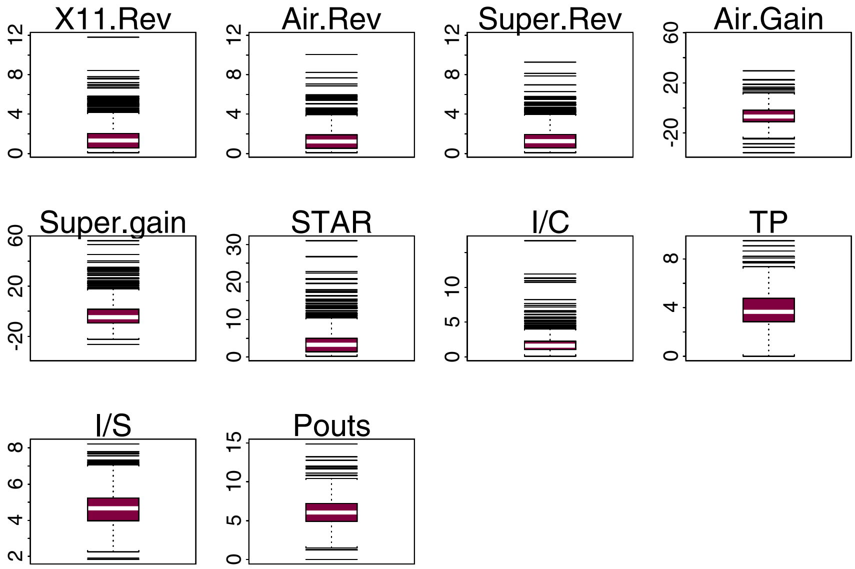
0. Descriptive Statistics

Figure A3: Empirical distributions of measures: R_0, Revision Gains (%), STAR, TP, I/C, I/S, and Pouts



The skewed distributions of R_0s, I/C and STAR indicate a logarithmic transformation is required when a linear regression analysis is performed.

Figure A4: Box plots of R_0, Revision Gains (%), STAR, TP, I/C, I/S, and Pouts



To compare the distribution properties of the two ARIMA models against the standard X-11, we create a binary response variable, Y as:

$$Y = \begin{cases} 0 \text{ (FALESE)} & \text{if RR} > 100\text{, ie. the revision of X-11 is smaller} \\ 1 \text{ (TRUE)} & \text{if RR} < 100\text{, ie. the revision of X-11 is larger.} \end{cases}$$

Figure A5: Box plots - Airline vs X11 against STAR, TP, I/S, I/C, and Pouts

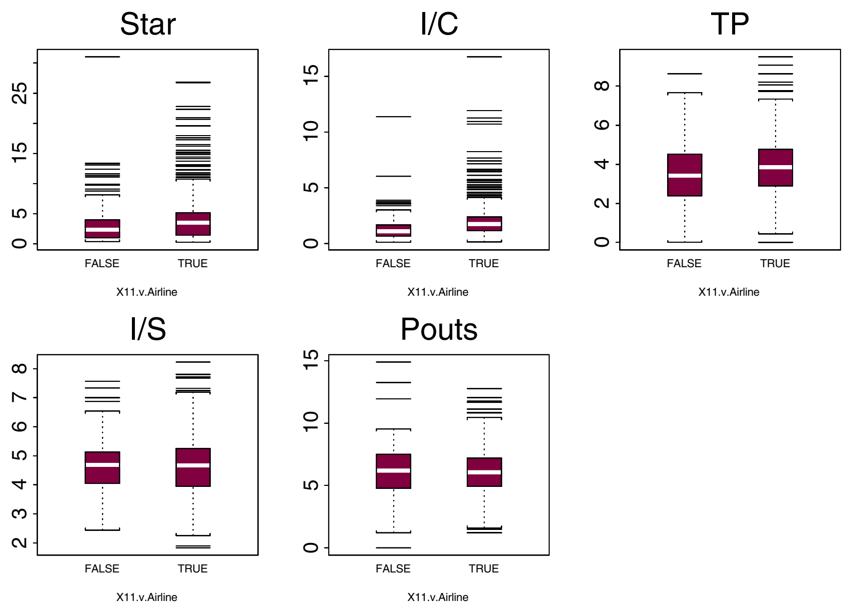


Figure A6: Super vs X11 against STAR, TP, I/S, I/C, and Pouts

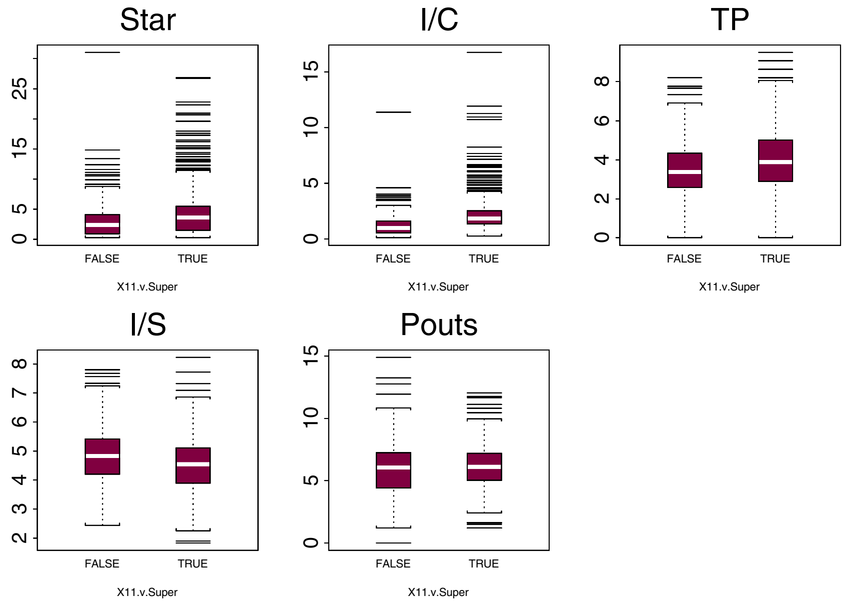


Figure A5 and A6 show that the medians, and inter-quartile ranges of Pout of a FALSE and TRUE pair are more or less the same for both the "airline" and "super" models. This indicates that Pout is unlikely to have any effect on revision gain. However, the medians and inter-quartile ranges of STAR and I/C have distinguishable differences. This indicates that STAR and I/C may have certain explanatory power.

We use three regression methods (1) binary tree recursive partition regression (2) simple linear regression (with non-linear transformation) and (3) logistic regression to explore and test the relationships between revision measures (R_0, RR or (RR-100)) and the five statistical measures.

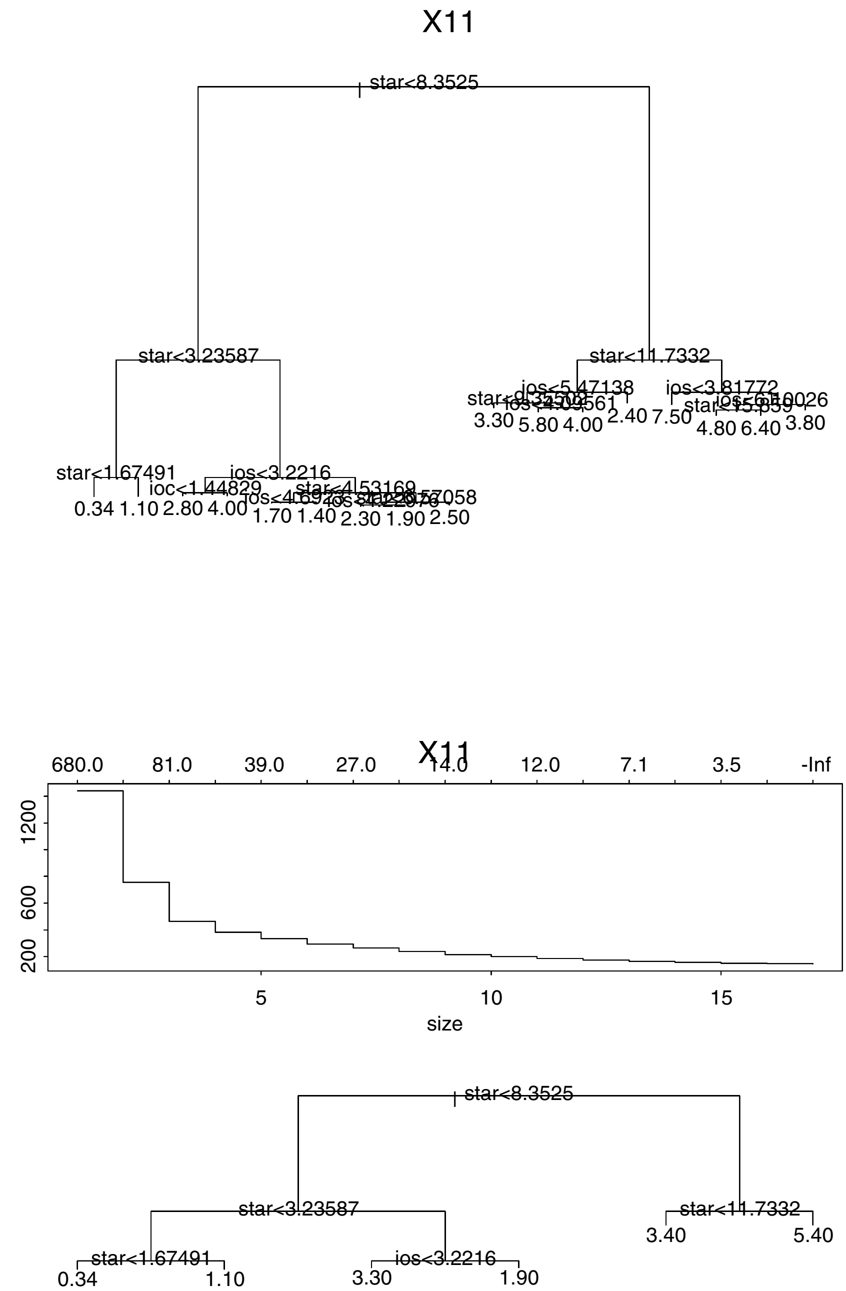
1. Recursive Partition Regression Tree Analysis

A tree based recursive partition regression is a statistical technique to explore the possible relationships between a dependent variable to its explanatory variable (Breiman et. al, 1984). Our primary purpose is to explore the possible relationship between**:** (1) revision R_0 and (2) the binary response variable Y with the five statistical measures.

Revision measure R_0:

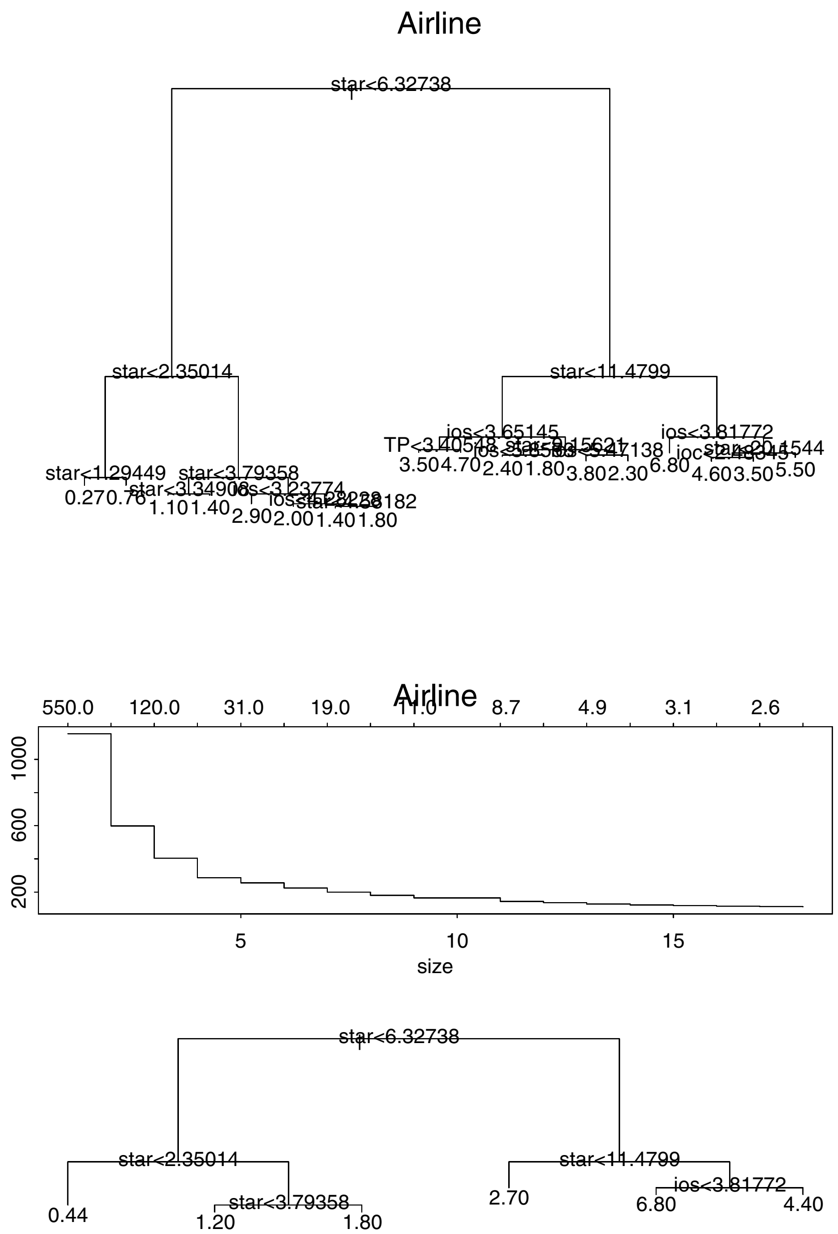
The following tree analyses used the revision measure R_0 from the standard X-11, X-11 with "airline" and "super" models respectively as the dependent variable and the five statistical measures as independent variables. The primary split on STAR suggests that the STAR value of a series is the best possible measure to explain the size of revisions.

Figure A7: X-11 R_0 regression tree



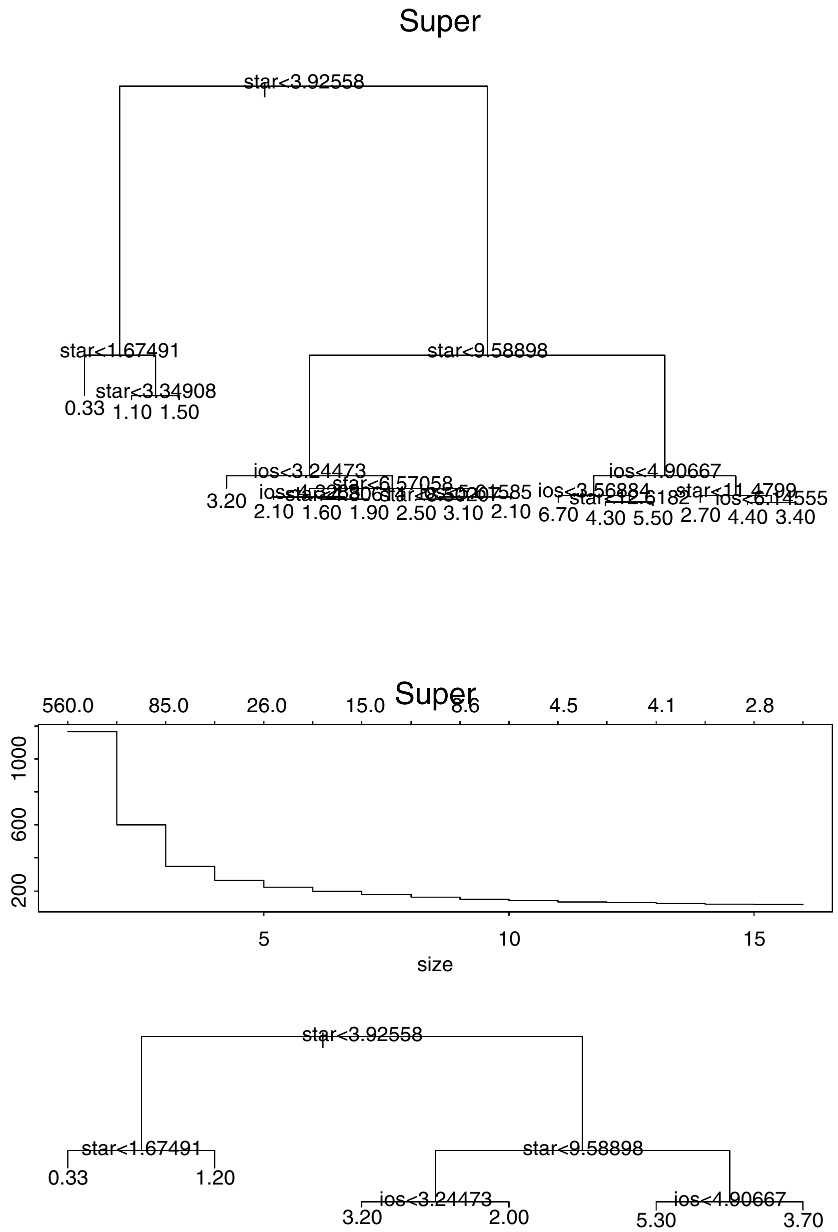
Note: the length of a vertical branch pair indicates the relative explanatory power of the split.

Figure A8: Airline model R_0 regression tree



Note: the length of a vertical branch pair indicates the relative explanatory power of the split.

Figure A9: Super model R_0 regression tree

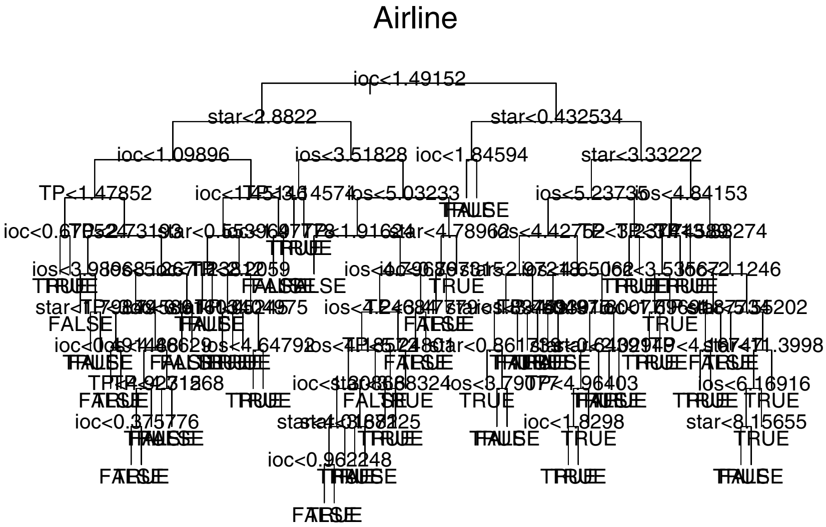


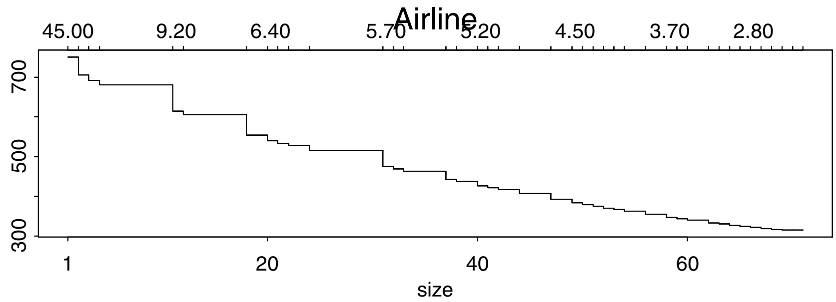
Note: the length of a vertical branch pair indicates the relative explanatory power of the split.

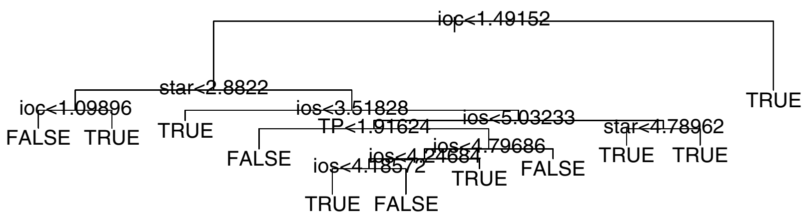
Binary response variable Y:

The following tree analyses use the binary response variable Y (i.e., the TRUE and FALSE factor variables) from Airline and Super model respectively as dependent variable and the five statistical measures as independent variables. The primary splits on I/C and then STAR suggest that the I/C and STAR values of a time series are the best possible explanatory variables to explain the possibility that X-11 with ARIMA extension produces less revisions than the standard X-11.

Figure A10: Airline model Y regression tree

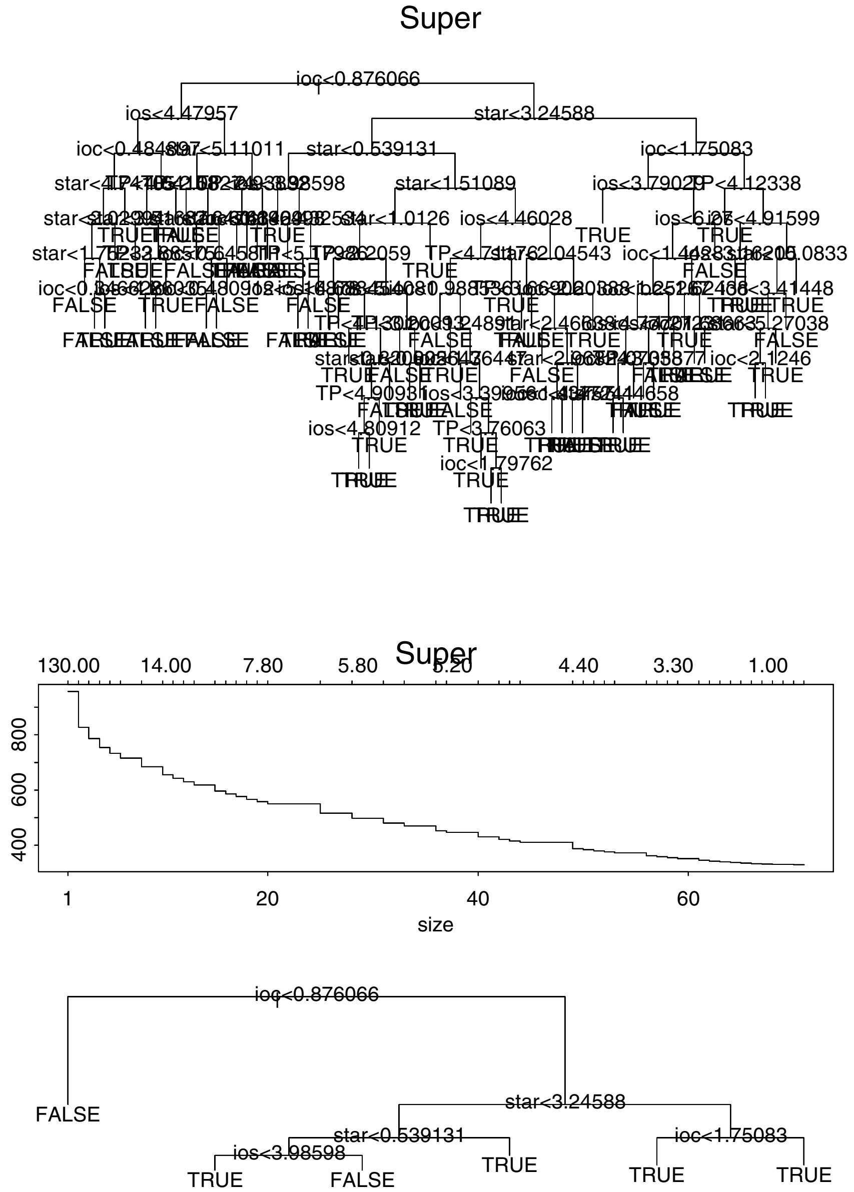






Note: the length of a vertical branch pair indicates the relative explanatory power of the split.

Figure A11: Super model Y regression tree



Note: the length of a vertical branch pair indicates the relative explanatory power of the split.

2. Linear regression analysis

Linear regression models (with log-trasformation for some measures) are used to explore the relationships between the five statistical measures and three revision measures, including revision R_0 , revision ratios RR (or revision gains). The skewed distributions of R_0, STAR and I/C suggest non-linear transformations are needed. We take a log-transformation for R_0, STAR and I/C in the linear regression analysis.

Revision R_0:

Table A3 shows the results of log R_0, which are derived from the standard X-11, X-11 with "airline" and "super" models, regressed on log(STAR), log(I/C), I/S, TP and Pout. All estimated coefficients are significant at the 5% level and the fitness measure R^2s show that the model adequately explains the variations of R_0 from the five explanatory measures.

Table A3: Regression model R_0 = a + b log(STAR) + b log(I/C) + c (I/S) + d (TP) + e (Pout)

	R_0 of X11		R_0 of Airline model		R_0 of Super model	
	coef. value	t-value	coef. value	t-value	coef. value	t-value
(Intercept)	0.3016	5.5605*	0.2130	4.0529*	0.1451	2.8533*
log(STAR)	0.9497	121.0524*	0.9407	123.7473*	0.9391	127.6983*
log(I/C)	-0.0384	-3.6700*	-0.0673	-6.6378*	-0.1127	-11.4862*
I/S	-0.2112	-26.1683*	-0.2083	-26.6345*	-0.1812	-23.9476*
TP	0.0151	3.4165*	0.0152	3.5561*	0.0171	4.1286*
Pouts	-0.0418	-9.7437*	-0.0373	-8.9757*	-0.0388	-9.6388*
R^2	0.9566		0.9581		0.9602	

Note that the sign of the estimated coefficients are the same across each R_0. This indicates that the five explanatory measures have the same effect on R_0, although with different magnitudes. STAR, TP positively contribute to the revision measure R_0, whilst I/C, I/S, and Pouts negatively contribute to the revision measure R_0. These can be interpreted as

- the more volatile (STAR), the larger the revision;
- the more turning point (TP), the larger the revision;
- the less movement in trend than in irregular (I/C), the smaller the revision;
- the less movement in seasonal than in irregular (I/S), the smaller the revision;
- the more data points modified by X-11 as outliers(Pouts), the smaller the revision.

All the above results are intuitive and provide insightful information about how the five measures contribute to R_0. Although all five measures have a significant effect on revision R_0, STAR is the most dominant explanatory measure.

Revision gain (RR - 100):

Table A4 and A5 show the results of regression analyses of revision gain, RR - 100, on four statistical measures for the "airline" and "super" models. The results demonstrate that I/C is the most significant measure to explain the variation of revision ratios RR (or revision gains). However, the R^2 statistics indicate that the models used do not necessary fit well.

Table A4: Linear regression using revision gain as response variable: Airline model

Fitted Model	Coefficient (t Value)	R^2
revision gain =a+b log(star)	a: -5.061 (-12.823*) b: -1.191 (-4.158*)	0.0218
revision gain =a+b log(I/C)	a: -5.135 (-16.974*) b: -2.873 (-7.649*)	0.0702
revision gain =a+b (I/S)	a: -5.218 (-3.874*) b: -0.222 (-0.738)	0.0008
revision gain =a+b log(star) +c log(I/C)	a: -4.264 (-10.715*) b: -0.935 (-3.342*) c: -2.714 (-7.212*)	0.0834
revision gain =a+b log(star) +c log(I/C) + d (I/S)	a: -4.4736 (-3.4000*) c: -2.7215 (-7.1744*) b: -0.9350 (-3.3413*) d: 0.0458 (0.1672)	0.0835

Figure A12: Airline model - STAR contribution to RR - 100

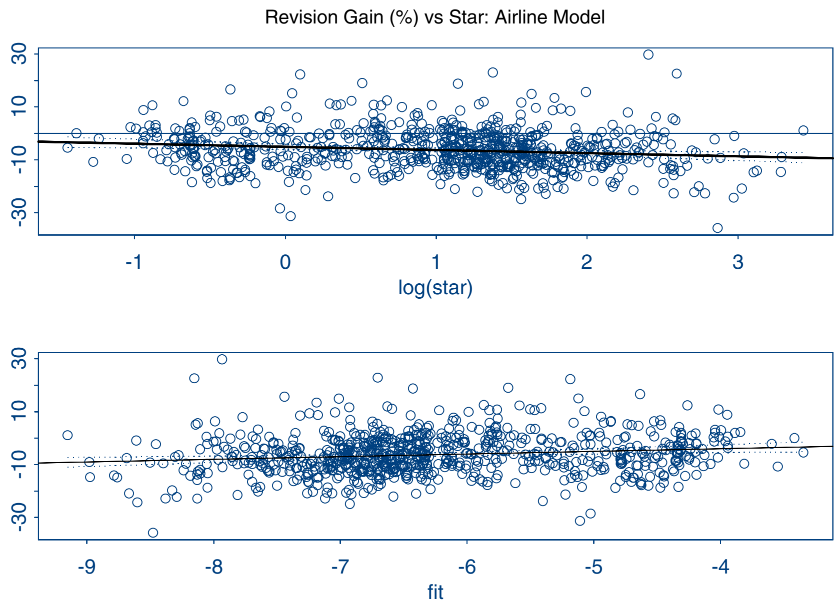


Figure A13: Airline model - I/C contribution to RR - 100

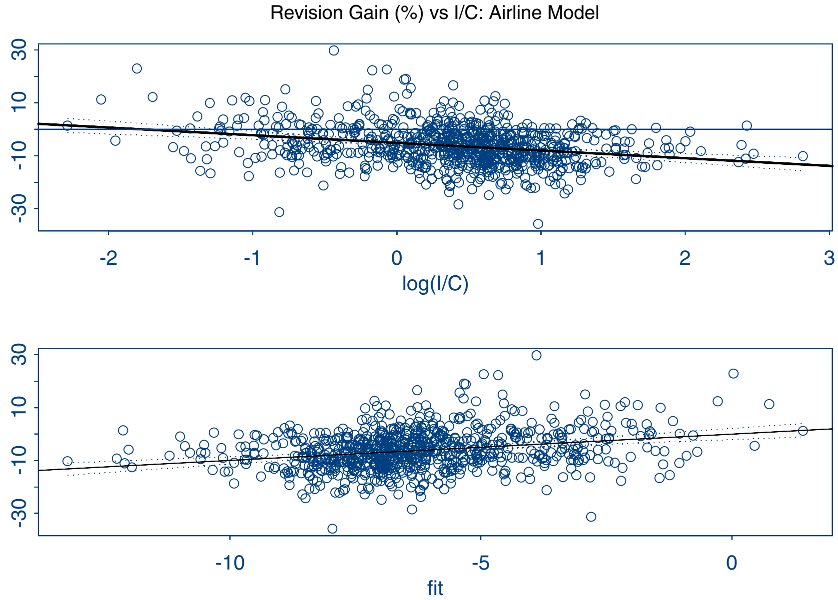


Figure A14: Airline model - I/S contribution to RR - 100

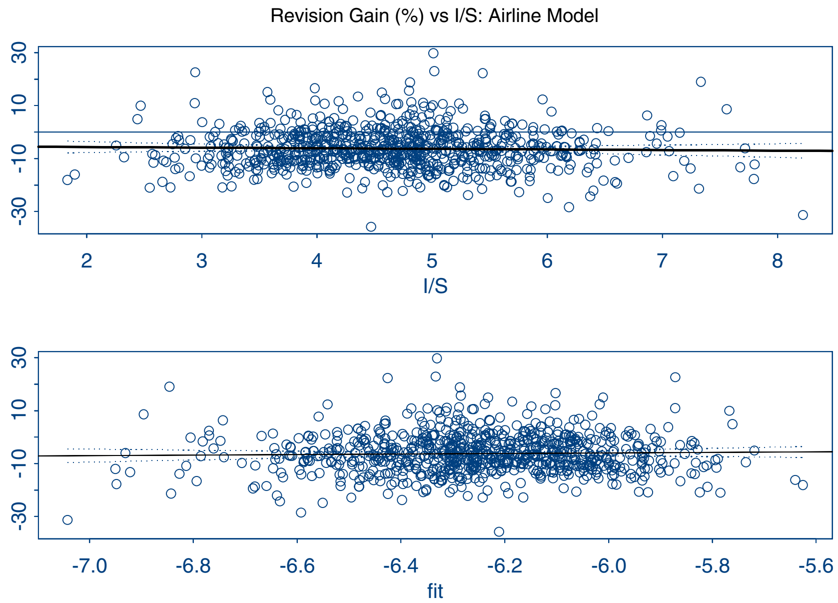


Figure A15: Airline model - Estimated revision gain

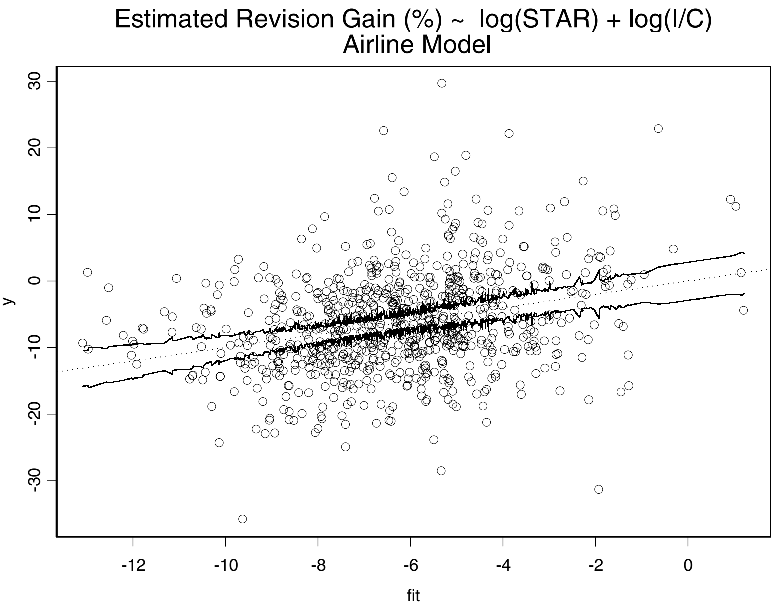


Table A5: Linear regression using revision gain as response variable: Super Model

Fitted Model	**Coefficient (t Value)**	**R^2**
revision gain =a+b log(star)	a: -0.857 (-1.609) b: -1.625 (-4.203*)	0.0223
revision gain =a+b log(I/C)	a: 0.337 (0.916) b: -7.265 (-15.898*)	0.246
revision gain =a+b (I/S)	a: -12.806 (-7.194*) b: 2.220 (5.925*)	0.043
revision gain =a+b log(star)+c log(I/C)	a: 1.228 (2.531*) b: -0.955 (-2.802*) c: -7.101 (-15.483*)	0.254
revision gain =a+b log(star)+c log(I/C)+d (I/S)	a: -12.175 (-8.000*) b: -0.981 (-3.031*) c: -7.595 (-17.310*) d: 2.928 (9.240*)	0.328

* indicates t statistics are significant at the 5% level.

Figure A15: Super model - STAR contribution to RR - 100

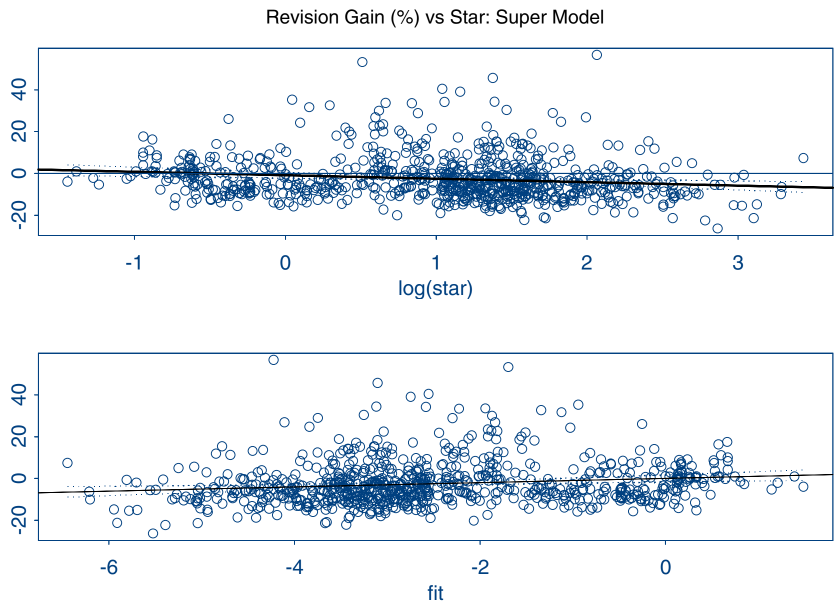


Figure A16: Super model - I/C contribution to RR - 100

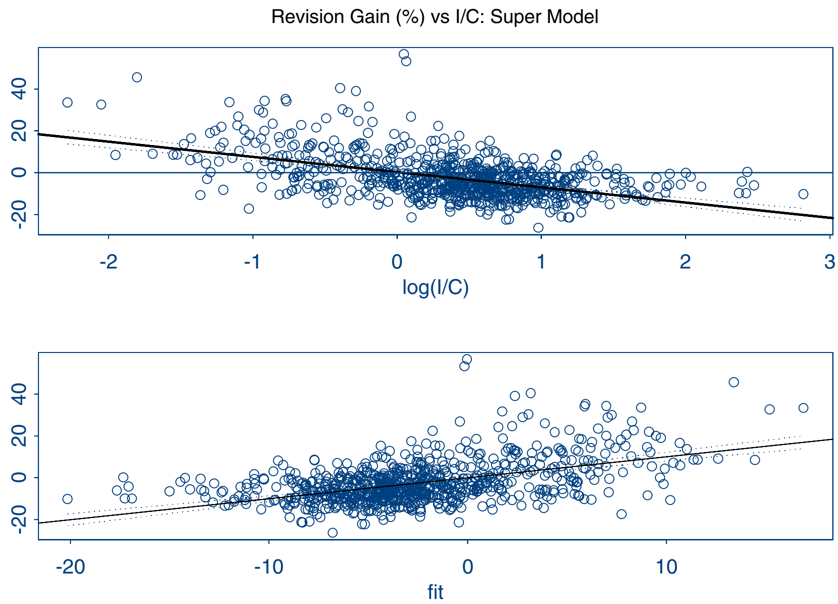


Figure A17: Super model - I/S contribution to RR - 100

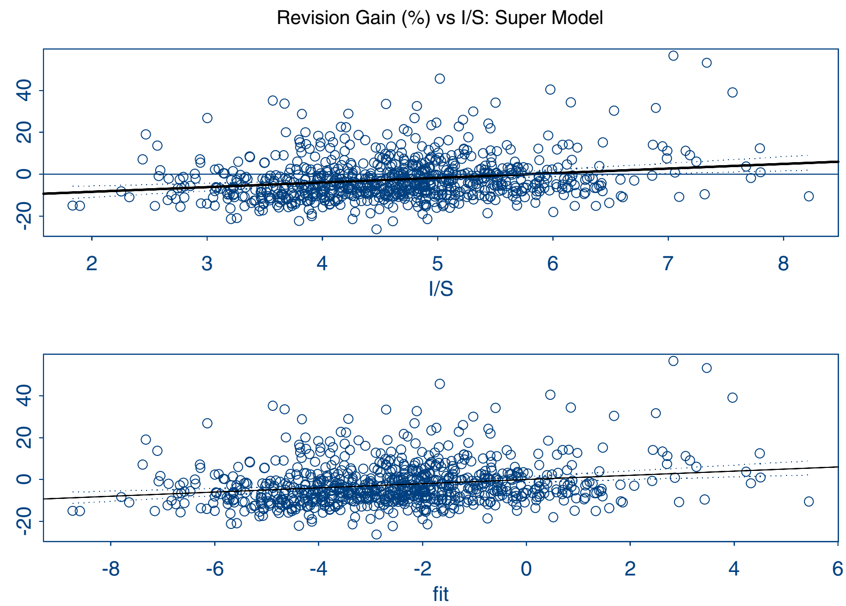
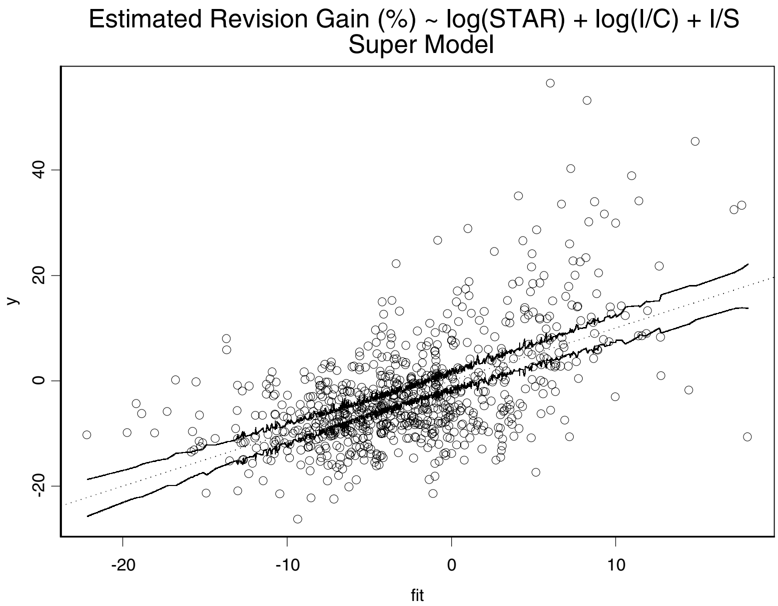


Figure A18: Super model - Estimated revision gain



From the above linear regression analyses, we conclude that the "airline" and "super" models are likely to have less revisions than the standard X-11 has when I/C and STAR increase. The positive coefficients of I/S show that the I/S has a positive effect (although it is not significant for the "airline" model). This indicates that the implied seasonal component model embedded in the both ARIMA models may not perform as well as the X-11's seasonal asymmetric filter when a series has relatively less seasonal pattern changes comparing to its irregular. The significant positive I/S coefficient for the "super" model also shows this static model performs worse than the more dynamic "airline" model for the seasonal component. In comparison with the standard X-11 on the 820 series, the net gain from an ARIMA model comes from better trend forecasts. This gain is usually larger than the loss from its poor seasonal forecasts that occurs when the volatility of a series increases.

3. Logistic Regression Analysis

Logistic regression analysis is often used to investigate the relationship between a discrete response variable and a set of explanatory variables (Collett,1991). For the created binary response variable, Y (see definition on page 25), a logistic regression model with the 4 statistical explanatory measures allow us to robustly test which explanatory measures significantly affect the probability of the response variable.

Statistical results of logistic regression models are reported in Table A6, which suggest that I/C and STAR significantly affect the probability of the binary response variable Y.

Table A6: Logistic Regression of binary Y on four explanatory measures

Method	Intercept (|t|)	Log(STAR) (|t|)	TP (|t|)	I/S (|t|)	Log(I/C) (|t|)
Airline	1.3046 (2.59)*	0.2954 (2.84)*	0.0548 (0.91)	-0.1155 (1.18)	0.9142 (6.53)*
Super	2.6411 (5.18)*	0.4339 (4.48)*	0.0504 (0.91)	-0.6105 (5.98)*	1.7503 (10.98)*

* indicates statistics are significant at the 5% level.

From the above logistic regression analyses, the positive coefficients of log(STAR) and log(I/C) indicate that the two ARIMA models are likely to have a higher probability of smaller revision than the standard X-11 has when I/C and STAR increase. The negative coefficients of I/S indicates that the two ARIMA models are likely to have a lower probability of smaller revision than the standard X-11 has when I/S (although it is not significant for the "airline" model) increases. These results are consistent with the linear regression (with transformation) analyses.