BUREAU OF THE CENSUS

STATISTICAL RESEARCH DIVISION REPORT SERIES

SRD Research Report Number: CENSUS/SRD/RR-88/14

OUTLIERS IN TIME SERIES

by

J. Peter Burman
University of Kent
at Canterbury
Canterbury, England

Mark C. Otto
Statistical Research Division
Bureau of the Census
Washington, D.C. 20233

This series contains research reports, written by or in cooperation with staff members of the Statistical Research Division, whose content may be of interest to the general statistical research community. The views reflected in these reports are not necessarily those of the Census Bureau nor do they necessarily represent Census Bureau statistical policy or practice. Inquiries may be addressed to the author(s) or the SRD Report Series Coordinator, Statistical Research Division, Bureau of the Census, Washington, D.C. 20233.

Recommended by: Nash J. Monsour

Report completed: May 18, 1988

Report issued: May 20, 1988

CENSUS BUREAU RESEARCH PROJECT

OUTLIERS IN TIME SERIES

by

J. Peter Burman

and

Mark C. Otto

* J. Peter Burman is an honorary professor at the Mathematical Institute, University of Kent, Canterbury, Kent, CT2 7NF, England and Mark C. Otto is a Mathematical Statistician with the Statistical Research Division of the Census Bureau, Washington, DC 20233. This paper is based upon work supported by the National Science Foundation under grant SES 84–01460, "On–site Research to Improve the Government–Generated Social Science Data Base." The research was conducted during September–December 1984 and July–September 1985, at the U.S. Bureau of the Census while the authors, respectively, Research Fellow and Associate in the American Statistical Association/Census Bureau Research Program. The program is supported by the Census Bureau and through the NSF grant. Any opinions, recommendations expressed here are those of the authors and do not necessarily reflect the views of the National Science Foundation or the Census Bureau. The authors would like to thank William R. Bell and David F. Findley for their support and helpful suggestions.

ABSTRACT

We study the effects of outliers on short–term forecasting errors and on autoregressive–integrated moving–average (ARIMA) model characteristics such as the Ljung–Box statistics and estimates of the seasonal moving–average parameter. We have fitted sixty Census Bureau monthly time series with ARIMA models, identified additive point outliers, and sought their external causes. Modification of outliers was found not increase the mean absolute forecasting error (of one, two, and three steps–ahead forecasts over the last three years of the data) in 31 out of 44 series with identified outliers. We also discuss consequences of different methods of outlier modification, choice of outlier identification threshold, and effects on the seasonal adjustment of time series.

KEYWORDS: Time Series; ARIMA model; Outlier; Signal Extraction; Forecasting; Seasonal adjustment; Census X–11.

1. OUTLINE OF PROBLEM

It is often taken for granted that modification of outliers improves the forecasting performance of a time series model, because:

1. It is believed that outliers occur at places where the process generating the series has temporarily broken down, so that modification of outliers is needed to compensate for this in the forecasts calculations.
2. If so, modification should bring the parameter estimates closer to their true values, resulting in improved forecasts.

In practice, this is far from clear. The threshold used to define outliers is arbitrary and for a given threshold, when more data are used, the outlier set is often not the same. Also, if the innovations of the model–generating process are not Normal (but from a fat–tailed distribution), the threshold may be detecting outliers more often than intended. In this paper we study the empirical effect of outlier modification on post–sample forecast errors — not those within sample, which are bound to be reduced.

Different types of outliers were defined by Fox (1972), and Denby & Martin (1979): additive outliers (AO), which affect only a single observation, and innovative outliers (IO), where an unusual innovation in the generating process affects all later observations. Further types of outliers, which Pierce (1987) calls "mixed", arise when only one characteristic of the series (e.g. trend or seasonal component) is changed by the innovation.

In the context of seasonal adjustment, automatic outlier identification and modification has been practiced for many years, e.g., in the Census X–11 program (Shisken, Young, and Musgrave 1963). When seasonally adjusting with X–11, a series is decomposed into trend, seasonal, and irregular components, regardless of the series generating process. Outliers are identified from the irregular series, using a threshold which is a multiple of the root mean square of an appropriate span of the series. But the default option, which is commonly used, leads to a large proportion of outliers, perhaps 10 percent of the observations, and it seems unlikely that the generating process has broken down so often. A further arbitrary feature is the need to choose a moving average for the seasonal component (e.g. [3] [5] or [3] [9]), which determines what the irregular series looks like.

Hillmer, Bell, and Tiao (1983) (hereafter called HBT) and Bell (1983) applied a more rigorous approach to the treatment of outliers in model–based seasonal adjustment. They showed how the residual errors of the model can be used, with a given threshold, to identify outliers of different types. Regression estimates of their magnitudes provide starting values in an extended model, which includes dummy variables to represent the outliers. Burman (1983) indicates, for the simpler models, the link with the traditional outlier–detection method of X–11.

In this paper, the full HBT method is called "simultaneous estimation" of outliers: it is an extension of Intervention Analysis – see Box and Tiao (1975). In our view, unless external causes can be identified, this extension suffers from the conceptual difficulty that the hypothesis being tested has an indefinite number of parameters. An alternative considered here is to use the regression estimates of the outliers, as they stand, to modify the series, and then refit the original model.

Another question is the effect of outliers on the quality of a seasonal adjustment. A number of authors have suggested criteria for the quality of a seasonal adjustment procedure, e.g., maximum smoothness of the seasonal component, sensitivity to change, minimizing average revisions over the last few years — see HBT and Burman (1980). But there is no agreement over the ranking of these criteria; in particular, for comparison of revisions between X–11 and model–based adjustments, the methods are targeting different final adjusted series. So we side–step the problem in this paper by concentrating on short–term forecasting performance of a fitted model, since this is often the main purpose of seasonal adjustment.

2. IDENTIFICATION OF OUTLIERS

HBT showed that outliers could be identified in a way not involving decomposition into trend, seasonal, and irregular. Assume that the current observation z_t (if necessary, transformed) can be expressed as a weighted average of past values plus an innovation which is Normally distributed white noise: i.e. $z_t = a_t - \pi_1 z_{t-1} - \pi_2 z_{t-2} \cdots$ or $(1+\pi_1 B + \pi_2 B^2 \cdots)z_t = a_t$, where $Bz_t = z_{t-1}$. We write this as: $\pi(B)z_t = a_t$. The series of coefficients is infinite, unless the model is purely autoregressive.

Following Denby and Martin (1979), HBT classify two types of outliers: additive (AO) and innovative (IO). An IO at time t_0 affects the innovation a_t and is built into the future level, slope, and seasonality of the series; it is estimated directly from the innovation. An isolated AO at time t only affects z_t itself and not future values z_{t+k}. But it produces a large forecast error $\hat{a}_t$ and this affects future forecast errors $\hat{a}_{t+k}$ to a diminishing extent. Bell (1983)—extending the outlier procedure in HBT—identifies two other types of outliers affecting the z_t series, changes in level and changes in the seasonal pattern. Changes in level can be treated as isolated AO's in the differenced series, Δz_t, and changes in the seasonal pattern can be treated as AO's in the seasonally differenced series, $\Delta_{12} z_t$. HBT show that the best estimate of an AO outlier magnitude, $\alpha(t)$, in a series of length n is obtained from a linear regression of the $\hat{a}_t$, $\hat{a}_{t+1}$, $\cdots$, and $\hat{a}_n$ on the π–weights $\alpha(t) = (\hat{a}_t+\pi_1\hat{a}_{t+1}+\pi_2\hat{a}_{t+2}+ \cdots \pi_{n-t}\hat{a}_n)/(1+\pi_1^2+\pi_2^2+ \cdots \pi_{n-t}^2)$. Now, define the numerator as $\hat{I}_t$ and the denominator as w_{n-t}: if the threshold for an outlier is a constant multiple of the standard error of $\hat{I}_t$, it tapers quite sharply downwards as t approaches n. If the $\hat{a}_t$ are estimates of independent Normal variates, the $\hat{I}_t$ are Normal and can be given a *t*–test. The $\hat{a}_t$ are unknown beyond t = n, but their expected values are 0. So we can write $\hat{I}_t = \pi(F)\hat{a}_t$, where $F\hat{a}_t = \hat{a}_{t+1}$, the inverse operator to B. In the discussion on HBT, the first author pointed out that, providing t is not too small, $\pi(B)z_t = \hat{a}_t$, and $\hat{I}_t$ can be written in a symmetric form: $\hat{I}_t = \pi(F)\pi(B)z_t$. This is to be understood as a doubly infinite moving average, in which missing values are replaced by forecasts and backcasts, and the technique of Signal Extraction will evaluate it—see Burman (1980).

If (n–t) is not small (in practice more than 2 years), w_t is almost equal to $w_\infty = \sum_{i=0}^{\infty} \pi_i^2$. So the threshold for an outlier is proportional to $w_\infty^{1/2}$ in the central part of the series. By the symmetry of $\hat{I}_t$, the threshold must also taper near the start of the series. To see how this happens, we note that the same model can be expressed in terms of the backcast errors, e_t: $\pi(F)z_t = e_t$ and hence $\hat{I}_t = \pi(B)\hat{e}_t$. So, for small t, the threshold is proportional to $w_t = \{\sum_{i=0}^{t} \pi_i^2\}^{1/2}$.

If several outliers are tentatively identified, a multiple regression gives estimates which allow for interactions between any that are close together. For example, if there are outliers at t and (t+2), the forecast errors have not recovered from the first shock before the second one is upon them. Interaction also occurs for monthly seasonal series with outliers at t and (t+12). For detailed formulae, see the Appendix.

We automatically identify and estimate modifications for outliers using an iterative process similar to HBT:

(1) Calculate a robust version of the root mean square error (RMSE) of $\hat{a}_t$.

(2) Estimate $\hat{I}_t$ and the the threshold values ($w_{n-t}^{1/2}$·RMSE) and identify, using a *t*–statistic, all those values beyond their thresholds.

(3) Re–estimate the effects of the all the outliers—including the newly identified outliers—using a multiple regression. Then revise $\hat{I}_t$, the RMSE, and the threshold values, using the residuals from the outlier regression. If this is the first pass, replace the robust RMSE by the regression RMSE.

(4) Repeat (2) and (3) until no more outliers are found.

(5) Do backward elimination (Draper and Smith 1981, pp. 305–307) of outliers until only outliers with *t*–statistics over the threshold remain.

(6) Estimate modifications for time points that have $\hat{I}_t$ values between a partial outlier threshold and the full outlier threshold.

As does Bell (1983), we use a robust version for the initial RMSE estimate (in (1)). W.S. Cleveland in his discussion of HBT pointed out that the initial estimate of σ_a will be biased upwards if the series contains several large outliers. Practically, we may fail to detect any outliers on our initial pass over $\hat{I}_t$ because of this misspecification of the RMSE. Bell adopted Cleveland's suggestion to use $1.48^*\text{median}|\hat{a}_t|$ which is based on the relation between the quartiles and the standard deviation of the Normal distribution. We use this only for the initial pass over $\hat{I}_t$, then after an initial set of outliers has been identified we use the parametric RMSE.

The backward elimination procedure is used in (5) because of situations such as "shadows". It often happens that a large outlier is not identified on the first pass but has an adjacent shadow (of opposite sign) which is identified. When the two are in the outlier regression together, the shadow drops out.

Finally, in (6), because of the uncertainty over identification of outliers, irregularities with t–values between the partial outlier threshold and the full outlier threshold (2.5 to 3.0 in our study) are treated as partial outliers (as in X–11). The amount of modification is determined linearly by the value of $\hat{I}_t$ and the position of the t–statistic between the thresholds, so $\alpha(t)$, as defined above is multiplied by $(t-2.5)/(3.0-2.5)$. Partial outliers modifications are not estimated in the regression (steps (2) and (3)) and their modifications do not affect the value of the RMSE used to detect other outliers.

Bell (1983) not only identifies and adjusts for point (AO) outliers but also changes in level and changes in the seasonal pattern. He tests for these three types of outliers simultaneously at each time point and chooses the most probable type or combination of types. We identified such outliers by looking at the irregular of the differenced series.

After tentative identification, two alternatives are open:

(1) Modify the series, using the regression estimates for the outliers, and re–fit the model.

(2) Introduce dummy variables to represent the outliers and re–fit the extended model.

The second is the one adopted by Bell (1983). He has an outer iteration loop as well, returning to outlier identification after the re–fit, and repeating until no more outliers are found. In our work, this outer loop was omitted, but both (1) and (2) were tried. Also automatic identification was only used for isolated AO's. Step changes were identified from the irregular of the differenced series and added to the model manually and IO identification was not attempted.

3. SERIES DESCRIPTIONS

60 Census Bureau monthly series were modeled: 17 Business Division retail and wholesale sales series, 15 Construction Division housing start and building permit series, 9 Foreign Trade Division import and export series, and 19 Industry Division value shipped, total inventory, and unfilled orders series (see Table 1). They are by no means a random sample, but consist mainly of series used by HBT, plus some Foreign Trade series already analyzed by members of the Statistical Research Division at the Census Bureau. The Retail Sales of Services series used by HBT have been discontinued. Most series were updated to 1982, though some ended in 1981 or 1983. The aim was to obtain 20 years' data, except for Business Division, whose series began in 1967.

4. MODEL IDENTIFICATION

Models were identified initially from the autocorrelations (ACF) and partial autocorrelations (PACF) of z_t (the logarithm of the original series), Δz_t, $\Delta_{12} z_t$, and $\Delta\Delta_{12} z_t$. For series where Trading Day (TD) effects are suspected, a 3–term periodicity in the ACF will usually be noticed, with r_4, r_7, and r_{10} being prominent. This can be tested by a regression of $\Delta\Delta_{12} z_t$ on differenced TD variables (see HBT), and model identification based on the residuals of this regression.

We estimate the models using Burman's (1980) exact likelihood estimation and signal extraction program. If the initial model is over–parameterized, the estimation program automatically reduces it, e.g. canceling a common factor between AR and MA, reducing the order of AR or MA, changing an AR factor into a difference, or replacing moving seasonality by fixed seasonal means. However, experience showed that it is prudent not to reduce the order of AR or MA (non–seasonal) until re–estimation after outlier identification. Also, even if the first estimation gave a fixed seasonal pattern, a moving pattern was tried again on re–estimation (see Section 8). With these exceptions, the same model was fitted on both estimations.

Model choice was also influenced by its potential use for seasonal adjustment. The optimal linear filters to extract the Trend, Seasonal, and Irregular are derived from the decomposition of the model spectrum (or pseudo–spectrum, when the model is non–stationary) – see Box, Hillmer, and Tiao (1978) and Burman (1980). Not all models have a valid 3–way decomposition, i.e. one with non–negative spectra; and, when two models fit equally well, one with a valid decomposition and monotonic Trend spectrum (i.e. most power at low frequencies) is preferred. These criteria tend to conflict: the second leads to a preference for MA rather than AR models, but sometimes the former have only a valid 2–way decomposition (i.e. seasonal and non–seasonal). If there is no valid decomposition at all (which can occur with models which cannot be factorized into seasonal and non–seasonal parts), the model is rejected.

The final choice of models included [23] (0 1 1)(0 1 1)$_{12}$ models, [6] others with 1 or 2 ARMA parameters, [20] with 3, and [10] with 4 or 5 parameters (see Table 2). In many cases the number of parameters was reduced by constraining insignificant ones to zero. The numbers quoted exclude the seasonal means, in cases where there is a fixed pattern.

5. DESIGN OF THE PROJECT

For each series, the last 36 observations were truncated, the chosen model fitted, and 3 post–sample forecasts made (the unadjusted or U estimates). Then outliers were identified (irregularities with t–values > 3), the series modified, the model re–fitted, and another set of forecasts made (the modified outlier or MO estimates). Then 3 observations were added to the series, and the process repeated 11 more times. At the 13th step, the full series was modeled and modified for outliers, to provide "actuals" for comparison: forecasts cannot be expected to anticipate an outlier. We call the results a chain run. The logarithms of all the series were modeled to stabilize the variance, and because the series were transformed, the comparisons of the forecasts were done in terms of the logarithms to avoid bias.

It seems to the authors that HBT's simultaneous estimation of outliers, although appealing as an 'optimal' solution, needs to be treated with caution. A model hypothesis should have a definite number of parameters, determined by an objective criterion, e.g. the AIC, whereas the number of outlier dummies depends on the threshold chosen. Moreover, if an outlier is close to the threshold (e.g. t–value < 3.5 with a threshold of t=3), it may not be identified on all the steps of the chain ("consistently identified"). The authors believe that it is better to try to assign external causes to identified outliers, and to confine the introduction of dummy variables to these cases; in fact, we were only partially successful in this quest, and some large outliers whose causes remain unknown were consistently identified. We re–estimated the chain runs with dummies for both the externally caused and consistently identified outliers: the corresponding forecasts are denoted SO (simultaneous outliers).

6. NATURE OF OUTLIERS

In attempting to link the identified outliers with external causes, two main sources were used. The first was the Chronology of Recent Noteworthy Events (1962–1984), compiled by the U.S. Bureau of Economic Analysis. It is a monthly economic memorandum that reports on major strikes and other economic and political events that effect the U.S. economy. This information was followed up by contacts with the Bureau of Economic Analysis and some market research firms in looking at particular industries. The second source was the monthly averages of temperature and precipitation recorded by the National Oceanic & Atmospheric Administration (NOAA). With one exception—the step change in the Variety Stores series discovered by HBT—only simple AO outliers were confidently identified; though there may have been a step change in the seasonality of the inventories of Oils and Fats (IFATTI). The absence of trend level changes is probably due to the Census Bureau's policy of adjusting backwards for known changes in data collection. Abrupt changes in seasonal pattern do not occur if it is caused by weather or custom. The identification of causes is discussed by Divisions (also see Table 2).

1. **Business Division** (19 outliers). These are very smooth series, only 9 out of 17 having any outliers (at t=3). All the series show Trading Day variation, but this is partly an artificial effect, because roughly 30 percent of businesses report in 4– and 5–week periods, and Business Division adjusts these figures to calendar months,

using the estimates of X–11 TD effects for those firms which do report in calendar months. For 7 outliers, external causes have been identified: unusually cold weather in the Northeast and North Central Regions, and a massive drop in the level of sales by Variety Stores in 1976, due to the closing down of W. T. Grant (noticed by HBT). A further 8 outliers are consistently identified, all with *t*–values below 4.

2. **Construction Division** (59 outliers). Most of these series are straightforward to model, because their noisiness submerges any complex correlation structure. The outliers were predominantly negative and 45 out of 59 were in the months of December–February: the prime cause was exceptionally cold weather. NOAA made available monthly averages of temperature and precipitation for various sub–regions for 1963–73, and Mr. Goodman (Federal Reserve Board) provided Census Region deviations from the 10–year averages for 1974–83, derived from NOAA data. Precipitation seems to have no effect, but both Housing Starts and Building Permits in the Northeastern and North Central Regions were affected by unusual cold (and occasionally unusual warmth). We were able to connect 37 out of 44 consistently identified outliers with extreme weather, all except one in the winter; the exception is a drop in CAOPVP in July 1977, the hottest July since 1932. However, the relation between temperature and irregularities in the series is not very close: some extremely cold months were not outliers for any of the series, and 5 out of 7 consistently identified outliers (4 negative) were in the months March–June. A possible explanation is that mean daily temperature is not the appropriate variable, but the number of days on which it is freezing at 8 a.m., and so the workforce sent home—a suggestion made by Mr. Goodman.

3. **Foreign Trade Division** (33 outliers) Some of these series were hard to model, in particular FUNKXU (Q jumped from 29 to 56 after outlier modification) and FEECXU (the latter was dropped from the study for this reason). Causes for 19 outliers were identified and large outliers in 1969 and 1971 were found in all series, except those for Canada. These were quickly identified from the Noteworthy Events sheets as due to national dock strikes. One negative outlier was found in a Canadian series in February 1967, which could be attributed to exceptionally cold weather. A negative outlier in August 1977 in exports of raw materials was probably due to a steel strike. One outlier was consistantly identified (in FIRMXU) leaving 13 outliers, which were only identified on some runs and mostly did not occur in the same month in different series, had t–values less than 4.

4. **Industry Division** (71 outliers). 34 of these could be linked with a variety of external causes: strikes in the Glass Container industry in 1966 and 1968; Communications Equipment affected by strikes at A.T.&T. in 1974 and 1983, and a jump in shipments in December 1983 just before the corporation's reorganization; the dip in Farm Machinery and Equipment shipments in October 1970, 1973, and 1976, when the 3–year labor contract at International Harvester was re–negotiated. Inventories of Oils and Fats have outliers only in August–October, due to large revisions to crop forecasts, just before harvest. Other causes of outliers include big changes in interest rates, anticipated price increases, end–season discount sales, unusual weather, and the first oil crisis in 1973. A further 18 outliers were consistently identified, but with no obvious cause, leaving 21 which were only identified on some chain runs.

7. RESULTS OF THE CHAIN RUNS

For the 45 series with identified outliers when the threshold is t=3, Table 3 shows the mean absolute forecasting errors (MAFE) over the 36 post–sample forecasts, in natural logarithms. The headings are U (unadjusted), MO (modified), and SO (simultaneous outliers). In a few cases, when only a subset of the 12 runs of an MO chain contain identified outliers, the MAFE(U) for the subset is also given, to enable comparisons to be made with the MAFE(MO). The MAFE(SO) all refer to complete chains, because the same outliers are specified on each run. The next 3 columns show the mean values of θ_{12}, the seasonal MA parameter, over the 13 runs in each chain (including the full series); and the next 3 columns give the standard deviation of θ_{12} over the 13 runs. When any run has θ_{12} above 0.96, the program automatically switches to fixed seasonal means; these cases are treated as if $\theta_{12} = 1$ in calculating the means, but the SD columns are blank.

Table 4 gives a couple of examples of the variation in θ_{12} during a chain run. It is difficult to draw general conclusions from these results. Table 5 summarizes the ratios of the MAFE's by Divisions: for 30 out of 45 series MAFE(MO) is less than or equal to MAFE(U), but in only 8 cases does it exceed MAFE(U) by more than 1 percent (including 3 of the Industry Division Inventories series, which are known to be of lower quality than Shipments). This suggests that outlier modification is usually worthwhile, and rarely harmful. MAFE(SO) is also less than or equal to MAFE(U) for 30 out of 45 series. However, only 9 of the remaining 15 are also MO failures. For 10 of the SO failures, the MAFE exceeded the MAFE(U) by more than 1 percent. On the whole there seems little to choose between the MO and SO methods of dealing with outliers: 23 series favor MO and 18 series SO, leaving 4 neutral. Simultaneous estimation takes longer, because it involves more parameters; but, if the purpose is seasonal adjustment, the fact that the same outliers are picked at each update should reduce the size of revisions.

8. INTERACTION BETWEEN FORECAST ERRORS, OUTLIERS AND MODELS.

Is there any way of predicting the effect of outliers or model characteristics on the MAFE? E.g. does the treatment of outliers become more important, as their number increases? Figure 1 shows no correlation between the ratio MAFE(SO)/MAFE(U) and the number of outliers. Does an improvement in model fit, after allowing for outliers, reduce the MAFE? Apparently not—see Figure 2, which plots the MAFE ratio against the ratio of the two Ljung–Box statistics for the 12^{th} run of the chain: the correlation (0.241) is not quite significant at the 10 percent level.

Finally, does the stability of the seasonal pattern give any information about the MAFE ratio? Outlier modification decreases θ_{12} in almost three quarters of the series, i.e. apparently the noise of the outliers had partially concealed a moving pattern. This is always the case when $\theta_{12}(U) > 0.9$, and usually when it exceeds 0.8; for several of these series all runs of the U–chain give a fixed pattern, but many of the MO and SO runs indicate slowly moving seasonality. The remaining quarter of the series, for which θ_{12} increases on modification, suggest a model in which the presence of outliers has obscured what was really a stable pattern. A priori, one would expect in the latter case MAFE(MO)/MAFE(U) < 1. But our admittedly small sample shows no inverse correlation between the MAFE ratio and the ratio of the corresponding mean values of θ_{12} (see Figure 3). Finally we note that the Standard Deviation of θ_{12} increased in 21 out of 45 series when using the SO method.

9. THRESHOLD

So far the threshold for full outliers has been taken as t=3, because that was the value used by HBT. A limited amount of work was done with the threshold at t=2.5, while making no allowances for partial outliers. Summary results are shown in Table 6. For 3 of the 60 series the number of identified outliers rose to more than 20 (the maximum that the program can handle at the moment) in more than half the chain runs. For 36 series no more outliers were identified; of the remaining 21, 15 favored the t=2.5 threshold and 6 supported the t=3 threshold. Clearly more work needs to be done in this area, but it is unlikely that there will be overwhelming evidence against t=3.

10. CONCLUSIONS

When using time series models for seasonal adjustment or short–term forecasting, the treatment of outliers is of prime importance. Two methods are explored: (1) automatic outlier identification and modification; (2) identification of external causes and simultaneous estimation of outliers, using dummy variables. Of the 60 series examined, 45 had outliers at the t=3 threshold; and two thirds of these had smaller post–sample forecast errors after allowing for the outliers. However, no characteristic (e.g. number of outliers, goodness of model fit, stability of seasonal pattern) was found which would enable us to predict whether outlier treatment will be beneficial for a particular series.

It seems that, in order to decide whether to ignore outliers in a particular case, we need to hold out some of the data and calculate the post–sample forecast errors from a sequence of runs. A priori, it is unlikely that we can afford to ignore identified outliers in the latest year, even if the test suggests that the rest of the outliers should be ignored. However, more research is needed, on a larger scale, to see whether the influence of an outlier on the forecasts (both direct and through the parameter estimates) varies with its distance from the end of the series.

Which of the two methods of outlier treatment is preferable is a matter of much less importance: it may well be decided by the amount of statistical and economic expertise available and the computer resources. Likewise the question of the choice of t–value for the threshold is not crucial; but, until more research has been done, t=3 is recommended.

Another area for investigation is whether heteroscedasticity of the innovations process between months (e.g. due to weather in Construction, harvest variations in Inventories of Oils and Fats) should be incorporated in the Likelihood function.

Finally, what are the practical lessons for the seasonal adjusters? There is evidence that the model–based method gives a "better" adjustment for some series (Bell and Hillmer (1984)), though X–11 may be satisfactory for many, perhaps the majority. It is recommended that the model–based algorithm should be used initially as an adjunct to X–11: first, to estimate θ_{12} and so decide whether the 3×5 seasonal moving average should be replaced by a less flexible smoother; second, to identify outliers and modify the series, before X–11 is run.

APPENDIX

A.1. MODELLING ADDITIVE OUTLIERS

An additive outlier (AO) only affects the series at t_0:

$$z_t = \frac{\theta(B)}{\phi(B)}a_t + \alpha\xi(t,t_0) \qquad (A.1)$$

where $\xi(t,t_0) = \{1 \text{ if } t = t_0, \text{ and } 0 \text{ otherwise}\}$ and $\phi(B)$ may contain difference operators. Following HBT, but using slightly different notation, let u_t be the innovations estimated from the model without taking account of outliers.

For an invertible model, multiply (A.1) by $\phi(B)/\theta(B) = \pi(B)$. Then

$$u_t = \alpha \cdot \pi(B)\xi(t,t_o) + a_t. \tag{A.2}$$

Using O.L.S., the impact $\hat{\alpha}$ of the intervention and its variance are:

$$\hat{\alpha} = (u_{t_o} + \pi_1 u_{t_o+1} + \cdots + \pi_{n-t_o} u_n)/w_{n-t_o},$$

$$\mathrm{var}(\hat{\alpha}) = \sigma_a^2/w_{n-t_o} \tag{A.3}$$

where $w_t = 1 + \pi_1^2 \cdots + \pi_t^2$.

These formulae reflect the influence of an AO on all innovations from t_o onwards.

A.2. DETECTION OF AN AO

If we put $u_t = 0(t>n)$, since they are unknown, (A.3) can be written:

$$\hat{\alpha} = \pi(F)u_{t_o}/w_{n-t_0}. \tag{A.4}$$

If $\{a_t\}$ is I.I.D. Normal, a t-test takes the form:

$$\frac{|\pi(F)u_{t_o}|}{w_{n-t_o}} \Bigg/ \frac{\sigma_a}{w_{n-t_o}^{1/2}} = \frac{|\pi(F)u_{t_o}|}{w_{n-t_o}^{1/2}\sigma_a} > \lambda$$

where λ depends on the significance level. If t_0 is unknown, and we define $I_t = \pi(F)u_t$, a pass over the data is made to find those points where $|I_t| > w_{n-t}^{1/2}\lambda\sigma_a$. Note that w is a function of t, so the test has a threshold that tapers downwards towards t=n.

The following developments arose from the first author's discussion on the HBT paper and subsequent collaboration with Dr. Bell. If t is not too near the start, $u_t = \pi(B)z_t$ so

$$I_t = \pi(B)\pi(F)z_t \ . \qquad \text{(A.5)}$$

This is a symmetric doubly–infinite moving average, in which missing values are replaced by forecasts and backcasts. $\{I_t\}$ behaves like the Irregular component in the traditional decomposition z_t = Trend + Seasonal + Irregular.

In practice, when (n–t) exceeds about 2 years (depending on the model and parameters), w_{n-t} closely approaches its final value w_∞. By symmetry, the taper near the end of the series is matched by a taper near the beginning: we cannot put $u_t = \pi(B)z_t$, but instead we use the model for the reversed series (which has the same ARMA parameters). If $v_t = \pi(F)z_t$, then $I_t = \pi(B)v_t$, which will be valid, except near the end of the series. Thus the test becomes: compute $\hat{I}_t = \pi(B)\pi(F)z_t$, the first estimate of I_t. We define $\tilde{w}_t = \{w_t \text{ if } t<t_L,\ w_{n-t} \text{ if } t>n-t_L,\ w_\infty \text{ otherwise}\}$ and t_L is the least t for which $w_t > 0.95w_\infty$, (say), and a potential outlier is identified when:

$$|\hat{I}_t| > \tilde{w}_t^{\,1/2}\lambda\sigma_a \ . \qquad \text{(A.6)}$$

A.3. MULTIPLE AO's

If a pass over the $\{\hat{I}_t\}$ indicates more than 1 potential outlier, we must take account of the interactions between them. If t_1 and t_2 are close together with $(t_1 < t_2)$, the residuals will not have recovered from the effect at t_1 before they are hit by the next outlier at t_2. We therefore need to apply a multiple regression. If outliers are detected at $t_1, t_2, \cdots t_k$, let $\mathbf{p}_j' = (0, \cdots 0, 1, \pi_1, \cdots \pi_{n-t_j})$ with 1 in position t_j. For this outlier alone, equation (A.2) becomes: $\mathbf{u} = \alpha_j \mathbf{p}_j + \mathbf{a}$ where $\mathbf{u}' = (u_1\ u_2 \cdots u_n)$, $\mathbf{a}' = (a_1\ a_2 \cdots a_n)$, and α_j = measure of outlier at t_j.

The effects of the outliers are additive so the general case is:

$$\mathbf{u} = \mathbf{P}\boldsymbol{\alpha} + \mathbf{a} \tag{A.7}$$

where $\mathbf{P} = (\mathbf{p}_1\ \mathbf{p}_2 \cdots \mathbf{p}_k)$ is an $n \times k$ matrix and $\boldsymbol{\alpha}' = (\alpha_1\ \alpha_2 \cdots \alpha_k)$. The normal equations are $\mathbf{P}'\mathbf{P}\boldsymbol{\alpha} = \mathbf{P}'\mathbf{u}$. But $\mathbf{p}_j'\mathbf{u} = u_{t_j} + \pi_1 u_{t_j+1} \cdots + \pi_{n-t_j} u_n = I_{t_j}$ and

$$\mathbf{P}'\mathbf{P} = \begin{bmatrix} w(t_1,t_1), & w(t_1,t_2), & w(t_1,t_3) & \cdots \\ w(t_2,t_1), & w(t_2,t_2), & w(t_2,t_3) & \cdots \end{bmatrix} = \mathbf{W} \text{ (say)}$$

where $w(t_j,t_j) = \sum_{i=0}^{n-t_j} \pi_i^2$, $w(t_h,t_j) = \sum_{i=0}^{n-t_j} \pi_i \pi_{i+t_j-t_h}$ if $t_j > t_h$, and $w(t_h,t_j) = w(t_j,t_h)$. Let $\mathbf{J}_0' = (I_{t_1}, I_{t_2}, \cdots, I_{t_k})$, so the normal equations are:

$$\mathbf{W}\hat{\boldsymbol{\alpha}} = \mathbf{J}_0. \tag{A.8}$$

The taper affects $\mathbf{W}$: for terms not near the beginning or end, we can replace $w(t_h,t_j)$ by $\tilde{w}(t_j-t_h) = \sum_{i=0}^{\infty} \pi_i \pi_{i+t_j-t_h}$ (assuming $t_j > t_h$), and for terms near the beginning the upper bound of summation in $\tilde{w}(t_h,t_j)$ is t_h. The covariance matrix of $\hat{\boldsymbol{\alpha}}$ is $(\mathbf{P}'\mathbf{P})^{-1}\sigma_a^2 = \mathbf{W}^{-1}\sigma_a^2$, whence the t–values of the $\hat{\alpha}_i$ can be calculated. The initial sum of squares of the dependent variable is $\sum_{t=1}^{n} u_t^2$, and the initial selection of outliers for inclusion in the regression is made with a robust version of σ_u, $1.48 \cdot \text{median}|u_t|$, instead of σ_a in (A.6). Revised residuals from (A.7) are needed to identify further outliers but, more conveniently, we can use revised $\hat{I}_t$, say $I_t(\hat{\alpha})$, e.g. from (A.7), $\hat{a}_t = u_t - (\pi_{t-1}, \pi_{t-2}, \cdots, \pi_{t-n})\hat{\boldsymbol{\alpha}}^*$ [$t = 1, 2, \cdots, n$]. The row vector is row t of an ($n \times n$) matrix $\mathbf{P}^*$, $\pi_j = 0$ if $j<0$, and $\hat{\boldsymbol{\alpha}}^*$ is an n–vector, which is $\hat{\boldsymbol{\alpha}}$ padded with zeros. So $I_t(\hat{\alpha}) = \pi(F)u_t - \pi(F)(\pi_{t-1}, \pi_{t-2}, \cdots, \pi_{t-n})\hat{\boldsymbol{\alpha}}^*$. Now $\pi(F)u_t = I_t$ and $\pi(F)\pi_{t-i} = \pi_0\pi_{t-i} + \pi_1\pi_{t-i+1} \cdots$. So

$$I_t(\hat{\alpha}) = I_t - \mathbf{W}^* \hat{\boldsymbol{\alpha}}^* \tag{A.9}$$

where $\mathbf{W}^*$ is ($n \times n$) and $w^*_{st} = \sum_{i=0}^{n-t} \pi_i \pi_{i+t-s}$ ($t>s$). Only columns ($t_1\ t_2 \cdots t_k$) of $\mathbf{W}^*$ are needed, because the rest are multiplied by zero elements of $\hat{\boldsymbol{\alpha}}^*$. Also, it is found that $\bar{w}_{st} \approx 0$ if $t-s > 5$, except for $t-s$ = MQ or 2MQ. (where MQ = periodicity of model, i.e. usually 12 or 4).

In other words, interaction between an outlier and other terms of the series I_t is negligible, unless they are close together or differ by exactly one or two years (for seasonal models). The advantages of this algorithm for outlier identification over that in older methods (e.g. the X–11 seasonal adjustment program) are the use of a formal test and the proper allowance for outlier interaction. The process of identification is iterative.

REFERENCES

Barnett, V. and Lewis, T. (1978), Outliers in Statistical Data, John Wiley & Sons: New York.

Bell, W. R. (1983), "A Computer Program (TEST) for Detecting Outliers in Time Series," Proceedings of the American Statistical Association Meetings, Chicago.

Bell, W. R. and Hillmer, S. C. (1984), "Issues involved with the Seasonal Adjustment of Economic Time Series," *Journal of Business & Economic Statistics*, 2, 291-349.

Box, G.E.P., Hillmer, S. C., and Tiao, G. C. (1978), "Analysis and Modeling of Seasonal Time Series," in *Seasonal Analysis of Economic Time Series*, ed. A. Zellner, Washington, D. C.: U.S. Department of Commerce, Bureau of the Census, 309-334.

Box, G.E.P. and Tiao, G. C. (1975). Intervention analysis with applications to economic and environmental problems, *Journal of the American Statistical Association*, 70, 70-79.

Burman, J. P. (1980), "Seasonal Adjustment by Signal Extraction," *Journal of the Royal Statistical Society*, Ser. A, 143, 321-337.

_______(1983). Comments on Modeling considerations in the seasonal adjustment of economic time series in *Applied Time*

Series Analysis of Economic Data by Hillmer, Bell & Tiao (see below).

Chronology of Recent Noteworthy Events (1962 to 1984), Monthly memorandum for the Statistical Indicators Division of the Bureau of Economic Analysis, Department of Commerce, Washington, DC 20230. George R. Green, Acting Chief.

Denby, L. and Martin, R. D. (1979), "Robust Estimation of the First Order Autoregressive Parameter," *Journal of the American Statistical Association*, 74, 140-146.

Draper, N., and Smith, H. (1984), *Applied Regression Analysis*, Second Edition, John Wiley & Sons: New York.

Hillmer, S. C., Bell, W. R., and Tiao, G. C. (1983a), "Modeling Considerations in the Seasonal Adjustment of Economic Time Series," in *Applied Time Series Analysis of Economic Data*, ed. A. Zellner, Washington, D.C.: U.S. Department of Commerce, Bureau of the Census, 74-100.

_______(1983b) "Response to Discussants," in *Applied Time Series Analysis of Economic Data*, ed. A. Zellner, Washington, D.C.: U.S. Department of Commerce, Bureau of the Census, 123-124.

Shisken, J., Young, A.H., and Musgrave, J.C. (1967), "The X-11 Variant of the Census Method II Seasonal Adjustment Program," Technical Paper No. 15, U.S. Dept. of Commerce, Bureau of Economic Analysis.

List of Tables

List of Figures

Figure 3. Forecast Error Ratios vs. Ratios of Mean Seasonal Moving-average parameters. The MAFE ratios are plotted against the mean of the θ_{12} estimates from the chain run. The level of θ_{12} indicates the rate of change in the seasonal pattern, values closer to one being more stable. The vertical reference line indicates no change in the stability of the seasonal pattern due to outlier modification; points to the right of this indicate that outlier modification produces a more stable seasonal pattern. Note that there is no relation between the MAFE ratios and the change in the stability of the seasonal pattern due to outlier modifications.

TABLE 1. Outlier Study Series

Series	Division Code	Years	Description
Business Division			
BAPPRS	S03000 570002	67-82	Retail sales of household appliances
BAUTRS	S03000 550000	67-82	Total retail sales of automobiles
BDPTRS	S03000 531100	67-82	Retail sales in department stores
BELGWS	S03000 506000	67-82	Retail sales of electrical goods
BFRNRS	S03000 570001	67-82	Retail sales of furniture
BFRNWS	S03000 502000	67-82	Wholesale sales of furniture
BGASRS	S03000 554100	67-82	Retail sales at gasoline stations
BGRCRS	S03000 541100	67-82	Retail sales at grocery stores
BGRCWS	S03000 514000	67-82	Wholesale sales at grocery stores
BHDWRS	S03000 507000	67-82	Retail sales at hardware stores
BHDWWS	S03000 525100	67-82	Wholesale sales at hardware stores
BLQRRS	S03000 592100	67-82	Retail sales at liquor stores
BMNCRS	S03000 561100	67-82	Retail sales of men's clothes
BSHORS	S03000 566100	67-82	Retail sales of shoes
BSPGWS*	S03000 504000	67-82	Retail sales of sporting, recreational an photographic goods
BVARRS	S03000 533100	67-82	Retail sales at variety stores
BWAPRS	S03000 560001	67-82	Retail sales of women's apparel
Construction Division			
C1FTBP	BP1FAM	64-83	Total 1 family dwelling building permits
C24TBP	BP24FA	64-83	Total 2 to 4 family dwelling building permits
C5PTBP	BP5PFA	64-83	Total 5+ family dwelling building permits
CAOPVP	PRAOTH	64-83	Value in place, all other private residences
CNC1BP	XABPNC1F	64-83	North Central 1 family building permits
CNC1HS	XAHSNC1F	64-83	North Central 1 family housing starts
CNC5HS	XAHSNC5F	64-83	North Central 5+ family housing starts
CNCTBP	BPNCRE	64-83	Total North Central building permits
CNCTHS	HSNC	64-83	Total North Central housing starts
CNE1BP	XABPNE1F	64-83	Northeast 1 family building permits
CNE1HS	HSNE1F	64-83	Northeast 1 family housing starts
CNETBP	BPNERE	64-83	Total Northeast building permits
CNETHS	HSNE	64-83	Total Northeast housing starts
CSOTHS	HSSO	64-83	Total South housing starts
CWSTHS	HSWT	64-83	Total West housing starts

TABLE 1 (continued)
Outlier Study Series

Series	Division Code	Years	Description
Foreign Trade Division			
FCANXU	XUCAN	66-82	Unadjusted exports to Canada
FCNCXU	XUCARSC	66-82	Unadjusted exports of cars to Canada
FEECXU	XUEEC	66-82	Unadjusted exports to the European Ecomomic Community
FFPPXU	XU058	66-82	Unadjusted exports of fruits, preserves and produce
FIRMXU	XU2	66-82	Unadjusted exports of industrial raw materials
FLARXU	XULAR	66-82	Unadjusted exports to Latin American Republics
FUNKXU	XUUK	66-82	Unadjusted exports to the United Kingdom
FWGRXU	XUWGER	66-82	Unadjusted exports to West Germany
FWHMXU	XUWH	66-82	Unadjusted exports to the Western Hemisphere
Industry Division			
IBEVTI	S62TI	64-83	Total inventories of beverages
ICMETI	N37TI	68-84	Total inventories of communications equipment
IFATTI	S63TI	64-83	Total inventories of fats and oils
IFMETI	S23TI	62-81	Total inventories of farm machinery and equipment
IGLCTI	S07TI	62-81	Total inventories of glass containers
IHAPTI	S35TI	62-81	Total inventories of household appliances
ITVRTI	S36TI	64-83	Total television and radio inventories
INEWUO	S80UO	64-83	Unfilled newspaper, periodical, and magazine orders
ITVRUO	S36UO	64-83	Unfilled television and radio orders

IAPEVS	N44VS	68-83	Value shipped of aircraft parts and equipment
IBEVVS	S62VS	64-83	Value shipped of beverages
ICMEVS	N37VS	68-83	Value shipped, communications equipment
IFATVS	S63VS	62-81	Value shipped of fats and oils
IFMEVS	S23VS	64-83	Value shipped of farm machinery and equipment
IFRTVS	S86VS	62-81	Value fertilizer shipped
IGLCVS	S07VS	62-81	Value of glass containers shipped
IHAPVS	S35VS	64-83	Value of household appliances shipped
IRREVS	S46VS	64-83	Value of railroad equipment shipped
ITOBVS	S65VS	64-83	Value of tobacco shipped
ITVRVS	S36VS	62-81	Value of televisions and radios shipped

TABLE 2. ARIMA Models and Number of Outliers

Series	Model	No. of Outliers Total[a]	Ext. Causes	Consistently Identified[b]
Business Division				
BVARRS	$(013)(011)_{12}$+TD+E, $\theta_2 = 0$	2	1	1
BGRCRS	$(013)(011)_{12}$+TD+E, $\theta_2 = 0$	4	2	2
BHDWRS	$(014)(011)_{12}$+TD, $\theta_2 = 0$	2	1	1
BAPPRS	$(010)(011)_{12}$+TD	1	-	1
BWAPRS	$(012)(011)_{12}$+TD+E	1	-	1
BSHORS	$(011)(011)_{12}$+TD+E	2	-	2
BAUTRS	$(110)(011)_{12}$+TD	1	1	-
BFRNWS	$(011)(011)_{12}$+TD	4	1	-
BSPGWS	$(011)(011)_{12}$+TD	2	1	-
BGASRS	$(011)(011)_{12}$+TD	-	-	-
BLQRRS	$(012)(011)_{12}$+TD	-	-	-
BMNCRS	$(101)(011)_{12}$+TD	-	-	-
BFRNRS	$(101)(011)_{12}$+TD	-	-	-
BDPTRS	$(101)(011)_{12}$+TD+E	-	-	-
BELGWS	$(011)(011)_{12}$+TD	-	-	-
BHDWWS	$(011)(011)_{12}$+TD	-	-	-
BGRCWS	$(013)(011)_{12}$+TD	-	-	-
		19	7	8

[a]Identified on at least one run.
[b]Identified on all runs, but no cause known.

Key: $(000)_{12}$ = seasonal means. TD = Trading Day adjustment. E = Easter adjustment. "$\theta_L' = 0$" ARMA parameters that are less than their standard errors and are constrained to zero.

TABLE 2 (continued)

Series	Model	Total[a]	No. of Outliers Ext. Causes	Consistently Identified[b]
Construction Division				
CAOPVP	$(310)(011)_{12}$	3	1	-
CNC1HS	$(200)(011)_{12}$	9	6	-
CNC5HS	$(011)(011)_{12}$	3	1	2
CNCTHS	$(101)(011)_{12}$	6	5	1
CNE1HS	$(300)(011)_{12}$	2	2	-
CNETHS	$(101)(011)_{12}$	4	2	-
CSOTHS	$(100)(000)_{12}$	1	1	-
CNC1BP	$(011)(011)_{12}$+TD	7	5	1
CNCTBP	$(100)(011)_{12}$+TD	3	3	-
CNE1BP	$(210)(000)_{12}$	6	5	-
CNETBP	$(011)(011)_{12}$+TD	4	4	-
C1FTBP	$(011)(011)_{12}$+TD	5	1	-
C24TBP	$(011)(011)_{12}$+TD	3	-	3
C5PTBP	$(013)(011)_{12}$, $\theta_2=0$	3	1	-
CWSTHS	$(013)(011)_{12}$	-	-	-
		39	37	7
Foreign Trade Division				
FIRMXU	$(011)(011)_{12}$	1	-	1
FCANXU	$(011)(011)_{12}$	-	-	-
FCNCXU	$(013)(011)_{12}$	3	1	-
FWHMXU	$(011)(000)_{12}$	6	4	-
FLARXU	$(013)(000)_{12}$	7	5	-
FWGRXU	$(011)(011)_{12}$	7	3	-
FUNKXU	$(011)(011)_{12}$	5	3	-
FFPPXU	$(011)(011)_{12}$	4	3	-
		33	19	1

TABLE 2 (continued)

Series	Model	No. of Outliers Total[a]	Ext. Causes	Consistently Identified[b]
Industry Division				
ICMEVS	$(210)(011)_{12}$	5	3	2
ICMETI	$(310)(011)_{12}$	3	-	3
IGLCVS	$(012)(011)_{12}$	9	7	-
IGLCTI	$(013)(000)_{12}$	4	3	-
IFMEVS	$(019)(000)_{12}$, $\theta_4=\cdots=\theta_8=0$	5	4	-
IFMETI	$(210)(011)_{12}$	10	7	1
IHAPTI	$(012)(011)_{12}$	3	1	1
ITVRVS	$(012)(011)_{12}$	1	-	1
ITVRTI	$(011)(011)_{12}$	2	1	-
IFATTI	$(014)(011)_{12}$, $\theta_2=\theta_3=0$	6	2	-
IFATVS	$(011)(011)_{12}$	2	-	1
ITOBVS	$(013)(011)_{12}$	11	2	5
INEWVO	$(011)(011)_{12}$	2	-	2
IAPEVS	$(011)(011)_{12}$	3	-	2
IFRTVS	$(011)(011)_{12}$	5	4	-
IHAPVS	$(011)(011)_{12}$	-	-	-
IRREVS	$(011)(011)_{12}$	-	-	-
IBEVVS	$(014)(011)_{12}$+TD	-	-	-
IBEVTI	$(012)(011)_{12}$	-	-	-
		71	34	1

[a]Identified on at least one run.
[b]Identified on all runs, but no cause known.

Key: $(000)_{12}$ = seasonal means. TD = Trading Day adjustment. E = Easter adjustment. "$\theta_L' = 0$" ARMA parameters that are less than their standard errors and are constrained to zero.

TABLE 3

MAFE (Mean Absolute Forecast Errors), Mean & S.D. of θ_{12} over chain runs, and Ljung-Box Q Statistic for 12th run of each chain)

Series	MAFE U	MAFE MO	MAFE SO	Mean (θ_{12}) U	Mean (θ_{12}) MO	Mean (θ_{12}) SO	SD(θ_{12}) U	SD(θ_{12}) MO	SD(θ_{12}) SO	Ljung-Box Q (24 terms) U	Ljung-Box Q (24 terms) MO	Ljung-Box Q (24 terms) SO
BVARRS	.0253[a]	.0257[a]	.0257	.743	.725	.765	.013	.019	.012	42.8	32.1	34.7
	.0257[b]	.0265[b]										
BGRCRS	.0099	.0097	.0095	.852	.878	.891	.049	.054	.048	34.0	30.0	31.4
BHDWRS	.0350	.0340	.0326	.432	.362	.358	.015	.023	.017	16.5	15.0	14.7
BAPPRS	.0411	.0408	.0409	.738	.687	.705	.013	.015	.019	30.0	27.6	25.6
BWAPRS	.0326	.0326	.0326	.641	.648	.649	.008	.007	.007	35.2	40.4	40.7
BSHORS	.0379	.0377	.0378	.720	.699	.687	.012	.015	.017	26.0	25.1	27.3
BAUTRS	.0704	.0669	.0670	.812	.782	.797	.026	.026	.025	35.5	47.4	44.7
BFRNWS	.0484	.0463	.0460	.766	.756	.754	.047	.063	.042	24.1	24.8	23.3
(8 runs)	.0494	.0473[c]										
BSPGWS	.0655	.0651	.0671	.680	.645	.634	.025	.033	.026	28.2	30.3	33.0
CAOPVP	.0675	.0679	.0660	.897	.894	.928	.009	.009	.014	23.3	29.2	27.4
(6 runs)	.0867	.0872[c]										
CNC1HS	.2110	.1883	.1661	.730	.800	.791	.055	.053	.034	30.3	16.2	23.7
CNC5HS	.3446	.3335	.2975	.772	.732	.733	.019	.022	.017	16.8	16.7	15.0
CNCTHS	.2270	.2238	.2282	.713	.701	.756	.027	.029	.025	21.6	35.0	35.0
CNE1HS	.1575	.1519	.1528	.660	.682	.691	.035	.056	.038	46.7	33.7	34.2

(a) With prior adjustment for step change. (b) Without prior adjustment for step change. (c) Scaled by ratio for U run. (d) All fixed seasonality. (e) Some fixed. (f) Most fixed.

TABLE 3 (continued)
MAFE (Mean Absolute Forecast Errors),
Mean & S.D. of θ_{12} over chain runs, and
Ljung-Box Q Statistic for 12th run of each chain)

Series	MAFE U	MAFE MO	MAFE SO	Mean (θ_{12}) U′	Mean (θ_{12}) MO	Mean (θ_{12}) SO	SD(θ_{12}) U	SD(θ_{12}) MO	SD(θ_{12}) SO	Ljung-Box Q (24 terms) U	Ljung-Box Q (24 terms) MO	Ljung-Box Q (24 terms) SO
CNETHS	.1771	.1767	.1725	.790	.839	.869	.042	.027	.021	20.5	22.4	23.1
CSOTHS	.1173	.1173	.1178	1.000	1.000	1.000	--(d)	--(d)	--(d)	24.5	26.4	26.8
(10 runs)	.1107	.1107(c)										
CNC1BP	.1493	.1456	.1548	.744	.798	.823	.029	.045	.039	20.5	26.4	25.8
CNCTBP	.1562	.1589	.1599	.725	.753	.780	.012	.012	.015	19.3	26.6	29.0
CNE1BP	.1351	.1299	.1302	1.000	.963	.958	--(d)	--(e)	--(e)	22.8	22.4	21.1
CNETBP	.1122	.1053	.1028	.846	.808	.801	.012	.014	.013	15.8	20.6	20.1
C1FTBP	.1204	.0909	.0905	.986	.945	.924	--(f)	--(e)	--(e)	14.4	15.9	9.3
C24TBP	.0898	.0903	.0902	.912	.895	.874	.028	.038	.029	18.9	19.8	18.6
C5PTBP	.1077	.1207	.1125	.791	.801	.792	.014	.039	.018	19.9	26.7	24.8
FIRMXU	.0874	.0874	.0873	.927	.920	.918	.027	.018	.017	35.8	30.0	32.2
FCNCXU	.2356	.2245	.2303	.683	.687	.688	.012	.017	.013	16.3	23.6	17.7
FWHMXU	.0714	.0689	.0708	.992	.935	.938	--(f)	--(f)	--(f)	31.4	27.1	22.4
FLARXU	.0966	.0972	.0988	1.000	.950	.940	--(d)	--(f)	--(f)	32.5	35.9	33.5
FWGRXU	.0622	.0607	.0640	.908	.887	.942	.011	.020	.026	27.9	21.2	39.9
FUNKXU	.0808	.0814	.0814	.880	.855	.841	.010	.009	.013	29.5	56.3	34.4
FFPPXU	.0858	.0855	.0851	.766	.728	.677	.011	.009	.015	41.4	31.7	22.8

(a) With prior adjustment for step change. (b) Without prior adjustment for step change. (c) Scaled by ratio for U run. (d) All fixed seasonality. (e) Some fixed. (f) Most fixed.

TABLE 3 (continued)

MAFE (Mean Absolute Forecast Errors),
Mean & S.D. of θ_{12} over chain runs, and
Ljung-Box Q Statistic for 12th run of each chain)

Series	MAFE U	MAFE MO	MAFE SO	Mean (θ_{12}) U	Mean (θ_{12}) MO	Mean (θ_{12}) SO	SD(θ_{12}) U	SD(θ_{12}) MO	SD(θ_{12}) SO	Ljung-Box Q (24 terms) U	Ljung-Box Q (24 terms) MO	Ljung-Box Q (24 terms) SO
ICMEVS	.0346	.0325	.0325	.795	.687	.719	.019	.016	.018	17.3	27.6	23.5
ICMETI	.0078	.0079	.0077	.694	.683	.701	.025	.034	.024	22.3	24.8	25.8
IGLCVS	.0450	.0456	.0445	.877	.846	.836	.009	.012	.010	9.9	26.1	28.1
IGLCTI	.0221	.0229	.0232	1.000	.950	.948	--(d)	--(e)	--(e)	23.0	28.9	26.5
IFMEVS	.0934	.0911	.0913	1.000	1.000	1.000	--(d)	--(d)	--(d)	37.2	41.5	40.8
IFMETI	.0202	.0205	.0225	.806	.753	.756	.015	.043	.011	26.7	13.2	13.1
IHAPTI	.0221	.0240	.0241	.847	.848	.840	.023	.030	.023	20.5	30.8	25.5
ITVRVS	.1061	.1069	.1061	.647	.637	.642	.023	.027	.022	23.6	24.7	23.7
ITVRTI	.0374	.0370	.0371	.822	.821	.816	.006	.007	.007	22.3	29.6	31.8
IFATTI	.0855	.0849	.0854	.647	.648	.693	.046	.046	.020	33.9	31.1	28.6
IFATVS	.0718	.0719	.0710	.737	.787	.783	.019	.026	.027	38.1	35.2	34.6
ITOBVS	.0674	.0656	.0571	.749	.712	.717	.013	.016	.014	45.5	16.7	15.9
INEWUO	.0571	.0566	.0495	.593	.573	.580	.024	.018	.015	22.1	22.8	20.6
IAPEVS	.0740	.0832	.0778	.973	.849	.851	--(f)	.046	.032	25.6	22.6	24.4
IFRTVS	.0864	.0908	.0847	.411	.355	.348	.036	.082	.044	14.7	22.8	19.7
(11 runs)	.0823	.0865(c)										

(a) With prior adjustment for step change. (b) Without prior adjustment for step change. (c) Scaled by ratio for U run. (d) All fixed seasonality. (e) Some fixed. (f) Most fixed.

TABLE 4

Estimated θ_{12} on Unadjusted (U), Modified Series (MO) and Simultaneously Estimated Outliers (SO) Runs

Run No.	CNC1HS U	CNC1HS MO	CNC1HS SO	ICMEVS U	ICMEVS MO	ICMEVS SO
1	.704	.820	.796	.794	.667	.696
2	.713	.827	.780	.774	.668	.697
3	.720	.842	.794	.771	.664	.694
4	.746	.853	.808	.784	.680	.712
5	.731	.827	.801	.775	.672	.703
6	.667	.711	.748	.786	.684	.717
7	.656	.710	.737	.797	.697	.731
8	.666	.737	.747	.818	.706	.742
9	.686	.750	.762	.807	.693	.735
10	.786	.842	.827	.798	.691	.730
11	.810	.835	.835	.798	.697	.733
12	.809	.834	.836	.794	.701	.733
13	.791	.815	----[a]	.839	.715	----[a]

Note: The estimated model of the first run for each series uses data with the last three years removed, the 12th run only has three months removed, and the last, 13th run, is on the full data. The three columns for each series are U = unadjusted series, MO = modified series, and SO = outliers simultaneously estimated in model.

[a]N.A. because the 13th SO run was made purely to provide "actuals" for comparison with the post-sample forecasts, and the former were modified for <u>all</u> identified outliers.

TABLE 5

	B	C	F	I	Total
MAFE(MO)/MAFE(U):					
> 1.01	0	4	2	9	15
≅	2	1	1	-	4
< 1.00	7	9	4	6	26
	9	14	7	15	45
MAFE(SO)/MAFE(U):					
> 1.01	1	6	3	4	14
≅	2	-	-	1	3
< 1.00	6	8	4	10	28
	9	14	7	15	45
MAFE(SO)/MAFE(MO):					
> 1.01	4	6	4	4	18
≅	2	-	1	1	4
< 1.00	3	8	2	10	23
	9	14	7	15	45

NOTE: Columns are labeled by division: B = Business Division, C = Construction, F = Foreign Trade, I = Industry. Also, '> 1.01' indicates the numerator is more than one percent larger than the denominator, '≅' indicates the numerator is less than one percent larger, and < 1 indicates the numerator is less than the denominator.

TABLE 6

Effect of Changing Threshold from t = 3 to t = 2.5

		B	C	F	I	Total
No extra outliers		13	10	5	8	36
Extra outliers:	MAFE(2.5)<MAFE(3.0)	2	3	4	6	15
	MAFE(2.5)≥MAFE(3.0)	2	1	-	3	6
Too many outliers at t=2.5 on at least half the runs		-	1	-	2	3
		17	15	9	19	60

Figure 1.

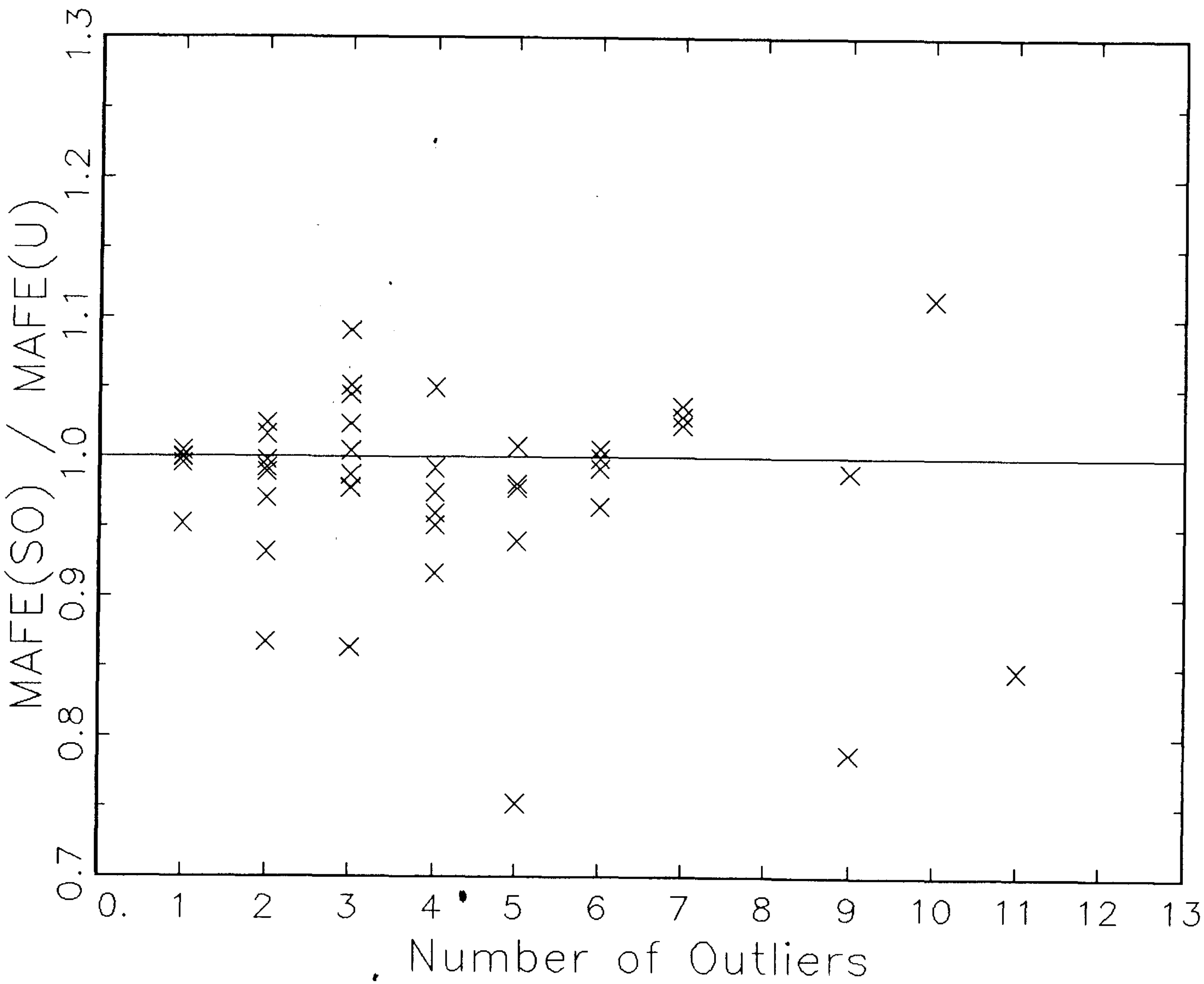
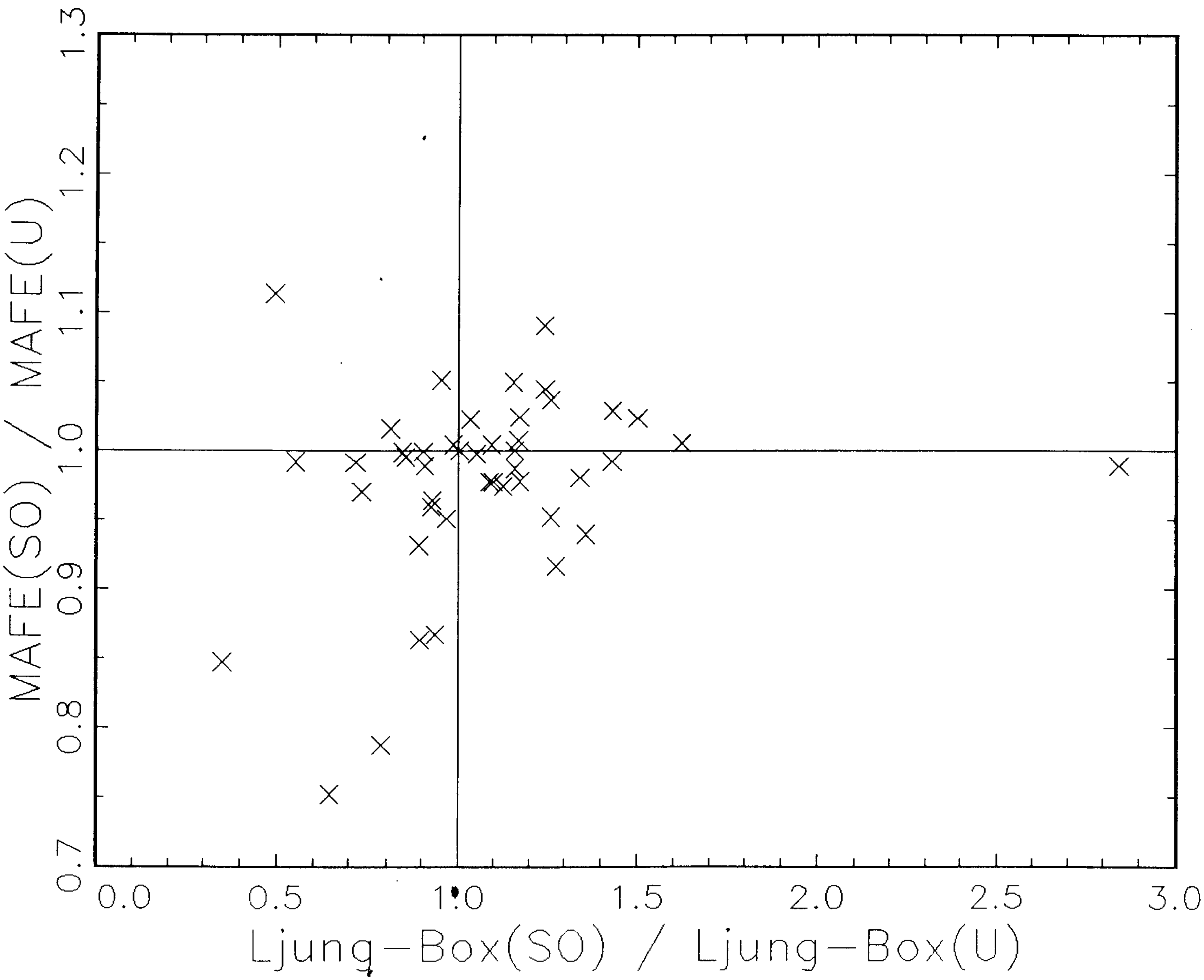
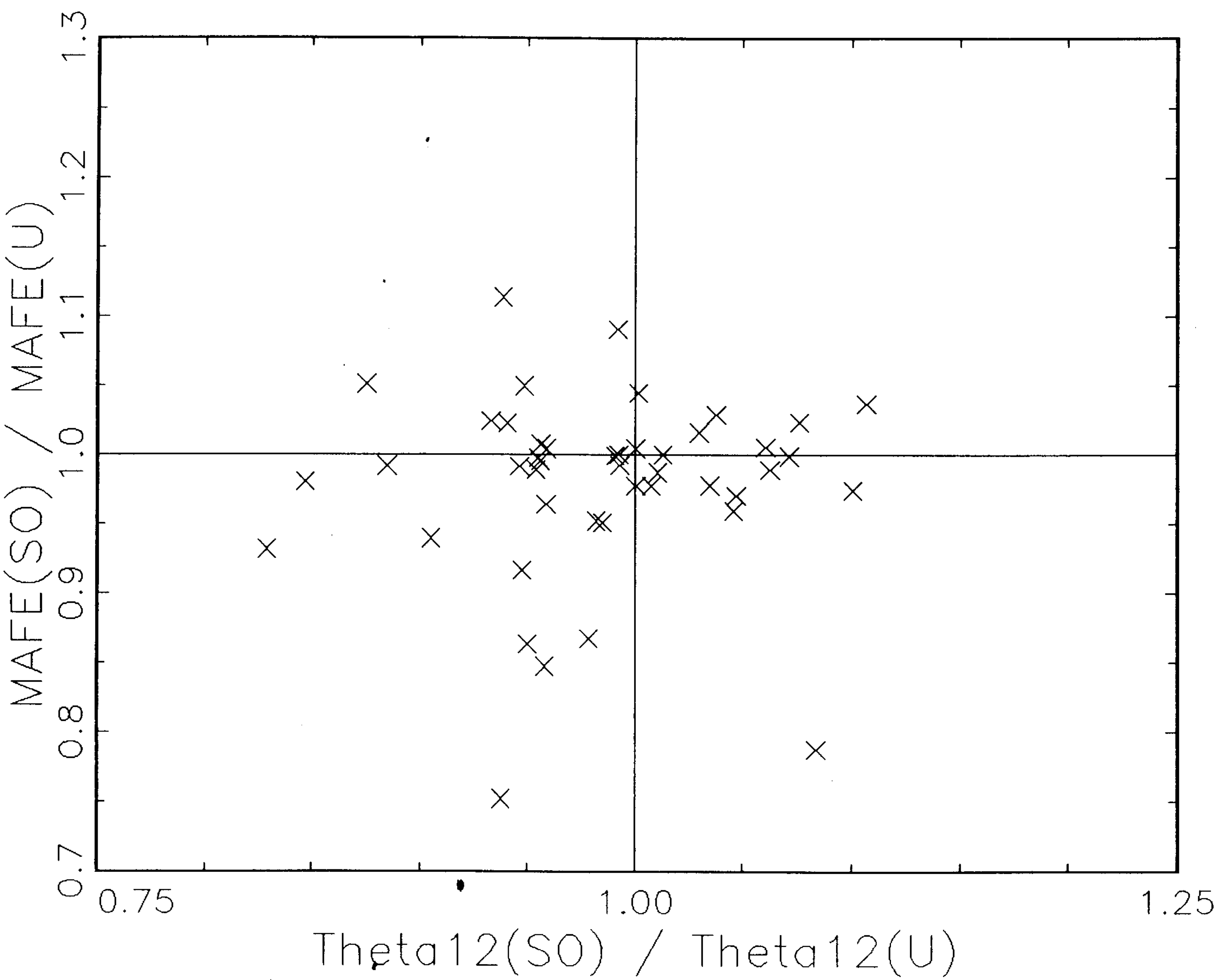

MAFE(SO) / MAFE(U)
1.3
1.2
1.1
1.0
0.9
0.8
0.7
0. 1 2 3 4 5 6 7 8 9 10 11 12 13
Number of Outliers

Figure 2.



MAFE(SO) / MAFE(U)
0.7
0.8
0.9
1.0
1.1
1.2
1.3
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Ljung-Box(SO) / Ljung-Box(U)

Figure 3.



1.3
1.2
1.1
1.0
0.9
0.8
0.7
MAFE(SO) / MAFE(U)
0.75
1.00
1.25
Theta12(SO) / Theta12(U)