BUREAU OF THE CENSUS

STATISTICAL RESEARCH DIVISION REPORT SERIES

SRD Research Report Number: CENSUS/SRD/RR-84/30

NEW TECHNIQUES FOR DETERMINING IF A TIME SERIES CAN BE SEASONALLY ADJUSTED RELIABLY, AND THEIR APPLICATION TO U.S. FOREIGN TRADE SERIES

by

David F. Findley and Brian C. Monsell
Statistical Research Division
Bureau of the Census
Room 3524, F.O.B. #3
Washington, D.C. 20233

Recommended by: Nash J. Monsour

Report completed: December 12, 1984

Report issued: December 14, 1984

To appear in Issues in Economic Modeling: The State of the Art

(Edited by M. Ray Perryman and James R. Schmidt) Amsterdam:

Kluwer-Nijhoff (1985).

NEW TECHNIQUES FOR DETERMINING IF A TIME SERIES CAN BE SEASONALLY ADJUSTED RELIABLY, AND THEIR APPLICATION TO U.S. FOREIGN TRADE SERIES.

David F. Findley and Brian C. Monsell
Statistical Research Division
U.S. Bureau of the Census
Washington, D.C. 20233 (U.S.A.)

David F. Findley has been Principal Researcher for Time Series Analysis in the Statistical Research Division of the Census Bureau since 1980. Brian C. Monsell has been a member of this Division's time series research group since 1982.

December 11, 1984

NEW TECHNIQUES FOR DETERMINING IF A TIME SERIES CAN BE SEASONALLY ADJUSTED RELIABLY, AND THEIR APPLICATION TO U.S. FOREIGN TRADE SERIES

by David F. Findley and Brian C. Monsell

1. INTRODUCTION

Deciding when a series is a good candidate for seasonal adjustment can be difficult. There are situations where a series may show evidence of seasonality, but because of a dominating irregular component, for example, or a volatile seasonal component, many of its seasonal factors cannot be estimated reliably. In these circumstances, the estimates of a given month's seasonal factor can change substantially when more data are added to the series and earlier data deleted.

Some seasonal adjustment programs, such as X-11 and X-11-ARIMA, provide diagnostics which can be used to help the analyst make this decision. We have found, however, that the diagnostics provided by X-11 and X-11-ARIMA are sometimes inadequate. In this article, we will discuss two new sets of measures which help to determine when a series can be seasonally adjusted reliably by a proposed seasonal adjustment methodology.

The first set, described in section 3, compares seasonal (and trading day) adjustments performed on sliding spans of data. These enable the analyst to see how stable the estimates are of seasonal factors and of month-to-month changes in the seasonally adjusted data. If too many months have unstable estimates, it is an indication that the adjustment method used cannot reliably adjust the series being examined.

The second set, described in section 5, uses the revisions history of a series to provide measures of (a) how much the initial seasonal adjustments get revised in later years and (b) how rapidly these adjustments converge to their final value. We will use these measures to determine whether the seasonal adjustments of the series being analyzed are subject to excessive amounts of revision and to help ascertain if the final adjustments are merely artifacts of the finite lengths of the adjustment filters used. In either of these situations, the seasonal adjustments are likely to be unreliable.

In the remaining sections, we use these methods in conjunction with others to analyze a number of Census Bureau series. In sections 4 and 6, a Census Bureau series called XU3 (exports of mineral fuels, lubricants and related materials) serves as an example for a detailed illustration of the use of our new techniques to determine if a series is a candidate for seasonal adjustment using X-11 (or X-11-ARIMA without the ARIMA forecasts). Then, in section 7, thirty regional foreign trade series are analyzed with these new techniques and with some conventional diagnostics. The reader who is chiefly interested in our conclusions regarding the adjustment of these series can proceed directly to sections 3, 7.1 and 7.4.

Although we use X-11 and X-11-ARIMA in this study[1], the techniques presented here can be adapted for use with other seasonal adjustment methods.

2. CONVENTIONAL ANALYSIS OF XU3

The graph of the series XU3, given in Figure 2.1, does not reveal any obvious persistent seasonal pattern, apart from a trough each December. Also, it suggests that the series undergoes a significant change around 1974. The analyst should carefully consider the question of what data span to use. For illustrative purposes, we will begin with an analysis of the full series (January 1966 to December 1983) and later give a summary of an analysis performed on the shortened series (January 1974 to December 1983).

(Insert Figure 2.1 about here)

X-11-ARIMA was used, without forecasting, to seasonally adjust the full series using 3x9 seasonal filters. Some of the diagnostics from X-11-ARIMA support seasonally adjusting the series, but others are cautionary. A summary of conventional diagnostics is given in Table 2.1. The F-tests used to detect stable seasonality tentatively suggest that there is significant seasonal variation in this series. The F-test for (linearly) moving seasonality indicates that there is no linear movement in the pattern of seasonality which would prevent its reliable estimation. (For more information on these F-tests, see [2], [3], [4].)

(Table 2.1 goes near here)

However, there are also indications that this series may not be a good candidate for seasonal adjustment. There are several signs that the series is highly irregular. The proportion of the sum of squared percent changes attributed to the irregular component is high for this adjustment (59.0 percent at lag one, 27.6 percent at lag 3 according to table F 2.B of the X-11-ARIMA analysis for XU3[2]). A graph of the SI ratios[3] for the last six years of data (given in Figure 2.2) shows how this irregularity is reflected in the spread of the values of the SI ratios for the individual calendar months. This kind of spread can lead to degraded estimates of seasonal factors obtained as weighted averages of these SI ratios.

(Insert Figure 2.2 about here)

Table 2.1 Diagnostics from X-11-ARIMA Tables for XU3

	full series (1966-83)	abridged series (1974-83)
F-test for stable seasonality, table B1	11.1	5.8
F-test for stable seasonality, table D8	15.2	7.7
F-test for moving seasonality, table D8	1.6	2.5
F-test for trading day, table C15	4.1	7.7
Relative contribution of the irregular component to the sum of squared percent changes at lag 1, table F2.B	59.0	58.7
Relative contribution of the irregular component to the sum of squared percent changes at lag 3, table F2.B	27.6	27.2

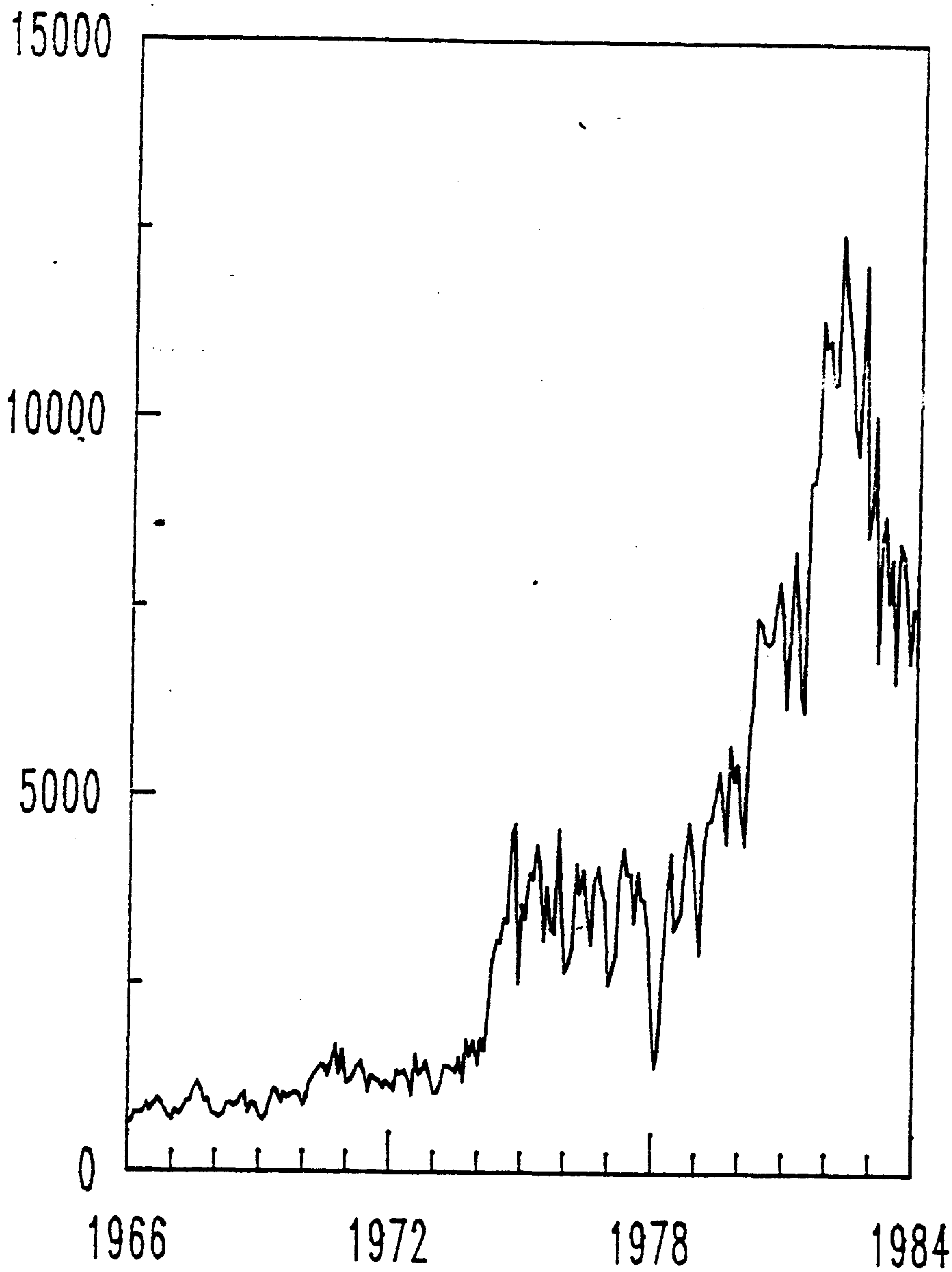
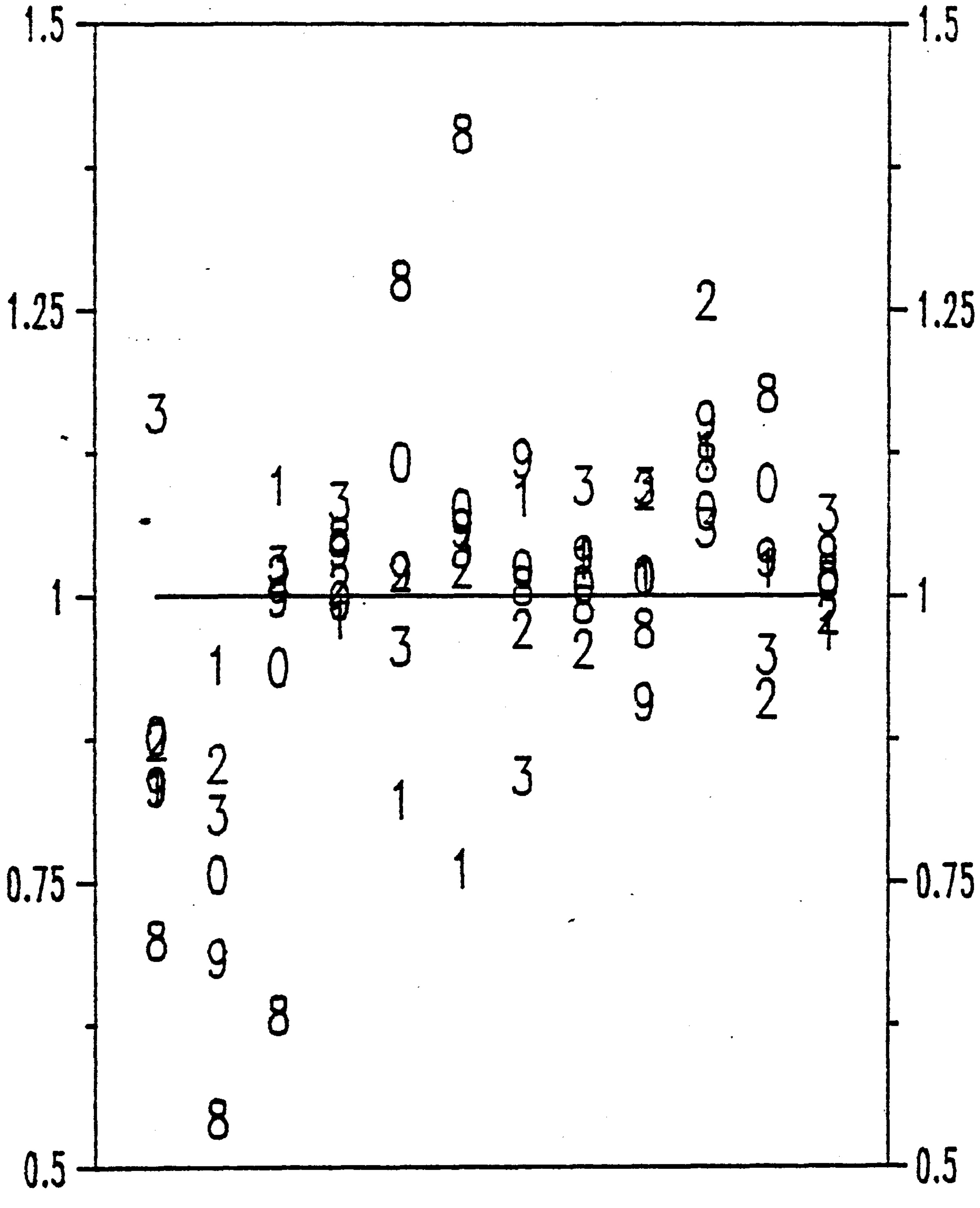


Figure 2.1 Graph of XU3



8 = 1978, 9 = 1979, 0 = 1980
1 = 1981, 2 = 1982, 3 = 1983

Figure 2.2 Year Over Year Plot of SI Ratios of XU3

Trading day adjustment was also performed on this series using X-11-ARIMA[4]. While this adjustment was accepted by X-11-ARIMA because of a significant F-statistic for trading day variation, the irregularity in the series causes concern about its reliability. Young [6] states that "[trading day] estimates made from highly irregular series cannot be expected to be useful."

Finally, the quality control statistics of X-11-ARIMA are not encouraging. X-11-ARIMA provides eleven quality control statistics to help the user evaluate the acceptability of a seasonal adjustment performed by X-11-ARIMA. These eleven statistics are combined in a weighted average to derive Q, an overall measure of the acceptability of the seasonal adjustment (see [5] for more details). If Q is less than one, the adjustment is deemed acceptable by X-11-ARIMA's criterion; if Q is greater than one, the adjustment is unacceptable[5]. For XU3, the value of Q is 1.08, casting some doubt on the adjustability of the series. (The authors have revised this measure due to an anomaly found in one of the eleven quality control statistics. Our revision of Q is described in Appendix A. The original Q's value is 0.87).

3. METHOD 1 : SLIDING SPANS ANALYSIS

A technique which we find particularly helpful for testing the reliability of a seasonal adjustment of a series is the examination of the results of seasonal adjustment for months common to a sequence of "sliding spans" within the series. These reveal how the seasonal (and trading day) adjustments vary according to which span is used.

To obtain these sliding spans, an initial span is selected based on criteria described below. Then a second span is obtained by deleting the earliest year of data from the first span and appending the year of data immediately following the last year of the first span. A third span is obtained from the second in like manner, and the process continues until there is no "future" data with which to create a new span.

(Insert Figure 3.1 about here)

Figure 3.1 illustrates the appropriate procedure for X-11 adjustment with 3x5 filters of a series which begins in January, 1974 and ends in December, 1983. Three eight-year sliding spans can be formed; one using data from 1974 to 1981, another with data from 1975 to 1982, and a third with data from 1976 to 1983. The number and length of these sliding spans will depend upon the length of the series being examined and on the length of the seasonal adjustment filter chosen by the analyst, as we explain in section 4 below. Each span is seasonally adjusted as though it were a complete series and each month common to more than one span is examined to see if its seasonal adjustments vary excessively from span to span. In Figure 3.1, for the seasonal factor of the observation occurring in January of 1981, $X_{1/81}$, we have three estimates, $S_{1/81}(1)$, $S_{1/81}(2)$ and $S_{1/81}(3)$, obtained from consecutive spans which overlap in the manner indicated. By comparing these three estimates, we can get an idea of how reliably the seasonal adjustment method is able to adjust $X_{1/81}$.

To describe how we make such a comparison, let:

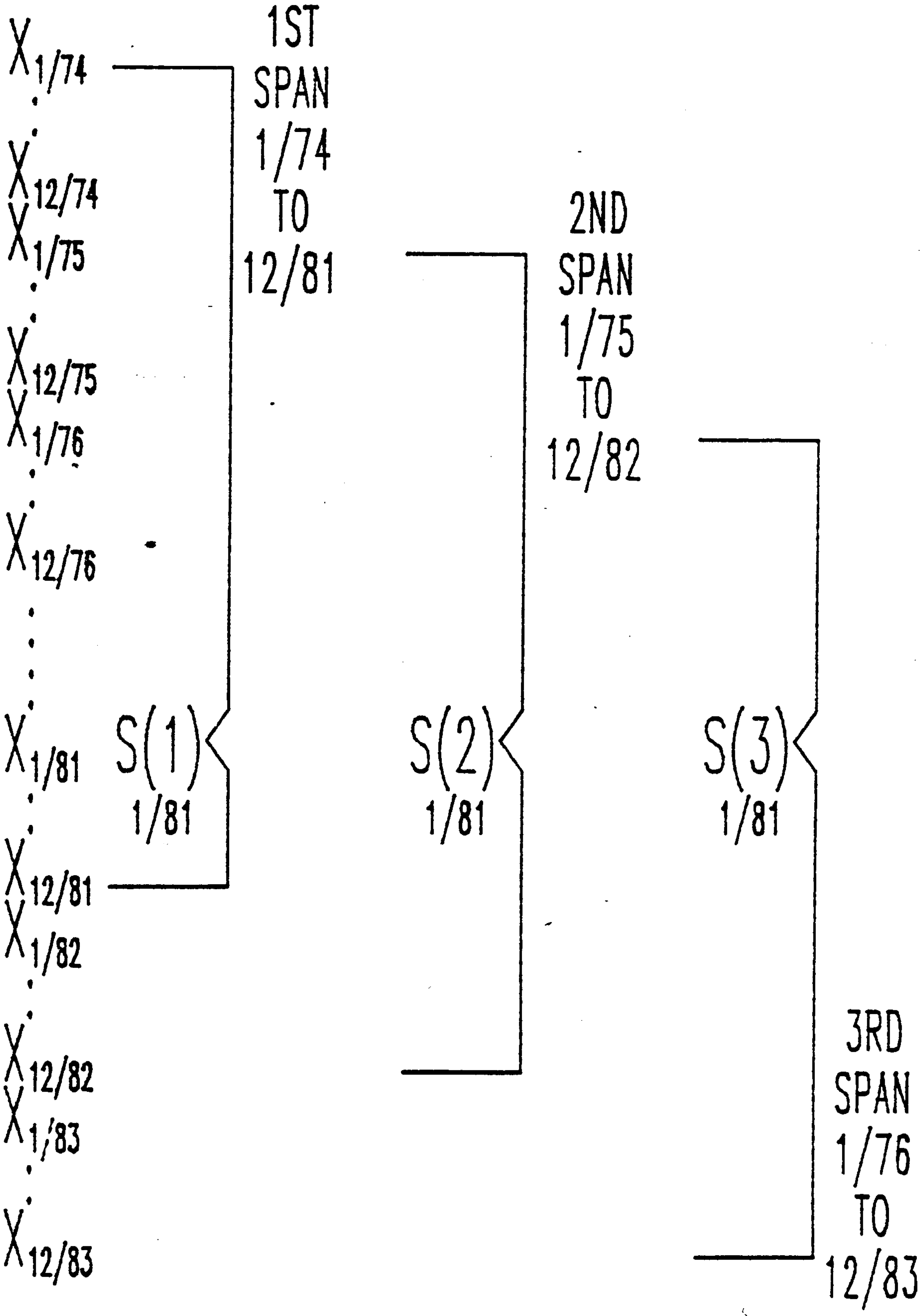


Figure 3.1 Illustration of Sliding Spans

$S_t(k)$ = the seasonal factor estimated from span k for month t;

$A_t(k)$ = the seasonally (and, usually also, trading day) adjusted value from span k for month t;

N_t = $\{k$: month t is in the k-th span$\}$

We will flag (the time series value associated with) month t as having an unreliable seasonal factor if

$$\frac{\max_{k \in N_t} S_t(k) - \min_{k \in N_t} S_t(k)}{\min_{k \in N_t} S_t(k)} > 0.03, \qquad (3.1)$$

and as having an unreliable estimate of month-to-month percentage change of the seasonally adjusted data if

$$\max_{k \in N_t} \frac{A_t(k) - A_{t-1}(k)}{A_{t-1}(k)} - \min_{k \in N_t} \frac{A_t(k) - A_{t-1}(k)}{A_{t-1}(k)} > 0.03 . \qquad (3.2)$$

Equation (3.1) tests whether the maximum percentage difference in the seasonal factors for month t is greater than 3 percent. When no trading day adjustment is done, this can be interpreted as testing whether the estimates of the level of the seasonally adjusted data vary substantially. Equation (3.2) tests whether the largest difference in the month-to-month percentage change in the seasonally (and trading day) adjusted data is greater than three percent for a month t. Often, users will seasonally adjust series mainly to get a "seasonally adjusted" value of the month-to-month percentage change. With this test, we assess the reliability of the estimate of month-to-month percentage change obtained from the seasonal adjustment method employed.

Based on our experience, the threshold values of 0.03 used above seem adequate for use with X-11 or X-11-ARIMA on most series. It may be appropriate to use different threshold values if other seasonal adjustment methods are used. Also, specific user requirements for reliability might dictate different values.

Once all months common to more than one span have been analyzed in this manner, the results can be summarized usefully in a series of tables. One table we use gives a summary for each category (seasonal factors, month-to-month percentage changes) of how many months were flagged as excessively variable as well as the percentage of months flagged. Another table shows how many times each calendar month (January, February, etc.) was flagged and how many months in each calendar year were flagged for each category. A further breakdown (a histogram) of the values of the test statistics which exceed the threshold is also given for each statistic.

If too many months are flagged (see below), it means that enough of the seasonal adjustments are unreliable to cast doubt upon the wisdom of seasonally adjusting the series. Note that an unreliable estimate of a month's seasonal factor can give rise to unreliable estimates of the two associated month-to-month changes. For this reason, there are almost always more months flagged for unreliable month-to-month changes than for the unreliable seasonal factors. One should look for frequent unreliable adjustments associated with certain calendar months and years as well. For example, problems with early years can sometimes be a sign that seasonal adjustments should be calculated from a segment of the series which does not include these years.

Trading day factors can also be analyzed in a similar manner. Let

$TD_t(k)$ = the trading day factor estimated from span k for month t

N_t = {k : month t is in the k-th span }

We will flag a month t as having an unreliable trading day factor if

$$\frac{\underset{k \varepsilon N_t}{\text{Max}} \ TD_t(k) - \underset{k \varepsilon N_t}{\text{Min}} \ TD_t(k)}{\underset{k \varepsilon N_t}{\text{Min}} \ TD_t(k)} > 0.02 . \tag{3.3}$$

Equation (3.3) tests to see if the maximum percentage variation in the trading day factor estimates associated with a given month t is greater than two percent. Again, summaries of the months flagged can be produced and broken down by year, by calendar month, and by magnitude, as we will illustrate in section 4.

If a large number of months have estimated trading day factors which have been deemed unstable, the results of the trading day regression must be considered suspect. Another frequent sign of a troublesome trading day adjustment is a high number of unacceptable month-to-month changes relative to the number of unacceptable seasonal factors. This is because the trading day factors are used to obtain the (seasonally and trading day) adjusted data, and their irregularities will usually be reflected in unstable estimates of month-to-month change in the adjusted data. In fact, in this study, we found only two series (XUASIA and XUUK, see section 7 below), both not amenable to seasonal adjustment, in which a disproportionately large number of erratic month-to-month changes were observed where this number could not be significantly reduced by eliminating the trading day adjustment.

In our investigations, we found that series which seemed to have good characteristics for seasonal adjustment usually had fewer than 15 percent of their months flagged for erratic seasonal factors, while series which had more than 25 percent of their months flagged could not be reliably adjusted. We found a "gray area" between 15 and 25 percent where a small proportion of the series in question probably could be adequately adjusted. These same threshold values are tentatively used for the percentage of months flagged for erratic trading day factors pending further investigation. Varying data user requirements make it difficult to determine analogous threshold values for erratic estimates of adjusted month-to-month changes but we certainly recommend that seasonal adjustment not be performed if more than forty percent of the estimates are flagged. The adjustor may decide to change any of the threshold values depending upon his or her own sense of how much variability can be tolerated in the adjustment, but we caution against raising the upper limits without careful study of the type of series being adjusted. Our recommendations are summarized in Table 3.1.

(Table 3.1 goes near here)

4. SLIDING SPANS ANALYSIS OF XU3

In this section, we present an analysis of the export series XU3 using the sliding spans methodology. The length of the spans was chosen to be

Table 3.1 Threshold Values for Evaluating the Sliding Spans Analysis

If less than 15 percent of the months tested are flagged for erratic seasonal factors and if the other sliding spans measures are acceptable, then	the series can probably be reliably adjusted
If more than 25 percent of the months tested are flagged for erratic seasonal factors, then	the series cannot be reliably adjusted
If between 15 and 25 percent of the months tested are flagged for erratic seasonal factors, then	the series may be adjustable : the series should be examined carefully before adjusting

The same criteria apply to the trading day factors. (provisional)

If more than 40 percent of the months tested are flagged for erratic estimates of month-to-month change, then	the series cannot be reliably adjusted

11 years, this being the length of the 3x9 seasonal filters selected for the X-11-ARIMA adjustment (ARIMA model forecasts were not used, so the adjustment procedure is essentially that of X-11). The length of the sliding spans should be at least as great as the length of the seasonal adjustment filter, since X-11 and X-11-ARIMA can produce poor adjustments even for a good series if the length of the filter selected exceeds the series length. Four sliding eleven-year spans were used for XU3. Table 4.1 shows the January, 1970 - December, 1973 section of a month-by-month analysis of the estimates of seasonal factors of XU3 for months common to several spans. The maximum percentage differences are given, along with symbols flagging months whose estimates' maximum percent difference exceeds the limit set for reliability. Different symbols correspond to the different levels of excess described in the breakdowns given in the analysis summary in Table 4.2 below.

(Tables 4.1 and 4.2 go near here)

Table 4.2 shows the results of the first analysis, which was done using four eleven-year spans. Trading day adjustment was performed if significant residual trading day variation was found in the irregulars using an F-test. First of all, note that the total number of months tested is not the same for the seasonal factors (144), the month-to-month changes (143) and the trading day factors (135). Since we are testing all months common to two or more spans, the set of months which can be examined begins after the first year of the first span and ends before the last year of the last span. For this example, the months common to more than one span fall between 1971 and 1982, a total of 144 months. There are at least two seasonal factor estimates for every month in this set. There is not, however, a month-to-month change for January 1971

Table 4.1 Example of Sliding Spans Output

Sliding Spans Analysis of Seasonal Factors for XU3

	70 80	71 81	72 82	73 83	Maximum % Diff.	
1-70	83.62	**********	**********	**********	**********	
2-70	84.78	**********	**********	**********	**********	
3-70	96.20	**********	**********	**********	**********	
4-70	102.82	**********	**********	**********	**********	
5-70	107.40	**********	**********	**********	**********	
6-70	105.29	**********	**********	**********	**********	
7-70	93.46	**********	**********	**********	**********	
8-70	105.01	**********	**********	**********	**********	
9-70	99.85	**********	**********	**********	**********	
10-70	113.69	**********	**********	**********	**********	
11-70	104.95	**********	**********	**********	**********	
12-70	103.41	**********	**********	**********	**********	
1-71	83.48	84.57	**********	**********	1.30	
2-71	84.27	84.04	**********	**********	.27	
3-71	95.37	94.49	**********	**********	.94	
4-71	103.03	104.03	**********	**********	.97	
5-71	107.80	109.98	**********	**********	2.02	
6-71	105.78	105.21	**********	**********	.54	
7-71	93.65	91.88	**********	**********	1.93	
8-71	104.74	106.20	**********	**********	1.39	
9-71	99.60	99.72	**********	**********	.12	
10-71	114.05	112.54	**********	**********	1.34	
11-71	106.00	107.59	**********	**********	1.49	
12-71	103.16	100.61	**********	**********	2.53	
1-72	82.90	83.90	80.34	**********	4.42	%%
2-72	83.51	83.40	79.90	**********	4.51	%%
3-72	94.30	93.83	92.62	**********	1.81	
4-72	103.73	104.45	103.79	**********	.69	
5-72	108.32	110.12	109.79	**********	1.66	
6-72	106.11	105.69	106.24	**********	.52	
7-72	94.18	92.56	95.73	**********	3.43	%
8-72	103.62	105.07	103.33	**********	1.68	
9-72	100.29	100.27	98.94	**********	1.36	
10-72	114.08	112.76	116.94	**********	3.71	%
11-72	107.12	108.46	112.23	**********	4.78	%%
12-72	103.30	101.14	100.66	**********	2.62	
1-73	81.90	82.68	79.79	77.95	6.06	%%%%
2-73	82.51	82.47	79.57	78.12	5.62	%%%
3-73	92.94	92.73	91.84	86.60	7.31	%%%%
4-73	104.49	105.08	104.40	104.88	.65	
5-73	109.23	110.45	110.42	109.54	1.12	
6-73	106.36	106.08	106.45	108.91	2.67	
7-73	95.10	93.77	96.30	99.69	6.31	%%%%
8-73	102.27	103.50	101.97	101.23	2.24	
9-73	100.94	101.00	99.82	98.26	2.79	
10-73	114.02	112.78	116.36	117.71	4.37	%%
11-73	108.27	109.45	112.73	113.51	4.83	%%
12-73	103.47	101.69	100.99	103.65	2.64	

Table 4.2 Sliding Spans Analysis for XU3 - with Trading Day

	S		M-M		TD	
TOTAL	51	(out of 144) (35.4 percent)	116	(out of 143) (81.1 percent)	84	(out of 135) (62.2 percent)
JANUARY	6		9		7	
FEBRUARY	7		7		2	
MARCH	9		9		7	
APRIL	0		11		8	
MAY	6		9		7	
JUNE	3		8		8	
JULY	6		9		8	
AUGUST	0		12		8	
SEPTEMBER	2		11		7	
OCTOBER	4		8		8	
NOVEMBER	6		12		7	
DECEMBER	2		11		7	
1971	0		5		0	
1972	5		7		5	
1973	9		11		7	
1974	8		11		9	
1975	4		10		8	
1976	1		12		9	
1977	2		7		7	
1978	3		9		7	
1979	5		11		7	
1980	7		12		9	
1981	6		12		9	
1982	1		9		7	

S : Number of months flagged for erratic seasonal factors
M-M : Number of months flagged for erratic month-to-month changes in the seasonally adjusted data
TD : Number of months flagged for erratic trading day factors

Breakdown of maximum % difference in seasonal factors :
- 18 months between 3 and 4 percent
- 16 months between 4 and 5 percent
- 10 months between 5 and 6 percent
- 7 months greater than or equal to 6 percent

Breakdown of maximum difference in month-to-month change :
- 40 months between 3 and 5 percent
- 26 months between 5 and 7 percent
- 35 months between 7 and 10 percent
- 15 months greater than or equal to 10 percent

Breakdown of maximum % difference in trading day factors :
- 61 months between 2 and 3 percent
- 23 months between 3 and 4 percent

in the span covering 1971 to 1981, since one cannot calculate a previous month-to-current month change for the first observation of a series. Therefore, one month fewer is tested for unreliable adjustment of month-to-month change.

While there are at least two trading day factor estimates for every month between 1971 and 1982, we must take into account the occurence of non-leap-year Februaries. Such months have the same number of Mondays, Tuesdays, Wednesdays, etc. Therefore, the trading day factor produced by X-11 (and X-11-ARIMA) for a non-leap-year February will be the same no matter what the coefficients are in the trading day regression. We adjust the total number of months tested by subtracting the number of non-leap-year Februaries, since the maximum percentage difference in the trading day factors for these months will always be zero.

One sees from Table 4.2 that a large number of months are flagged for unstable trading day factors, and an even larger number of months are flagged for unstable month-to-month changes. This immediately suggests a problem with the estimation of the trading day factors. An examination of the F-statistic used to identify the presence of trading day effects by X-11-ARIMA shows that in two of the four spans (1971-1981 and 1973-1983) the residual trading day variation was deemed not significant enough to warrant adjustment.

A spectral analysis was done of the final irregular modified for outliers for an X-11-ARIMA run of XU3 without trading day adjustment. (See [7] and [8] for more information about this analysis.) The spectrum is shown in Figure 4.1 and reveals that there are no relatively strong peaks at either of the primary trading day (alias) frequencies, 0.348 and 0.432. Combining this fact with the information from the F-tests and sliding spans, we concluded that there is not enough trading day variation present in the series for X-11-ARIMA to estimate trading day factors reliably.

(Figure 4.1 should be placed here)

We then repeated the sliding spans analysis, this time without incorporating a trading day adjustment. The results are summarized in Table 4.3. Note that although there is improvement, especially in the adjusted month-to-month changes, we still consider the percentage of erratic seasonal factor estimates to be too high (32.6 being greater than the threshold value of 25 percent mentioned earlier).

(Table 4.3 goes near here)

One final note. We mentioned before that a shortened version of XU3 might have better characteristics for adjustment. Some X-11-ARIMA diagnostics from a run done on the abridged series, using 3x5 seasonal filters, are given in the second column of Table 2.1. A sliding spans analysis was done on the abridged series, using 3 eight-year spans starting in January of 1974. However, the percentages of months flagged for erratic seasonal factors (35 percent) and erratic trading day factors (49 percent) were still too high for us to recommend adjusting the shortened series. For a more complete version of the analysis of the shortened series, see [9].

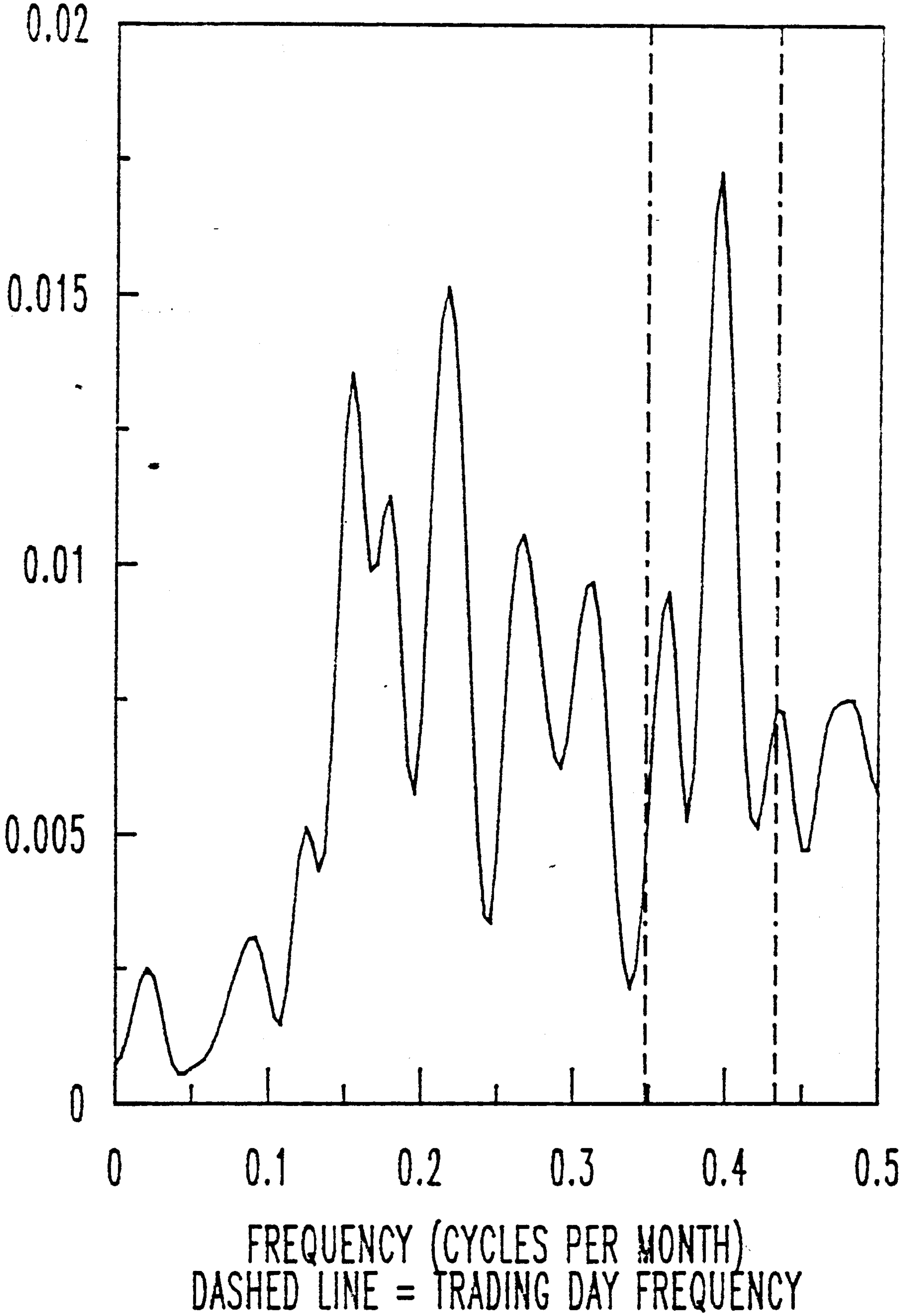


Figure 4.1 Spectrum of the Modified Irregular of XU3

Table 4.3 : Sliding Spans Analysis for XU3 - without Trading Day

	S		M-M	
TOTAL	47	(out of 144) (32.6 percent)	66	(out of 143) (46.2 percent)
JANUARY	7		8	
FEBRUARY	6		3	
MARCH	8		4	
APRIL	0		8	
MAY	5		5	
JUNE	2		2	
JULY	6		3	
AUGUST	0		8	
SEPTEMBER	3		5	
OCTOBER	3		5	
NOVEMBER	7		9	
DECEMBER	0		6	
1971	1		4	
1972	5		6	
1973	8		9	
1974	6		7	
1975	5		6	
1976	0		2	
1977	0		1	
1978	3		3	
1979	5		6	
1980	7		9	
1981	6		10	
1982	1		3	

S : Number of months flagged for erratic seasonal factors
M-M : Number of months flagged for erratic month-to-month changes in the seasonally adjusted data

Breakdown of maximum % difference in seasonal factors :
17 months between 3 and 4 percent
15 months between 4 and 5 percent
9 months between 5 and 6 percent
6 months greater than or equal to 6 percent

Breakdown of maximum difference in month-to-month change :
29 months between 3 and 5 percent
19 months between 5 and 7 percent
7 months between 7 and 10 percent
1 months greater than or equal to 10 percent

5. METHOD 2 : REVISIONS HISTORY ANALYSIS

In our second method for determining the reliability of an adjustment of a series, the revisions histories of a span of individual months of the series provide the data needed for computing two measures called CPREV and CONRAT. While this method has certain limitations which are not present in the sliding spans analysis, these measures can be quite informative.

To make it possible to produce full revisions histories for some of the months in the time interval over which the series is measured, this interval must be long enough that some (nearly) final seasonal adjustments can be calculated. Such adjustments are only available for months which are far enough away from the ends of the series that their seasonal adjustments are obtained by use of the symmetric versions of type of moving average (filter) specified. Hence, a "start up" period is required, which for XU3, we chose to be the first eleven years of the series because of the length of the 3x9 seasonal filter used in this analysis. Thereafter, for each month following the start up period for which a final adjustment can be obtained, its successive seasonal adjustments are calculated as later data are added to the series, one observation at a time.

Let $X_{i,t}$ be the seasonal adjustment for month i obtained from X-11 or X-11-ARIMA, say, using data up through month i+t. For example, $X_{i,0}$ is the concurrent adjustment for month i. Because of the finite length of the filters used, as t increases these $X_{i,t}$ converge to a final value. The number of months N until this final adjustment $X_{i,N}$ is reached depends upon the length of the seasonal filter used to adjust the data. With X-11 and X-11-ARIMA, truly final values are obtained only when N is twice the length of the seasonal filter used, but adequately "final" values can usually be obtained by choosing N to be the filter length, which is what we will do in this paper. Thus, for the analysis of XU3, we set N = 60.

We will analyze the revisions history $X_{i,0},\ldots,X_{i,N}$ for each month which falls within a preselected span after the start-up period. This span is called the experimental period.

The first quantity defined below measures the cumulative amount of revision undergone by the seasonal adjustment of a given month in the experimental period, expressed as a percentage of the concurrently adjusted value for that month. Let NOBS be the number of observations in the experimental period. Then we define

$$\mathrm{CPREV}(i) = \frac{1}{X_{i,0}} \sum_{t=0}^{N-1} |X_{i,t+1} - X_{i,t}| \; (60/N), \qquad (5.1)$$

where i = 1,...,NOBS.

The factor (60/N) at the end of (5.1) is applied to make it easier to set threshold values for CPREV which do not change with the length of the seasonal filters used. Lengthier filters take longer to produce a final estimate, and if no compensating factor were used to normalize CPREV, adjustments obtained from shorter filter lengths would usually have smaller CPREV values.

If CPREV(i) is large for a particular month i, this usually means that the seasonally adjusted value for month i undergoes frequent substantial revisions as more data become available. This is what one observes when seasonal adjustment is performed on a nonseasonal or erratically seasonal series. It is usually a sign that none of that month's adjustments can be counted upon to be reliable. If too many months have large values of CPREV(i), we will conclude that the series cannot be adequately adjusted by the methodology being used.

Another indication of the reliability of the final (and presumably best) seasonal adjustment of a given month's datum can be obtained by assessing how erratically (slowly) the preliminary seasonal adjustments converge to the final value. The intuitive reasoning goes as follows: the finite length of the adjustment filter ensures that the adjustments obtained for a given month will always converge to a final value as future data are added to the series. This occurs even if (a) there is no seasonality in the series, (b) there is a seasonal pattern which is changing too rapidly to be accurately estimated, or (c) there is a regular seasonal pattern which is too weak to measure relative to the "noise background" (irregular). Now in each of these three situations, the final adjustment is merely an artifact of the adjustment procedure and the manner in which preliminary adjustments converge to it should be noticeably more erratic than in a strongly and regularly seasonal series.

The measure CONRAT, which we use to assess the rate or manner of convergence to the final seasonal adjustment, is defined as follows. Let NOBS and N be the same as in equation (5.1). Then:

$$\text{CONRAT} = \frac{\sum_{t=0}^{N-1} \beta^{N-1-t} \left| \frac{X_{i,t} - X_{i,N}}{X_{i,N}} \right|}{\sum_{t=0}^{N-1} \beta^{t}}, \qquad (5.2)$$

where i = 1,...,NOBS, and $0 < \beta < 1$.

The weights β^{N-1-t} in (5.2) give more weight to deviations from the final seasonal adjustment occurring closer in time to the final value.

The authors have examined the value of these measures for a number of series regarded as candidates for both seasonal and trading day adjustment. Based on this experience, we consider values of CPREV > 0.18 or CONRAT > 0.01 to be signs of a series which one cannot reliably seasonally adjust with the X-11 procedure. (Our choice of β for CONRAT is described below.)

6. REVISIONS HISTORY ANALYSIS OF XU3

Table 6.1 contains the results of a revisions history analysis of XU3 using X-11 with 3x9 seasonal filters and trading day adjustment. Rather than examine each value of CPREV and CONRAT, we will consider only the minimum, maximum and mean value of these measures over the experimental period, and the number of months whose values exceed the threshold levels specified below.

(Table 6.1 goes near here)

In CONRAT, for XU3, β was selected so that

Table 6.1 Revisions History Analysis for XU3

REVISIONS MEASURES FOR XU3

MONTH	CPREV	CONRAT
1	.370	.025
2	.234	.018
3	.418	.015
4	.319	.027
5	.225	.021
6	.270	.016
7	.266	.014
8	.287	.009
9	.293	.011
10	.290	.006
11	.236	.011
12	.342	.012
13	.273	.025
14	.250	.023
15	.369	.031
16	.257	.007
17	.325	.027
18	.289	.024
19	.276	.013
20	.281	.009
21	.268	.007
22	.211	.016
23	.234	.013
24	.240	.005
AVE	.284	.016
MAX	.418	.031
MIN	.211	.005

24 out of 24 months tested (100.0 percent) had CPREV > 0.18
18 out of 24 months tested (75.0 percent) had CONRAT > 0.01

HISTOGRAM OF CPREV FOR XU3

(.180,	.195)	0	
(.195,	.210)	0	
(.210,	.225)	1	*
(.225,	.240)	4	****
(.240,	.255)	2	**
(.255,	.270)	3	***
(.270,	.285)	4	****
(.285,	.300)	4	****
(.300,	.315)	0	
(.315,	.330)	2	**
(.330,	.345)	1	*
(.345,	.360)	0	
(.360,	.375)	2	**
(.375,	.390)	0	
(.390,	.405)	0	
(.405,	.420)	1	*

HISTOGRAM OF CONRAT FOR XU3

(.000,	.002)	0	
(.002,	.005)	0	
(.005,	.007)	3	***
(.007,	.009)	3	***
(.009,	.011)	2	**
(.011,	.014)	3	***
(.014,	.016)	4	****
(.016,	.018)	1	*
(.018,	.020)	0	
(.020,	.023)	1	*
(.023,	.025)	3	***
(.025,	.027)	2	**
(.027,	.029)	1	*
(.029,	.032)	1	*

$$\beta^{N/2} = 1/2,$$

which ensures that the earlier terms in equation (5.2) are weighted substantially less than the later ones. Since N = 60, we solved $\beta^{30} = 0.5$, leading to $\beta = (0.5)^{1/30} = 0.97716$.

In examining Table 6.1, we see that the averages of CPREV and CONRAT for XU3 (0.2851 and 0.0188) are higher than the empirically derived limits given in section 5, so we are inclined to conclude that X-11's seasonal adjustment of XU3 is not acceptable, the same conclusion reached earlier on the basis of the sliding spans analysis.

Histograms of the individual values of CPREV and CONRAT are also given in Table 6.1. Note that the values for a substantial majority of the months in the experimental period are higher than the threshold values given above.

While the revisions history analysis gives valuable information, there are two drawbacks to its use. One is that this methodology requires a series with a large number of observations in order for the experimental period to have sufficient length for a meaningful analysis. The second is the fact that the most recent months are excluded from the experimental period because final adjustments are unavailable for them. For example, with XU3 the latest month in our experimental period was December, 1978. This means that we do not obtain direct information about the adjustment of data closer to the present, which would ordinarily be the data of greater interest. Because of these drawbacks, we usually analyze the revisions history only when the sliding spans analysis seems inconclusive. In the subsection 7.3 below, we will show that for the series in this study, the measured values from the sliding spans and revisions history analyses are quite highly correlated, something we have also observed with other sets of series, too.

7. ANALYSIS OF THIRTY REGIONAL FOREIGN TRADE SERIES

In this section, we present results from an analysis of 30 regional U.S. foreign trade series utilizing the sliding spans and revisions histories. We compare these two techniques with each other and with some commonly used measures found in X-11 or X-11-ARIMA.

The series investigated are from a representatively diverse set of 15 import and 15 export series. The definitions of these series are given in Appendix B. Graphs of the series can be found in Appendix C.

7.1 SLIDING SPANS RESULTS

Table 7.1 gives a summary of the sliding spans analyses performed on the 15 export series. All but two of these series begin in 1966. The exceptions, XUCOME and XUCOMEA, both begin in 1968.

For each of these series the sliding spans analysis utilized four eleven-year spans. X-11-ARIMA (without forecasts) was used to adjust the series, with 3x9 seasonal filters used in each span (as in our analysis of XU3).

(Tables 7.1 and 7.2 go near here)

Looking at the table, we see a mostly clear dichotomy between series which should and should not be seasonally adjusted. Series such as XUAFR, XUCOME, XUCOMEA and XUSASIA have a high percentage of seasonal and trading day factors flagged, while most others (XUASIA, XUDEVC, XULAR, XUOECD, XUOEEC, XUWEUR and XUWH) have a low percentage flagged. XUASIA, however, has a high percentage of unreliable estimates of adjusted month-to-month change. Only four series have erratic seasonal or trading day factor percentages in the "gray" areas between 15% and 25% (XUANEC, XUJAP, XUUK, XUWGER).

Seasonal adjustments using all of the available observations were also calculated for these series. A summary of the X-11-ARIMA diagnostics connected with these runs is given in Table 7.2. By rather lax standards, X-11's F-test for stable seasonality should exceed 7.0, the new Q statistic should be less than 1.20, and the contribution[6] of the irregular to the sum of squares of the series of percentage changes in the original data at lags one and three (F 2.B1 and F 2.B3) should be less than 50 and 30 percent, respectively, before seasonal adjustment should be contemplated. In comparing the conclusions one would draw from just examining this table with those suggested by the sliding spans analysis, one notes that two of the series, XUCOME and XUCOMEA, have acceptable values for some of the X-11-ARIMA diagnostics listed in Table 7.2 despite the striking instability of their adjustments. A further examination of Table 7.2 shows, however, that these two series both suffer from a high degree of irregularity. For example, F 2B.1, for XUCOME is calculated to be 60.71. These high values indicate that this series should not be seasonally or trading day adjusted with X-11-ARIMA. Thus, examination of only a limited set of conventional measures could lead to an incorrect conclusion. Finally, the X-11-ARIMA diagnostics strongly suggest that the four series for which the sliding spans analysis was inconclusive, XUANEC, XUJAP, XUUK and XUWGER, should not be seasonally adjusted and we accept this verdict.

Table 7.1 Results of Sliding Spans Analysis on Foreign Trade Export Series

SERIES	TD	S	S(%)	M-M	M-M(%)	TD	TD(%)	ADJUST?
XUAFR	YES	62	43.1	123	86.0	73	54.1	NO
XUANEC	YES	22	15.3	71	49.7	15	11.1	NO
XUANEC	NO	25	17.4	44	30.8	---	----	NO
XUASIA	YES	8	5.6	57	39.9	10	7.4	NO
XUASIA	NO	22	15.3	46	32.2	---	----	NO
XUCOME	YES	112	77.8	133	93.0	109	80.7	NO
XUCOMEA	YES	94	65.3	122	85.3	110	81.5	NO
XUDEVC	YES	0	0.0	6	4.2	0	0.0	YES
XUJAP	YES	25	17.4	82	57.3	31	23.0	NO
XUJAP	NO	28	19.4	48	33.6	---	----	NO(?)
XULAR	YES	3	2.1	42	29.4	17	12.6	YES*
XULAR	NO	1	0.7	4	2.8	---	----	YES*
XUOECD	YES	3	2.1	24	16.8	0	0.0	YES
XUOECD	NO	3	2.1	11	7.7	---	----	YES
XUOEEC	YES	5	3.5	61	42.7	7	5.2	NO
XUOEEC	NO	8	5.6	27	18.9	---	----	YES*
XUSASIA	YES	118	81.9	134	93.7	117	86.7	NO
XUUK	YES	27	18.8	87	60.8	36	26.7	NO
XUUK	NO	32	22.2	72	50.3	---	----	NO
XUWEUR	YES	4	2.8	21	14.7	1	0.7	YES
XUWEUR	NO	3	2.1	14	9.8	---	----	YES
XUWGER	YES	20	13.9	75	52.4	38	28.1	NO
XUWGER	NO	22	15.3	33	23.1	---	----	YES(?)*
XUWH	YES	0	0.0	2	1.4	0	0.0	YES

TD? : Are trading day factors estimated?
S : Number of months flagged for erratic seasonal factors
S(%) : Percentage of months flagged for erratic seasonal factors
M-M : Number of months flagged for erratic month-to-month changes in the seasonally adjusted data
M-M(%) : Percentage of months flagged for erratic month-to-month changes in the seasonally adjusted data
TD : Number of months flagged for erratic trading day factors
TD(%) : Percentage of months flagged for erratic trading day factors
ADJUST? : Based on our interpretation of (3)-(8) only, would we accept the X-11 seasonal adjustment (and trading day adjustment if (2) is YES) of this series? (?) indicates uncertainty. * indicates conflict with a decision based solely on our interpretation of X-11-ARIMA's quality control statistics (Table 7.2).

Table 7.2 X-11-ARIMA Diagnostics for Foreign Trade Export Series

SERIES	TD?	F-ST	F-TD	OLDQ	NEWQ	F 2.B1	F 2.B3	ADJUST?
XUAFR	YES	6.8	8.7	1.28	1.36	61.46	52.63	NO
XUANEC	YES	14.5	6.5	1.02	1.18	61.16	37.76	NO
XUANEC	NO	14.2	----	1.06	1.21	64.76	38.33	NO
XUASIA	YES	7.0	23.8	1.06	1.25	55.36	48.69	NO
XUASIA	NO	7.2	----	1.06	1.26	68.38	48.62	NO
XUCOME	YES	12.3	11.9	0.88	1.13	61.41	31.01	NO
XUCOMEA	YES	11.6	8.8	0.87	1.13	65.22	36.61	NO
XUDEVC	YES	40.3	5.0	0.51	0.76	30.57	11.56	YES
XUJAP	YES	6.6	7.4	1.08	1.30	66.57	46.66	NO
XUJAP	NO	6.1	----	1.14	1.35	73.84	48.01	NO
XULAR	YES	11.1	5.6	0.82	1.09	50.76	36.68	NO(?)*
XULAR	NO	11.0	----	0.82	1.09	53.77	37.51	NO(?)*
XUOECD	YES	23.0	8.0	0.78	1.00	47.11	23.68	YES
XUOECD	NO	21.7	----	0.86	1.07	52.77	23.72	YES(?)
XUOEEC	YES	19.3	3.4	0.89	1.10	53.00	26.97	NO(?)
XUOEEC	NO	18.6	----	0.88	1.08	54.56	27.91	NO*
XUSASIA	YES	2.2	1.0	1.57	1.57	77.28	67.29	NO
XUUK	YES	9.0	11.4	1.14	1.28	59.18	43.23	NO
XUUK	NO	7.6	----	1.33	1.43	67.48	52.45	NO
XUWEUR	YES	23.2	7.9	0.78	1.00	46.21	23.41	YES
XUWEUR	NO	21.9	----	0.85	1.06	51.53	23.56	YES(?)
XUWGER	YES	12.6	10.7	1.06	1.19	61.48	32.49	NO
XUWGER	NO	11.7	----	1.12	1.24	69.33	36.66	NO*
XUWH	YES	57.9	9.9	0.40	0.61	21.16	8.97	YES

TD? : Are trading day factors estimated?
F-ST : F-Statistic for stable seasonality from Table D 8
F-TD : F-Statistic for trading day variation from Table C 15
OLDQ : Conventional X-11-ARIMA Q statistic
NEWQ : Revised X-11-ARIMA Q statistic
F 2.B1 : Relative contribution of the irregular component to the variance of the original series at lag 1 from Table F 2.B
F 2.B3 : Relative contribution of the irregular component to the variance of the original series at lag 3 from Table F 2.B
ADJUST?: Based on X-11-ARIMA's quality control statistics alone, mainly those of this table (see text for others), would we accept X-11's seasonal adjustment (and trading day adjustment if TD? is YES) of this series? (?) indicates uncertainty. * indicates conflict with a decision based solely on our interpretation of the sliding spans analysis (Table 7.1).

Now we turn to the analysis of the fifteen import series using the sliding spans methodology. Most of these series start in 1974, the exceptions being FUDEVC and FULAR, which both begin in 1971. Three eight-year sliding spans (the first starting in 1974) were used in the analysis of these series, applying 3x5 seasonal filters for the adjustment of each span. The results of this analysis are presented in Table 7.3.

(Tables 7.3 and 7.4 go near here)

We observe that many more of these series experience problems with the trading day factors estimated by X-11-ARIMA than did the export series. Even when seasonally adjusted again without adjusting for trading day effects some of the series (FUAFR, FUCACM, FUCOME and FUSA) do not improve enough to encourage seasonal adjustment. In addition, other series (FUOECD, FUWEUR, FUWGER and FUWH), which already have acceptable values for the percentage of seasonal factors flagged when trading day adjustments are included, but have an unacceptable trading day adjustment, show a dramatic drop in the number of months flagged for erratic month-to-month changes, and little increase of erratic seasonal factors, when seasonal adjustments are recalculated without trading day adjustments. Only two series (FUASIA and FUOEEC) have acceptably low values of both percentages, erratic seasonal factors and erratic trading day factors, and even these series show substantially more stable estimates of adjusted month-to-month change when trading day adjustment is not performed.

This leaves five series (FUANEC, FUDEVC, FUJAP, FULAR and FUUK) having values for the percentage of the months flagged for extreme seasonal factors which fall in the "gray area". Two of these series, FUANEC and FUDEVC, have

Table 7.3 Results of Sliding Spans Analysis on Foreign Trade Import Series

SERIES	TD?	S	S(%)	M-M	M-M(%)	TD	TD(%)	ADJUST?
FUAFR	YES	49	51.0	72	75.8	54	60.0	NO
FUAFR	NO	39	40.6	54	56.8	--	----	NO
FUANEC	YES	22	22.9	45	47.4	1	1.1	NO*
FUANEC	NO	17	17.7	27	28.4	--	----	NO*(?)
FUASIA	YES	9	9.4	45	47.4	11	12.2	NO*
FUASIA	NO	9	9.4	19	20.0	--	----	YES
FUCACM	YES	48	50.0	75	78.9	50	55.6	NO
FUCACM	NO	37	38.7	50	52.6	--	----	NO
FUCOME	YES	57	59.4	74	77.9	64	71.1	NO
FUCOME	NO	56	58.3	63	66.3	--	----	NO
FUDEVC	YES	23	24.0	39	41.1	5	5.6	NO
FUDEVC	NO	17	17.7	28	29.5	--	----	NO(?)
FUJAP	YES	16	16.7	46	48.4	32	35.6	NO
FUJAP	NO	15	15.6	30	31.6	--	----	NO(?)
FULAR	YES	21	21.9	51	53.7	33	36.7	NO
FULAR	NO	18	18.8	38	40.0	--	----	NO
FUOECD	YES	11	11.5	43	45.3	28	31.1	NO*
FUOECD	NO	7	7.3	19	20.0	--	----	YES
FUOEEC	YES	11	11.5	42	44.2	11	12.2	NO
FUOEEC	NO	13	13.5	21	22.1	--	----	YES*
FUSA	YES	54	56.3	78	82.5	53	58.9	NO
FUSA	NO	45	46.9	56	58.9	--	----	NO
FUUK	YES	24	25.0	37	38.9	20	22.2	NO
FUUK	NO	20	20.8	20	21.1	--	----	NO(?)
FUWEUR	YES	12	12.5	36	37.9	20	22.2	NO(?)*
FUWEUR	NO	6	6.3	17	17.9	--	----	YES
FUWGER	YES	9	9.4	47	49.5	33	36.7	NO*
FUWGER	NO	5	5.2	11	11.6	--	----	YES
FUWH	YES	5	5.2	38	40.0	23	25.6	NO*
FUWH	NO	2	2.1	9	9.5	--	----	YES

TD : Are trading factors estimated?
S : Number of months flagged for erratic seasonal factors
S(%) : Percentage of months flagged for erratic seasonal factors
M-M : Number of months flagged for erratic month-to-month changes in the seasonally adjusted data
M-M(%) : Percentage of months flagged for erratic month-to-month changes in the seasonally adjusted data
TD : Number of months flagged for erratic trading day factors
TD(%) : Percentage of months flagged for erratic trading day factors
ADJUST? : Based on our interpretation of (3)-(8) only, would we accept the X-11 seasonal adjustment (and trading day adjustment if (2) is YES) of this series? (?) indicates uncertainty. * indicates conflict with a decision based solely on our interpretation of X-11-ARIMA's quality control statistics (Table 7.4).

Table 7.4 X-11-ARIMA Diagnostics for Foreign Trade Import Series

SERIES	TD?	F-ST	F-TD	OLDQ	NEWQ	F 2.B1	F 2.B3	ADJUST?
FUAFR	YES	2.3	6.7	1.24	1.51	55.44	43.07	NO
FUAFR	NO	2.3	---	1.26	1.52	61.83	45.78	NO
FUANEC	YES	10.5	4.7	0.85	0.97	46.46	20.64	YES*
FUANEC	NO	9.8	---	0.89	1.01	48.86	21.96	YES*
FUASIA	YES	8.2	5.0	0.77	0.92	36.16	20.18	YES*
FUASIA	NO	8.3	---	0.82	0.96	33.96	21.09	YES
FUCACM	YES	5.6	2.5	1.22	1.35	55.34	37.06	NO
FUCACM	NO	5.7	---	1.25	1.38	55.09	37.07	NO
FUCOME	YES	2.5	9.9	1.36	1.44	55.70	47.64	NO
FUCOME	NO	2.4	---	1.41	1.48	70.35	53.86	NO
FUDEVC	YES	2.6	7.6	1.26	1.41	46.46	29.92	NO
FUDEVC	NO	2.8	---	1.21	1.37	46.17	32.82	NO
FUJAP	YES	7.7	7.1	1.06	1.18	35.48	25.96	NO
FUJAP	NO	7.5	---	1.02	1.17	38.32	25.57	NO
FULAR	YES	2.9	3.3	1.27	1.38	50.33	41.43	NO
FULAR	NO	2.9	---	1.27	1.38	55.58	45.22	NO
FUOECD	YES	8.6	6.0	0.95	1.14	30.91	27.00	YES*
FUOECD	NO	9.1	---	0.95	1.14	35.73	25.00	YES
FUOEEC	YES	4.7	3.8	1.19	1.29	42.20	39.05	NO
FUOEEC	NO	4.6	---	1.21	1.30	47.42	40.70	NO*
FUSA	YES	2.4	6.8	1.56	1.56	57.79	41.99	NO
FUSA	NO	2.5	---	1.59	1.59	65.35	41.82	NO
FUUK	YES	5.5	6.3	1.13	1.23	44.35	40.38	NO
FUUK	NO	5.4	---	1.17	1.26	49.42	42.32	NO
FUWEUR	YES	9.9	5.4	0.91	1.10	30.94	25.29	YES*
FUWEUR	NO	10.2	---	0.96	1.15	35.18	24.30	YES(?)
FUWGER	YES	10.7	6.2	0.92	1.11	34.66	26.95	YES*
FUWGER	NO	11.2	---	0.95	1.15	31.17	28.42	YES
FUWH	YES	7.2	3.4	1.05	1.12	43.47	23.82	YES*
FUWH	NO	7.4	---	1.02	1.14	46.44	25.51	YES

TD? : Are trading day factors estimated?
F-ST : F-Statistic for stable seasonality from Table D 8
F-TD : F-Statistic for trading day variation from Table C 15
OLDQ : Conventional X-11-ARIMA Q statistic
NEWQ : Revised X-11-ARIMA Q statistic
F 2.B1 : Relative contribution of the irregular component to the variance of the original series at lag 1 from Table F 2.B
F 2.B3 : Relative contribution of the irregular component to the variance of the original series at lag 3 from Table F 2.B
ADJUST?: Based on X-11-ARIMA's quality control statistics alone, mainly those of this table (see text for others), would we accept X-11's seasonal adjustment (and trading day adjustment if TD? is YES) of this series? (?) indicates uncertainty. * indicates conflict with a decision based solely on our interpretation of the sliding spans analysis (Table 7.3).

trading day factor estimates which are acceptably stable, but unless trading day adjustment is omitted, FUANEC has too many erratic estimates of month-to-month change. For four of these five series, there is sufficient evidence in the X-11-ARIMA output for us to recommend against seasonally adjustment. FULAR and FUDEVC have a high degree of irregularity, coupled with a lack of stable seasonality (reflected by the F-test for stable seasonality). FUUK also is very irregular. FUJAP exhibits a deleterious amount of linearly moving seasonality (reflected by the moving seasonality F-test of the X-11-ARIMA program [4]). Therefore, we recommend that these four series not be adjusted. If stringent reliability requirements are not needed, an examination of the SI ratios suggests that FUANEC can be seasonally but not trading day adjusted. The summary measures from the X-11-ARIMA runs for these and the other import series can be found in Table 7.4. There is additional discussion in subsection 7.4.

7.2 REVISIONS HISTORY RESULTS

Now we turn to the results from the revisions history analysis of the regional foreign trade series. Only 13 of the 30 series, the export series excluding XUCOME and XUCOMEA, are long enough that a revisions history analysis can be performed. (The experimental period was required to contain at least two years of data.) The results for these series are given in Table 7.5. The X-11 runs used to produce the revision histories all used 3x9 seasonal filters and included a trading day adjustment if X-11's F-test did not reject it.

(Table 7.5 goes near here)

Table 7.5 Results of Revision History Analysis of Foreign Trade Series

SERIES	CPREV AVE	CPREV MAX	CPREV MIN	CPREV NOM	CONRAT AVE	CONRAT MAX	CONRAT MIN	CONRAT NOM
XUAFR	.389	.629	.166	23	.015	.006	.031	18
XUANEC	.144	.213	.104	1	.009	.005	.022	9
XUASIA	.099	.146	.066	0	.005	.014	.001	1
XUDEVC	.142	.211	.074	4	.006	.012	.003	5
XUJAP	.154	.262	.090	6	.009	.005	.020	8
XULAR	.151	.359	.072	6	.008	.020	.002	5
XUOECD	.114	.163	.074	0	.006	.013	.003	4
XUOEEC	.194	.361	.096	12	.009	.029	.003	5
XUSAISA	.435	.619	.274	24	.029	.052	.015	24
XUUK	.211	.390	.132	16	.013	.029	.004	15
XUWEUR	.120	.184	.080	1	.006	.014	.003	3
XUWGER	.194	.282	.133	16	.009	.021	.004	6
XUWH	.081	.134	.053	0	.005	.015	.002	2

NOM : Number of months in experimental period for which the measure exceeds its rejection threshold.

The thirteen series tested are all export series, and the experimental period for the series runs from January, 1977 through December, 1978. Using the criteria given in section 5, five series (XUAFR, XUOEEC, XUSASIA, XUUK and XUWGER) are found to be unsuitable for seasonal adjustment.

Two of the remaining series (XUANEC and XUJAP) have mean values of CONRAT equal to 0.009, near enough to the 0.01 threshold value for CONRAT to suggest careful examination of other measures. The values of CPREV and CONRAT for all the other series are quite acceptable. Except for XUASIA, whose revisions histories give no hint of difficulty with the seasonal adjustment, these results are quite similar to those obtained from the sliding spans analysis. However, a sliding spans analysis of XUASIA resricted to the time interval 1966-1978 containing the revisions history start-up and experimental period finds very few erratic estiamtes of level (1.5%), month-to-month change (6.1%) or trading day effect (0.0%), and the graph of XUASIA given in Appendix C suggests that the character of the series changes around 1978. In fact, the graph suggests there is little seasonality in this time interval, a conjecture supported by the X-11-ARIMA diagnostics (F-ST = 4.3, NEWQ = 1.30). Seasonal adjustment is therefore inappropriate for either the full or the shortened series.

7.3 COMPARISON OF MEASURES

In order to examine how the new measures relate to each other and to some of the more conventional measures, correlations between the measures were calculated from their values for the series analyzed in this paper. Two correlation tables are presented in Table 7.6: One, with all 31 series, shows the correlations between the sliding spans measures and the conventional X-11-ARIMA measures. The other, with 14 series (including XU3), contains correlations for those series for which we could calculate the revisions history measures. While any conclusions

Table 7.6 Correlations of Measures Used in Study

A. Correlations for all 31 series in study

	S(%)	M-M(%)	TD(%)	F-ST	F-TD	OLDQ
M-M(%)	.884					
TD(%)	.917	.897				
F-ST	-.441	-.645	-.453			
F-TD	-.086	-.091	-.096	.074		
OLDQ	.579	.670	.542	-.808	-.125	
NEWQ	.602	.697	.575	-.827	-.044	.971

B. Correlations for 14 series analyzed with revision history analysis

	S(%)	M-M(%)	TD(%)	CPREV	CONRAT	F-ST	F-TD	OLDQ
M-M(%)	.856							
TD(%)	.968	.903						
CPREV	.933	.864	.949					
CONRAT	.977	.810	.947	.944				
F-ST	-.529	-.766	-.562	-.535	-.478			
F-TD	-.427	-.257	-.379	-.445	-.516	-.093		
OLDQ	.792	.869	.762	.763	.741	-.875	-.085	
NEWQ	.719	.850	.703	.695	.668	-.940	-.030	.985

S(%) : Percentage of months flagged for erratic seasonal factors
M-M(%) : Percentage of months flagged for erratic month-to-month changes in the seasonally adjusted data
TD(%) : Percentage of months flagged for erratic trading day factors
CPREV : Revisions history measure CPREV
CONRAT : Revisions history measure CONRAT
F-ST : F-statistic for stable seasonality taken from X-11-ARIMA table D 8
F-TD : F-statistic for trading day taken from X-11-ARIMA table C 15
OLDQ : Conventional Q measure from X-11-ARIMA
NEWQ : Revised Q measure from X-11-ARIMA

drawn from such an analysis should be considered preliminary, we feel that three points should be made. First, there appears to be a strong positive correlation between the sliding spans measures and the revisions history measures. This would seem to say that these measures capture very similar information. This is reassuring, because for many series we can only compute the sliding spans measures due to the length of the series. Second, the conventional measures, with the exception of the F-test for trading day, display stronger relationships among themselves than with the new measures. Finally, the F-test for trading day and the percentage of months flagged for erratic trading day factors seem to have very little, if any, correlation. We will return to this point in a later paper where we shall show that direct trading day regression with seasonal ARIMA error models can sometimes provide more persuasively reliable trading day adjustments than X-11's simple trading day model.

(Table 7.6 goes near here)

7.4 THE UTILITY OF THE NEW METHODS

Unlike the conventional diagnostic measures of X-11-ARIMA, the new measures, with the exception of CONRAT, are directly interpretable as quantities of interest to many, perhaps most, producers of seasonally adjusted data. To the extent that they lead to similar conclusions, which usually happens, they render the decisions suggested by the more traditional diagnostics more intelligible. Even in these situations, they frequently offer valuable supplementary information, as they do in revealing that the series FUASIA, FUOECD, FUWGER, and FUWH, while seasonally adjustable, should not be trading day adjusted by X-11, in contradiction to what is suggested by X-11's F-test. Indeed, the results of Tables 7.1-7.4 show that this F-test has little value for predicting when the trading day adjustments calculated by X-11 or X-11-ARIMA will be reliable.

The series XULAR, XUOECD, XUOEEC, XUWEUR and FUWH seem marginal by the standards of X-11-ARIMA and quite adjustable according to the new measures, providing trading day adjustment is omitted in the case of XUOEEC and FUWH, and perhaps also of XULAR.

The greatest disparity occurs with FUOEEC, which is nonseasonal and quite irregular according to X-11-ARIMA and reasonably adjustable for seasonal but not for trading day variation, as far as the new methods can suggest. In this case, we mildly favor X-11-ARIMA's conclusion. Thus we do not recommend using only the new measures.

8. FINAL REMARKS

A subject which clearly needs to be examined further is the effect of trading day adjustment on estimates of the seasonally adjusted month-to-month change. In our sliding spans analysis, we noticed that some series with rather unstable estimates of the month-to-month change in the seasonally adjusted series have stable estimates for the seasonal factors. For example, XUOEEC had only 5 months flagged for erratic seasonal factors, but 61 months flagged for erratic month-to-month changes. When we repeated this analysis without adjusting for trading day, the number of months flagged for erratic month-to-month changes dropped dramatically, despite the fact that XUOEEC had only seven of its months flagged for erratic trading day factors. Similar occurences in series such as XULAR and XUOECD give the impression that sometimes even an apparently good trading day adjustment can lead to erratic estimates of the month-to-month changes. There is, of course, the possibility that the criterion (3.3) is not sensitive enough to problems with the trading day adjustment. At present, it seems more likely to us that there are some problems with X-11's approach to trading day adjustment which will be illuminated by a time series modeling approach we have begun to explore.

As we have mentioned before, the revisions history measures have some limitations. First, these measures cannot be calculated if the series is too short. Shorter start-up periods have been tried, but tend to implausibly inflate the resulting measure. Second, as pointed out in [10], there are some seasonal adjustment techniques that do not have seasonal factors which converge to a final value in any practical sense. In this case, CONRAT becomes meaningless, as there is no final value to converge to. A possible solution to this might be to pick a "pseudo-final" value a fixed number of months in the future from the concurrent estimate for each month in an experimental period selected by the user. Third, the values of CPREV and CONRAT sometimes seem to be affected significantly by whether or not the series has been trading day adjusted. For example, we recall from Table 6.1 that the values of CPREV and CONRAT for XU3 were 0.2851 and 0.0188, respectively. This was from seasonally adjusted data which was also calendar adjusted, because the X-11 F-statistic's value of 4.1 suggested the trading day effect was large enough. If this analysis is done without the calendar adjustment, we obtain CPREV = 0.18 and CONRAT = 0.016 for XU3. When trading day adjustment was omitted, such decreases were observed for almost all of the 14 series for which revisions histories were calculated. This suggests that the threshold values for CPREV and CONRAT presented earlier may not be suitable for series not adjusted for trading day effects. This phenomenon needs further scrutiny.

The opinions expressed in this paper are those of the authors and do not necessarily reflect Census Bureau policy or practice.

BIBLIOGRAPHY

[1] Monsell, B. C. (1984). The substantive changes in the X-11 procedure of X-11-ARIMA. Bureau of the Census, Statistical Research Report No. CENSUS/SRD/RR-84/10.

[2] Shiskin, J., Young, A. H. and Musgrave, J. C. (1967). The X-11 variant of the Census Method II seasonal adjustment program. Bureau of the Census, Technical Paper 15.

[3] Dagum, E. B. (1983). The X-11-ARIMA Seasonal Adjustment Method. Ottawa : Statistics Canada.

[4] Morry, M. and Lothian, J. (1978). A test for the presence of identifiable seasonality when using the X-11 program. Statistics Canada, Seasonal Adjustment and Time Series Staff, Research Paper.

[5] Lothian, J. and Morry, M. (1978). A set of quality control statistics for the X-11-ARIMA seasonal adjustment program. Statistics Canada, Seasonal Adjustment and Time Series Staff, Research Paper.

[6] Young, A. H. (1965). Estimating trading-day variation in monthly economic time series. Bureau of the Census, Technical Paper 12.

[7] Cleveland, W. S. and Devlin, S. J. (1980). Calendar effects in monthly time series : detection by spectrum analysis and graphical methods, Journal of the American Statistical Association, 75, pp. 487-498.

[8] Cleveland, W. S. (1983). Seasonal and calendar adjustment, in Handbook of Statistics Volume 3 : Time Series in the Frequency Domain, (eds. D. R. Brillinger and P. R. Krishnaiah), Amsterdam : North-Holland, pp. 39-72.

[9] Monsell, B. C. and Findley, D. F. (1984). Techniques for determining if a seasonal time series can be seasonally adjusted reliably by a given seasonal adjustment methodology. American Statistical Association Proceedings of the Business and Economic Statistics Section.

[10] Bell, W. R. and Hillmer, S. C. (1984). Issues involved with the seasonal adjustment of time series, Journal of Business and Economic Statistics, 2, pp. 291-320.

FOOTNOTES

1 Even if ARIMA forecasts are not used in the adjustment process, X-11-ARIMA uses a modified X-11 procedure which leads to (usually) slightly different seasonally adjusted values than are obtained from Census X-11. For more information on the differences between X-11 and X-11-ARIMA, see [1]. For specific information about these two programs, see [2] for Census X-11 and [3] for X-11-ARIMA.

2 Traditionally, analysts have used mainly the relative contribution of the lag one percent changes of the irregular to the sum of squared lag one percent changes of the original series to measure how much irregular is present in the series. However, an analysis of transfer functions performed by Lothian and Morry [5] led these authors to suggest that the third lag provides the best indication of how strong the irregular is relative to the seasonal component. We will give the values of both measures in this paper.

3 The SI ratios of a given X-11 (or X-11-ARIMA) seasonal adjustment are the detrended (and, when appropriate, also trading day adjusted) values. They are called "ratios" because, in the simple multiplicative model (series = trend x seasonal x irregular), these values are derived by dividing the original data adjusted for extremes by an estimate of the trend. The X-11 procedure obtains its seasonal factors as weighted averages of these SI ratios. For more information, see [2] and [3].

4 While the final version of this paper was being prepared, an error was discovered in the X-11-ARIMA program affecting the trading day adjustment of the treatment of leap-year Februaries. The computations listed in this paper were done using the uncorrected X-11-ARIMA. The number of leap-year

Februaries is small enough that different adjustments for them would not affect our conclusions. For more information on this error, contact the authors or Statistics Canada.

5 Our experience in using the Q-measure from X-11-ARIMA has suggested a different strategy for evaluating Q from the one described in [5]. A series is to be considered adjustable if Q is less than 0.8. If Q is greater than 1.2, then the series probably shouldn't be adjusted. If Q falls between 0.8 and 1.2 additional diagnostics should be examined before a recommendation is made regarding its adjustability.

6 These measures should probably be replaced by more robust ones. Being ratios of sums of squares, they can be strongly influenced by outliers, which makes it difficult to set threshold values for them.

FIGURE CAPTIONS FOR GRAPHS

Figure 2.1 Graph of XU3

Figure 2.2 Year Over Year Plot of SI Ratios of XU3

Figure 3.1 Illustration of Sliding Spans

Figure 4.1 Spectrum of the Modified Irregular of XU3

APPENDIX A. A REVISION OF X-11-ARIMA's Q-STATISTIC

The X-11-ARIMA quality control statistic M2 uses information from table F2.F of the X-11-ARIMA output to evaluate the relative contribution of the irregular component to the variation of the series about a fitted mean function (referred to as the "stationary portion of the variance" in X-11-ARIMA). In X-11-ARIMA, the mean function used is a straight line fit to the final trend estimates given in table D12 of the X-11-ARIMA output (or the logarithm of this table if the series is being adjusted multiplicatively). This mean function is then subtracted from the original data given in table B1 (we will call the result B1') and from the final trend cycle stored in D12 (we will call the result D12'). Then, the entries in table F2.F are calculated as follows:

$$RS_I = \frac{\text{Var(final irregular from table D13)}}{\text{Var(B1')}} ,$$

$$RS_S = \frac{\text{Var(final seasonal from table D10)}}{\text{Var(B1')}} ,$$

$$RS_C = \frac{\text{Var(D12')}}{\text{Var(B1')}} ,$$

$$RS_P = \frac{\text{Var(monthly prior factors from table A2)}}{\text{Var(B1')}} ,$$

$$RS_{TD} = \frac{\text{Var(final trading day factors from table C16.B)}}{\text{Var(B1')}} .$$

Using the values given above, M2 is defined as

$$M2 = (RS_I / (100 - RS_P)) / 0.10 . \qquad (a)$$

If M2 is greater than 1, the variation of the irregular component is deemed excessive. If the result of (a) is greater than 3, M2 is set equal to 3.

Lothian and Morry [4] state that "the average series adjusted had a cycle which contributes about 5 to 10% to the stationary portion of the variance. The threshold level for the M1 and M2 statistics are based on this assumption." However, many Census Bureau series which can be reliably adjusted for seasonal variation show much higher contributions by the trend cycle than are allowed for by the criteria of Lothian and Morry. After examining graphs of the series and the X-11-ARIMA output for such series, we felt that a straight line was not an adequate mean function for the purposes of the M2 statistic.

In our revised procedure, a linear spline (a continuous piecewise linear function) is fit by least squares to the final trend cycle estimates, instead of a straight line. The spline is constrained to be a linear function for each January through December period.

This "spline trend" is used instead of the X-11-ARIMA's linear mean function to produce new B1' and D12' series and the calculation of the F2.F table and the revised M2 then proceeds as before. A new value of Q is calculated, using the revised M2.

In the first two columns of Table A we compare the results of these two methods for a Census Bureau series WFURN (wholesale furniture sales from January, 1967 to December, 1979). Note how the X-11-ARIMA method does not eliminate the trend cycle variation in the stationary series as well as the "spline trend" does (RS_C is 28.97 versus the new value of 7.74). However, since this series shows a strong degree of seasonality (high RS_S values), the irregular component's contribution is not affected very much. Therefore, the impact on the value of Q is minimal.

The last two columns of Table A also compare the two methods as they are applied to XU3. There are large differences in the contribution of the trend cycle (73.53 versus 13.46 for the new method) and irregular (12.67 versus 35.80 for the new method) components. This has a significant effect on the values of M2 and Q, with Q going from 0.87 (a sign that the seasonal adjustment is acceptable) for the standard X-11-ARIMA method to 1.08 (a sign of an unacceptable adjustment) for the revised method.

It seems likely that we will recommend a further modification of Q after additional studies have been completed. The M1 component of Q, which, in our notation, is defined to be min{F 2.B3/10.0 , 3.0}, was motivated by Lothian and Morry's experience that the tabled F 2.B3 value for the irregular is usually less than 10.0 for series which can be well adjusted. The better series in this study suggest that a higher threshold than 10.0 could reasonably be used for F 2.B3. Such a change would usually lead to lower Q values.

(Table A goes near here)

Table A Comparison of M2 Results

	WFURN		XU3	
	X-11-ARIMA procedure	Revised M2	X-11-ARIMA procedure	Revised M2
RS_I	3.69	4.73	12.67	35.80
RS_C	28.97	7.74	73.53	13.46
RS_S	58.69	76.65	15.19	42.93
RS_P	0.00	0.00	0.00	0.00
RS_{TD}	12.45	15.99	0.32	0.90
Total	104.80	105.11	101.71	93.10
M2	0.369	0.473	1.267	3.000
Q	0.25	0.26	0.87	1.08

APPENDIX B. CLASSIFICATION OF ALL FOREIGN TRADE SERIES

XUAFR	Exports to Africa
XUANEC	Exports to Asia NEC
XUASIA	Exports to Asia
XUCOME	Exports to Communist Areas in Europe (1/68 to 12/83)
XUCOMEA	Exports to Communist Areas in Europe and Asia (1/68 to 12/83)
XUDEVC	Exports to Developed Countries
XUJAP	Exports to Japan
XULAR	Exports to Latin American Republics
XUOECD	Exports to the members of the Organization for Economic Cooperation and Development
XUOEEC	Exports to Members of European Common Market except the United Kingdom and West Germany
XUSASIA	Exports to South Asia
XUUK	Exports to the United Kingdom
XUWEUR	Exports to Western Europe
XUWGER	Exports to West Germany
XUWH	Exports to the Western Hemisphere
FUAFR	Imports from Africa
FUANEC	Imports from Asia NEC
FUASIA	Imports from Asia
FUCACM	Imports from the Central American Common Market
FUCOME	Imports from Communist Areas in Europe
FUDEVC	Imports from Developed Countries (1/71 to 12/83)
FUJAP	Imports from Japan
FULAR	Imports from Latin American Republics (1/71 to 12/83)
FUOECD	Imports from members of the Organization for Economic Cooperation and Development
FUOEEC	Imports from Members of European Common Market except the United Kingdom and West Germany
FUSA	Imports from South Asia
FUUK	Imports from the United Kingdom
FUWEUR	Imports from Western Europe
FUWGER	Imports from West Germany
FUWH	Imports from the Western Hemisphere

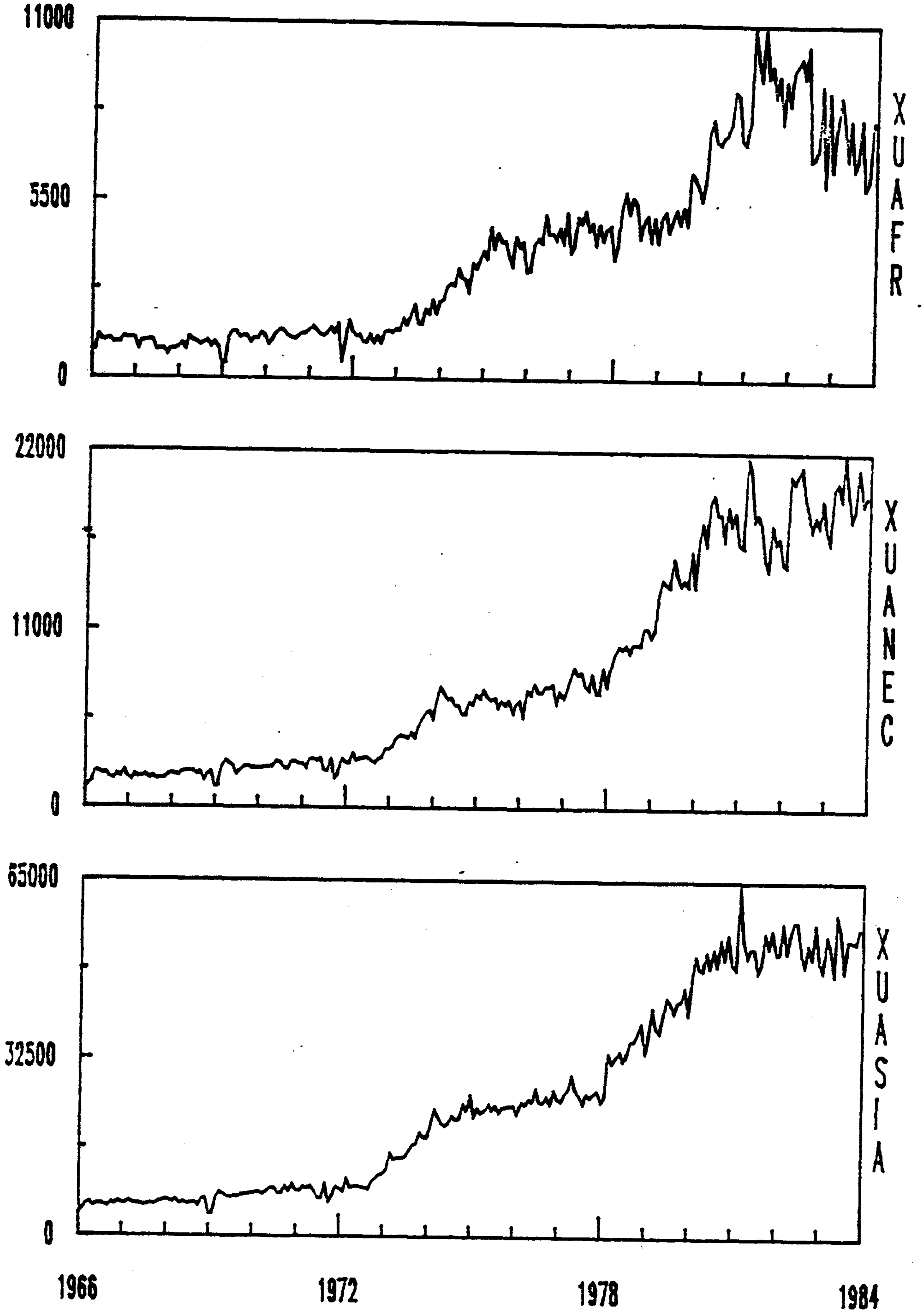
NOTES :

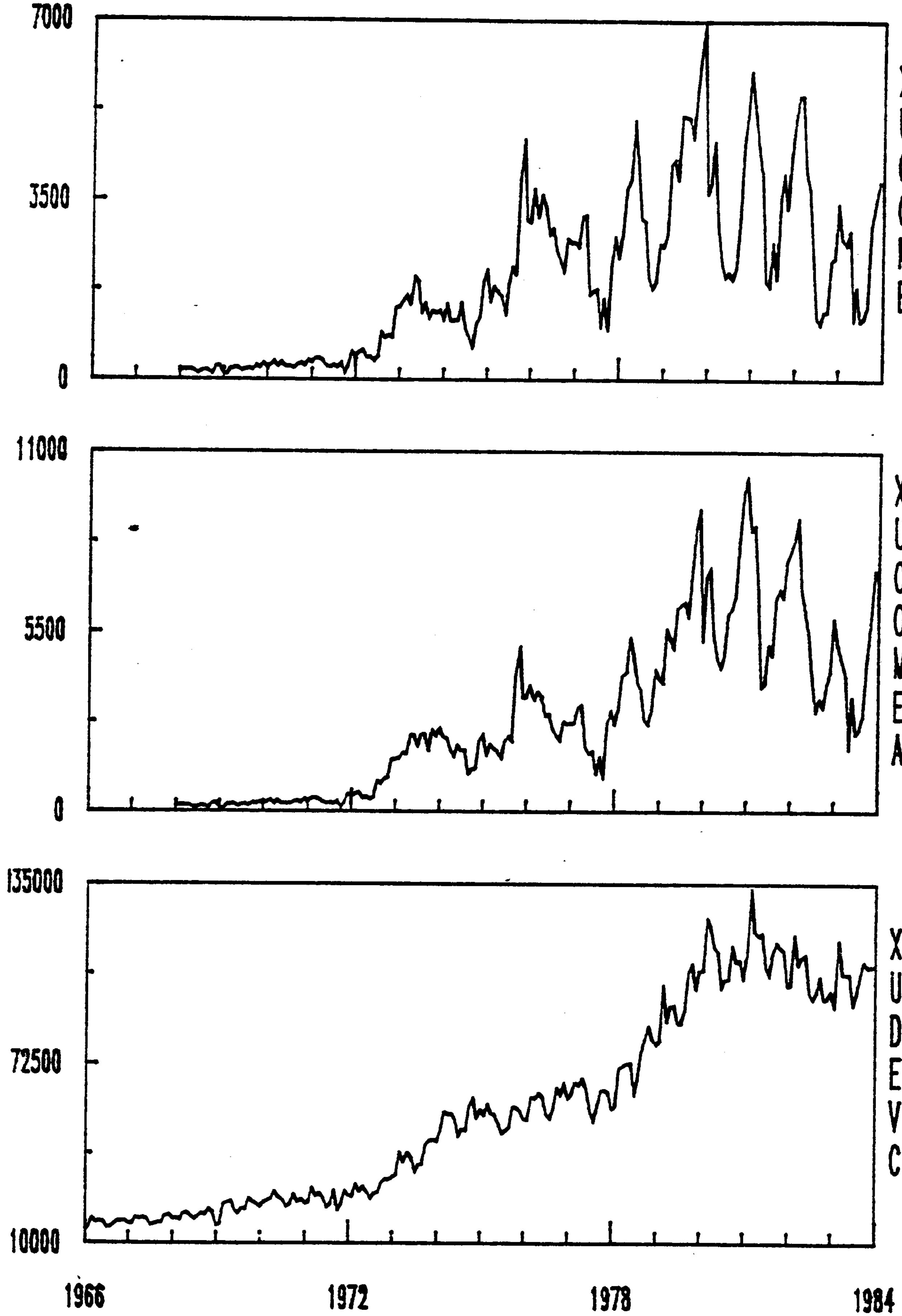
(1) - All series are expressed in terms of hundreds of thousands of dollars. Also, except as noted, all the export series start in January, 1966 and end in December 1983, and all of the import series start in January, 1974 and end in December, 1983.

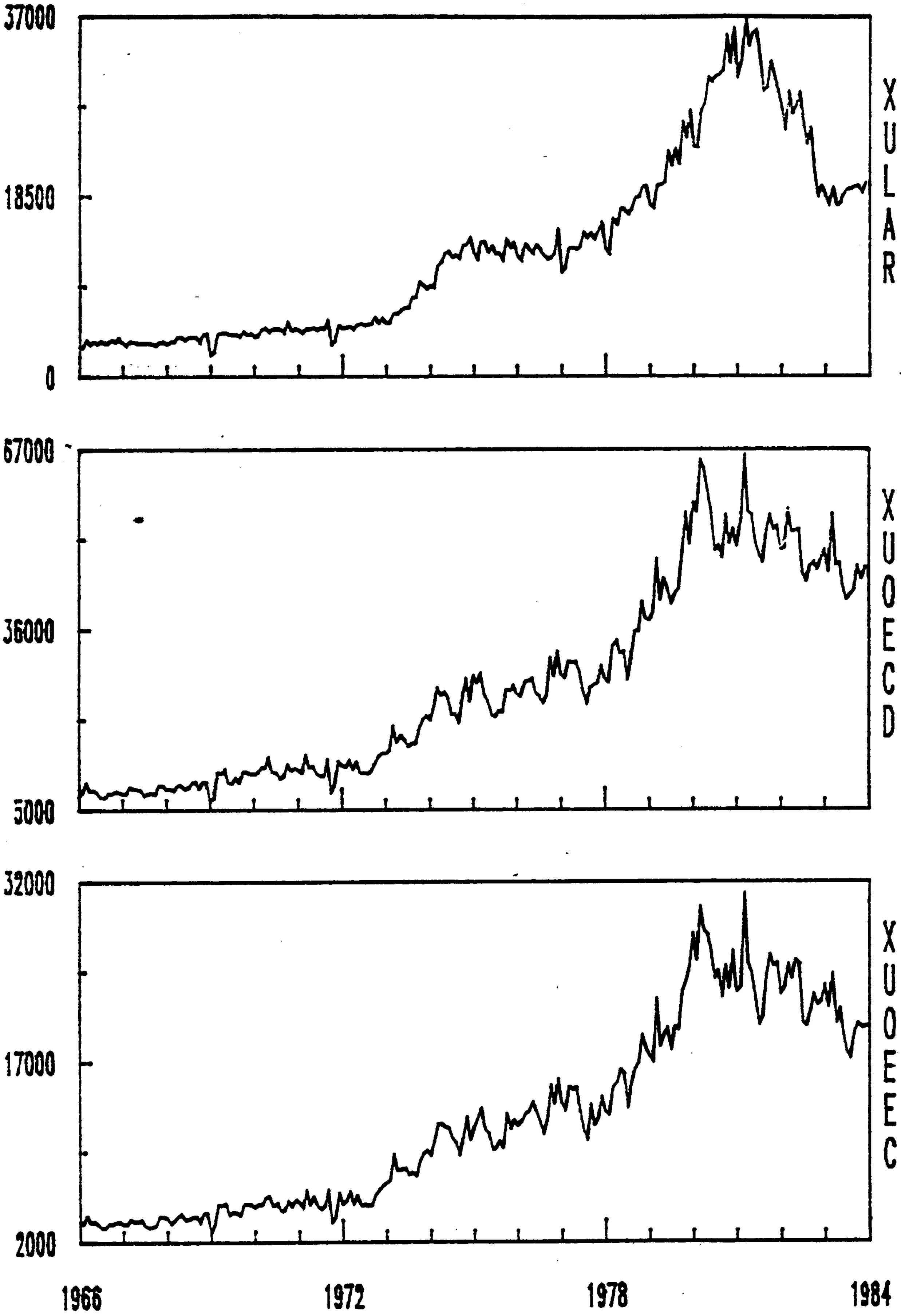
(2) - The NEC in XUANEC and FUANEC refers to countries not contained in certain subclassifications. Countries whose data are combined to form these two series include Burma, Malasia, Thailand, Campuchia, Korea, Hong Kong, 49. al.

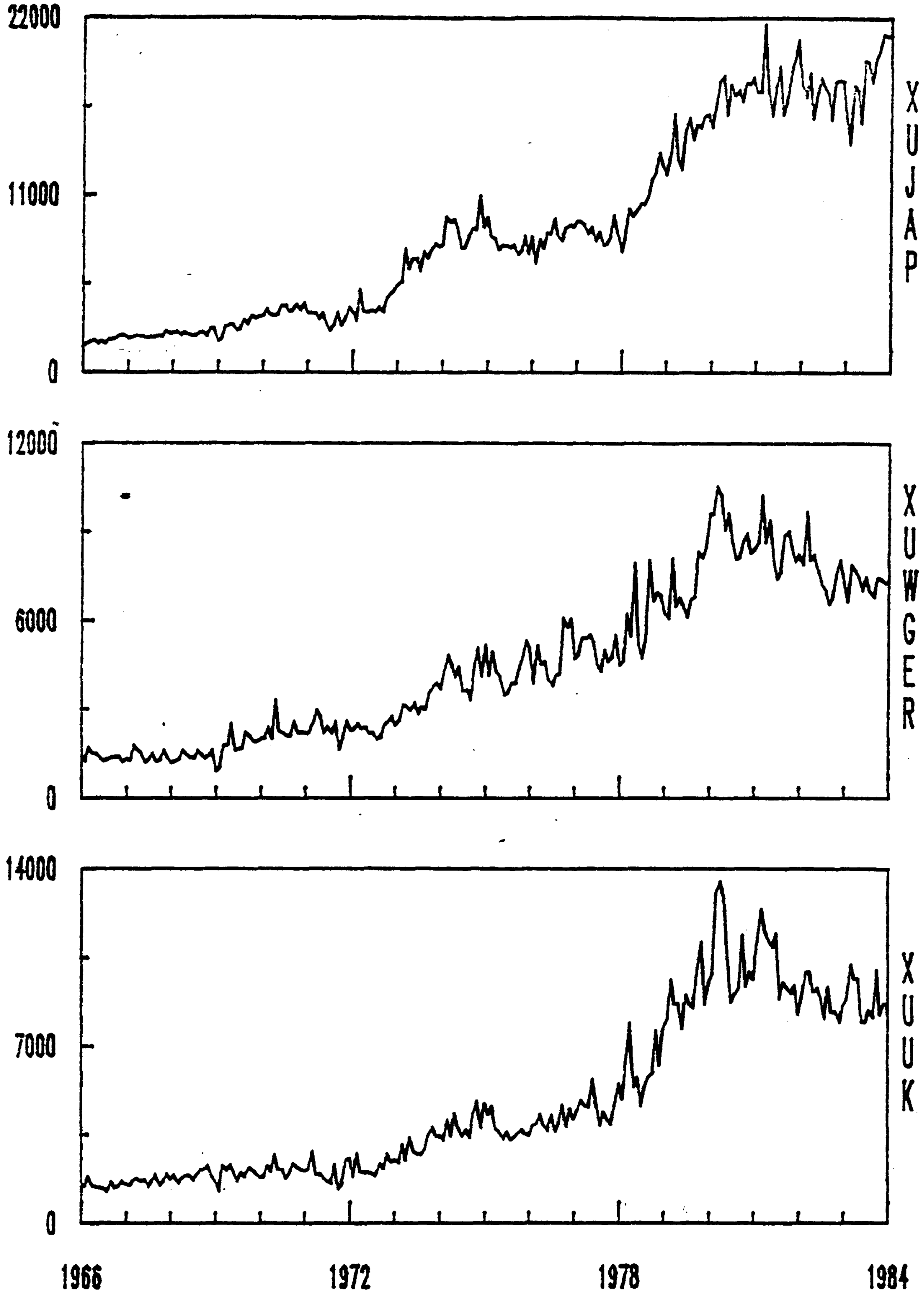
APPENDIX C. GRAPHS OF REGIONAL FOREIGN TRADE SERIES

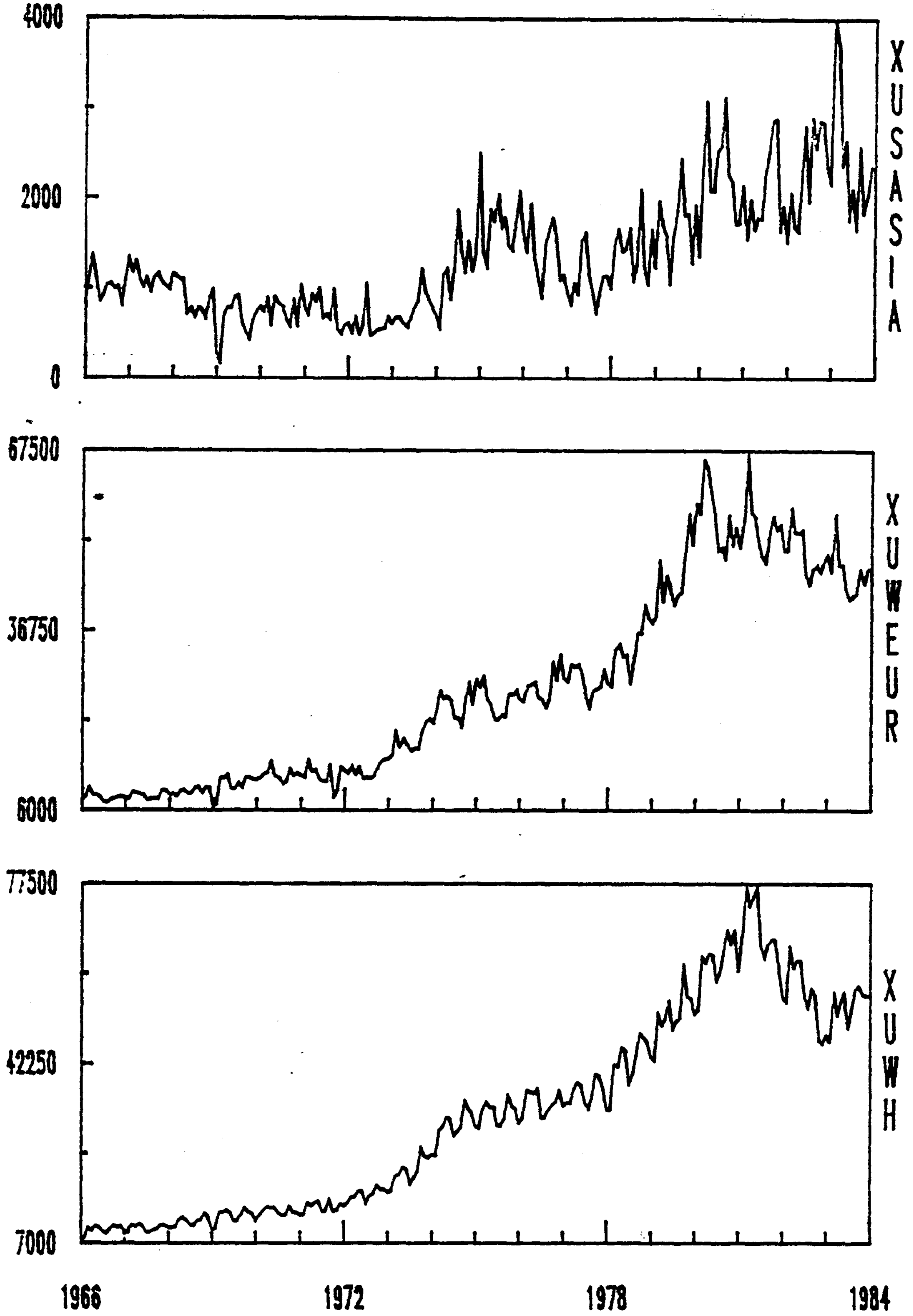
The graphs of this Appendix depict the data from which seasonal factors are calulated. Occasionally an individual month's datum in one of these graphs may differ from a published unadjusted value, reflecting a modification made to counteract the influence of a strike or a similarly atypical disruption.

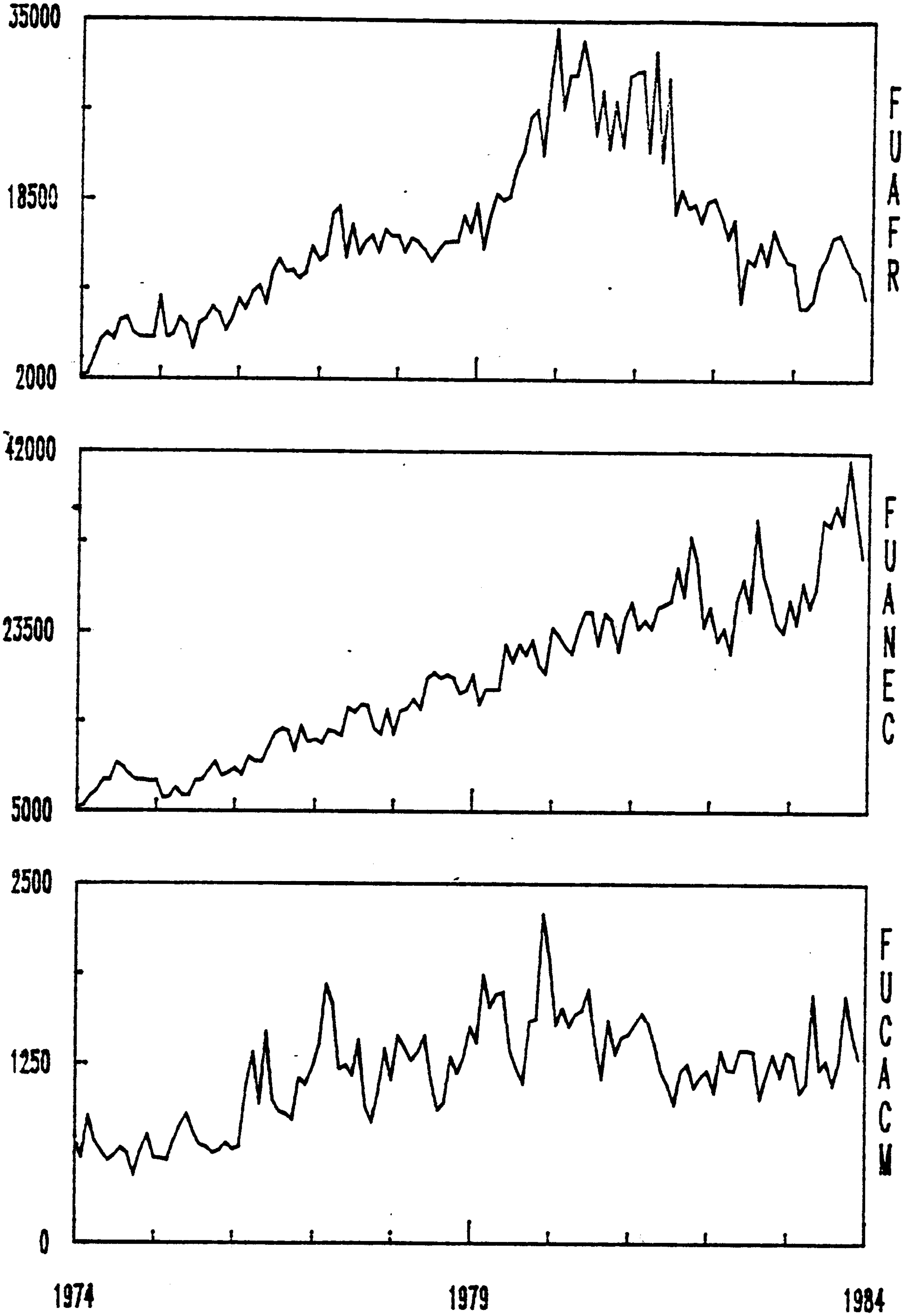

11000
5500
0
XUAFR
22000
11000
0
XUANEC
65000
32500
0
XUASIA
1966
1972
1978
1984

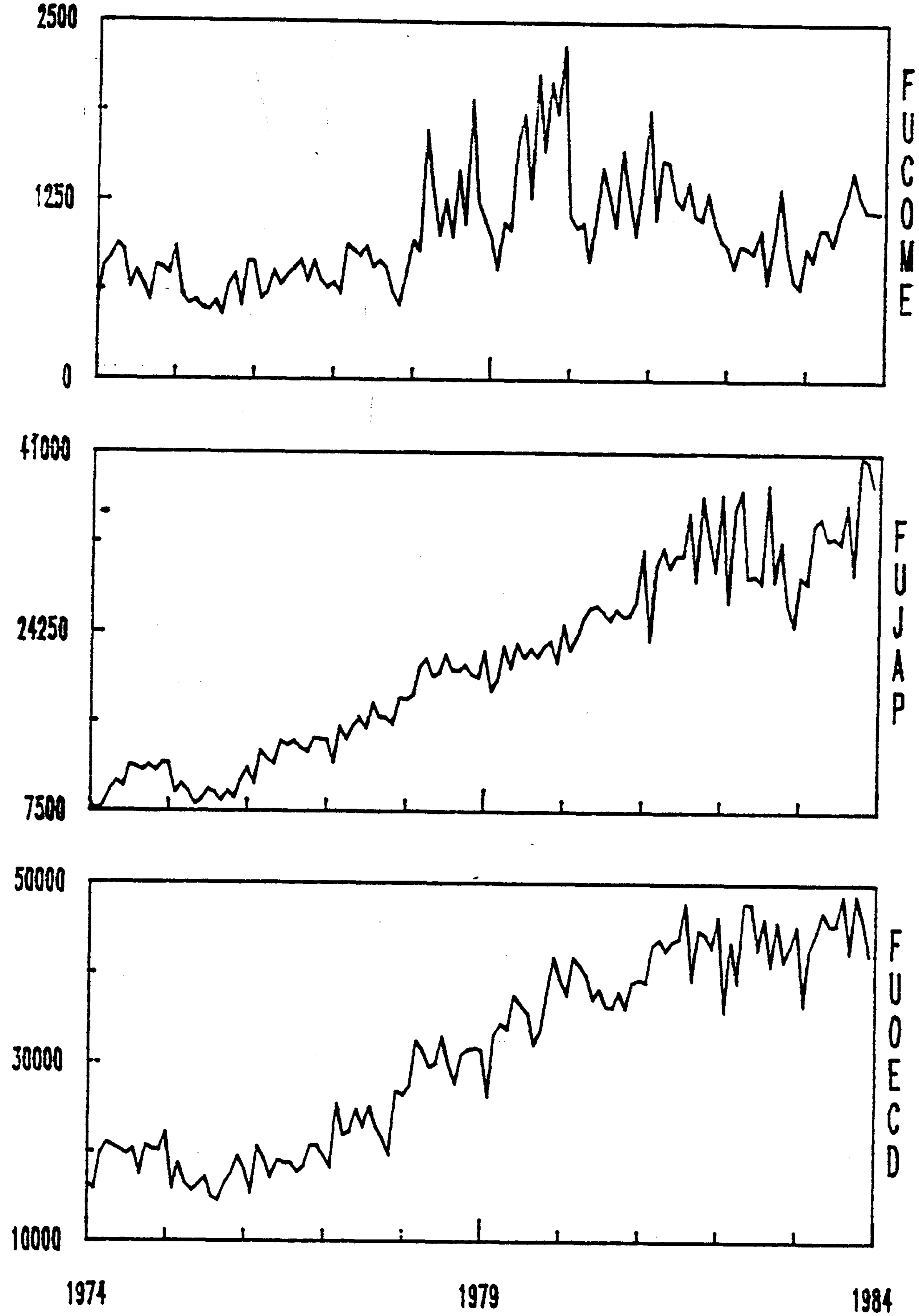

7000
3500
0
XUCOME
11000
5500
0
XUCOMEA
135000
72500
10000
XUDEVC
1966
1972
1978
1984

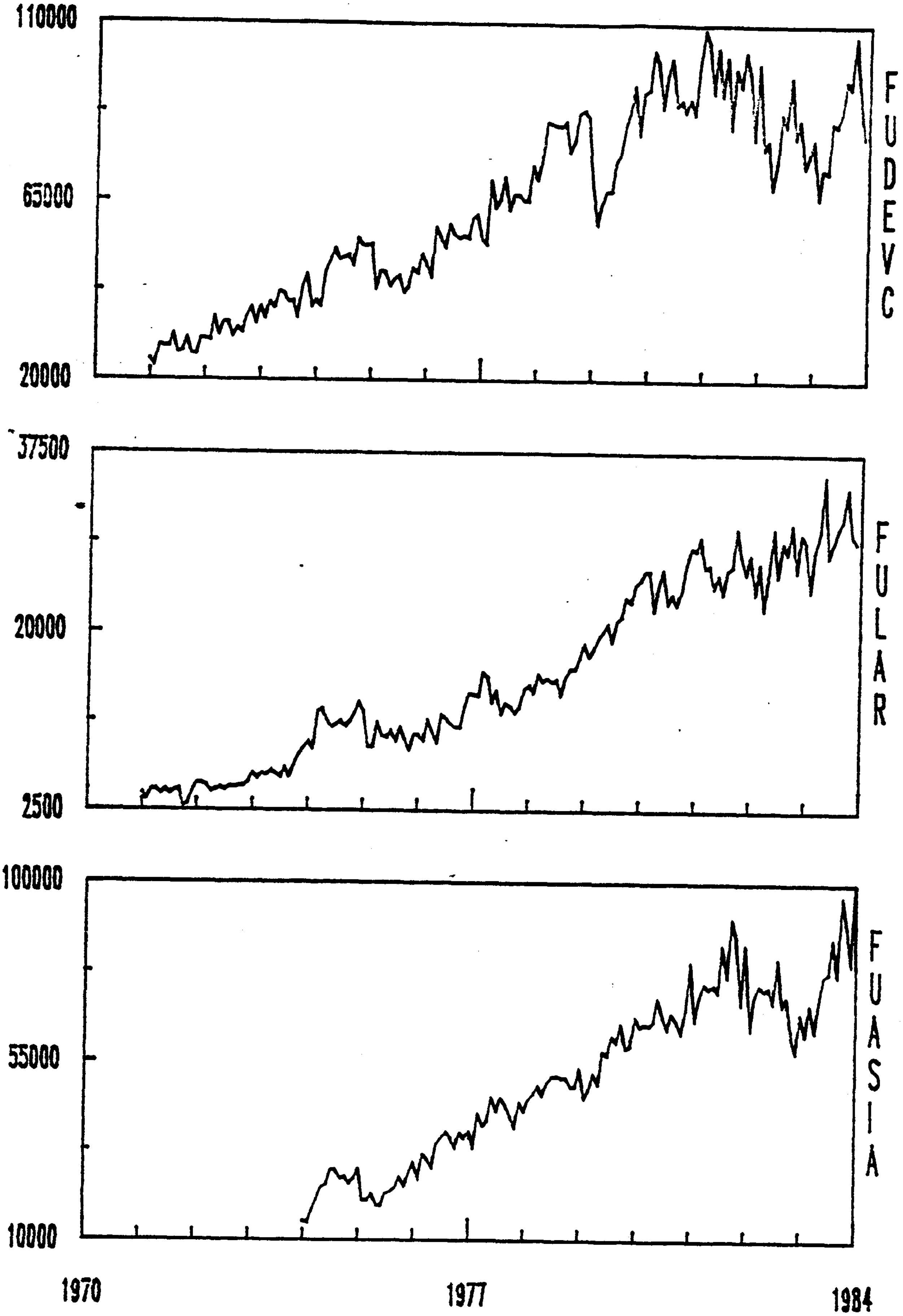

37000
18500
0
XULAR
67000
36000
5000
XUOECD
32000
17000
2000
XUOEEC
1966
1972
1978
1984

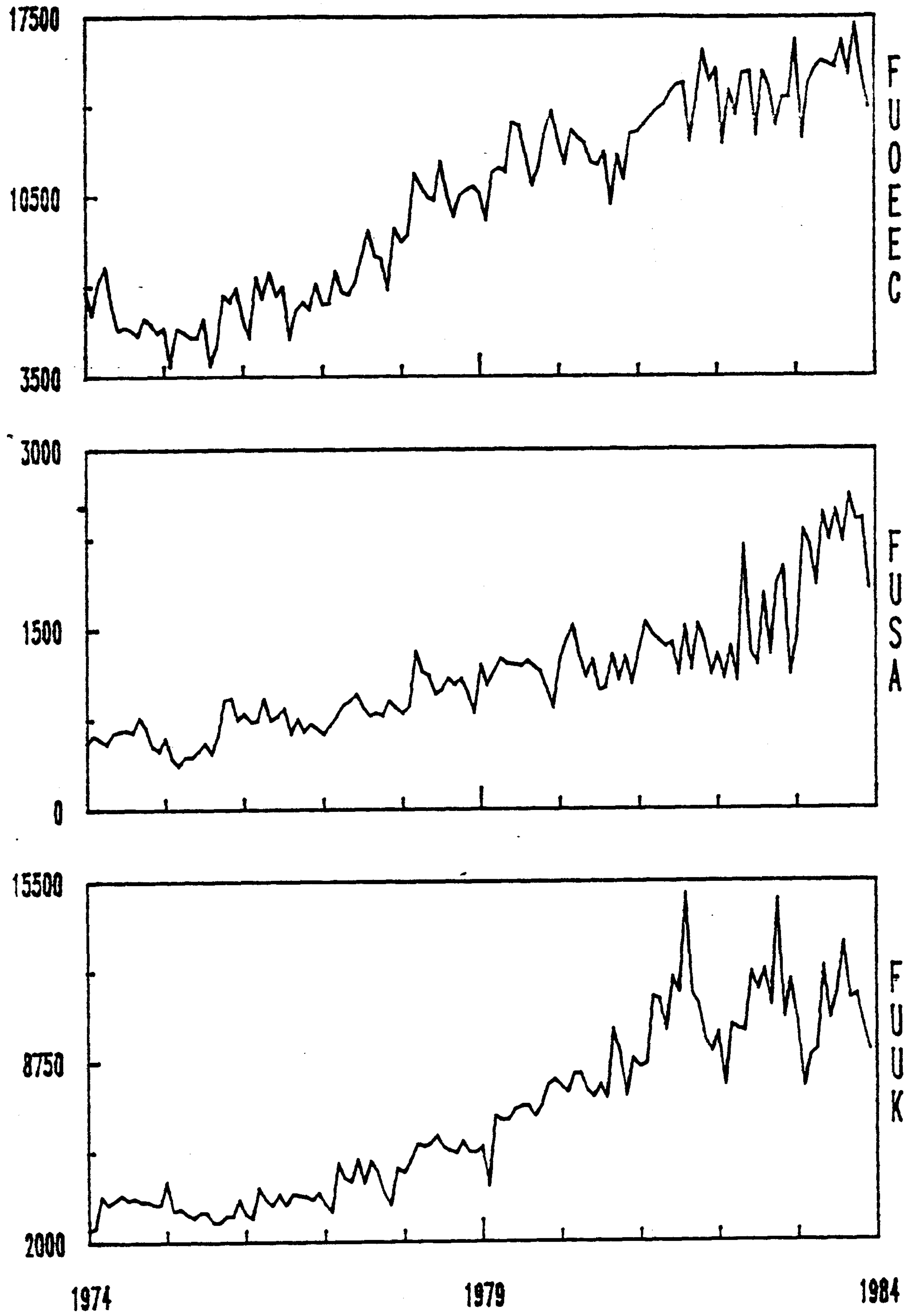

22000
11000
0
XUJAP
12000
6000
0
XUWGER
14000
7000
0
XUUK
1966
1972
1978
1984

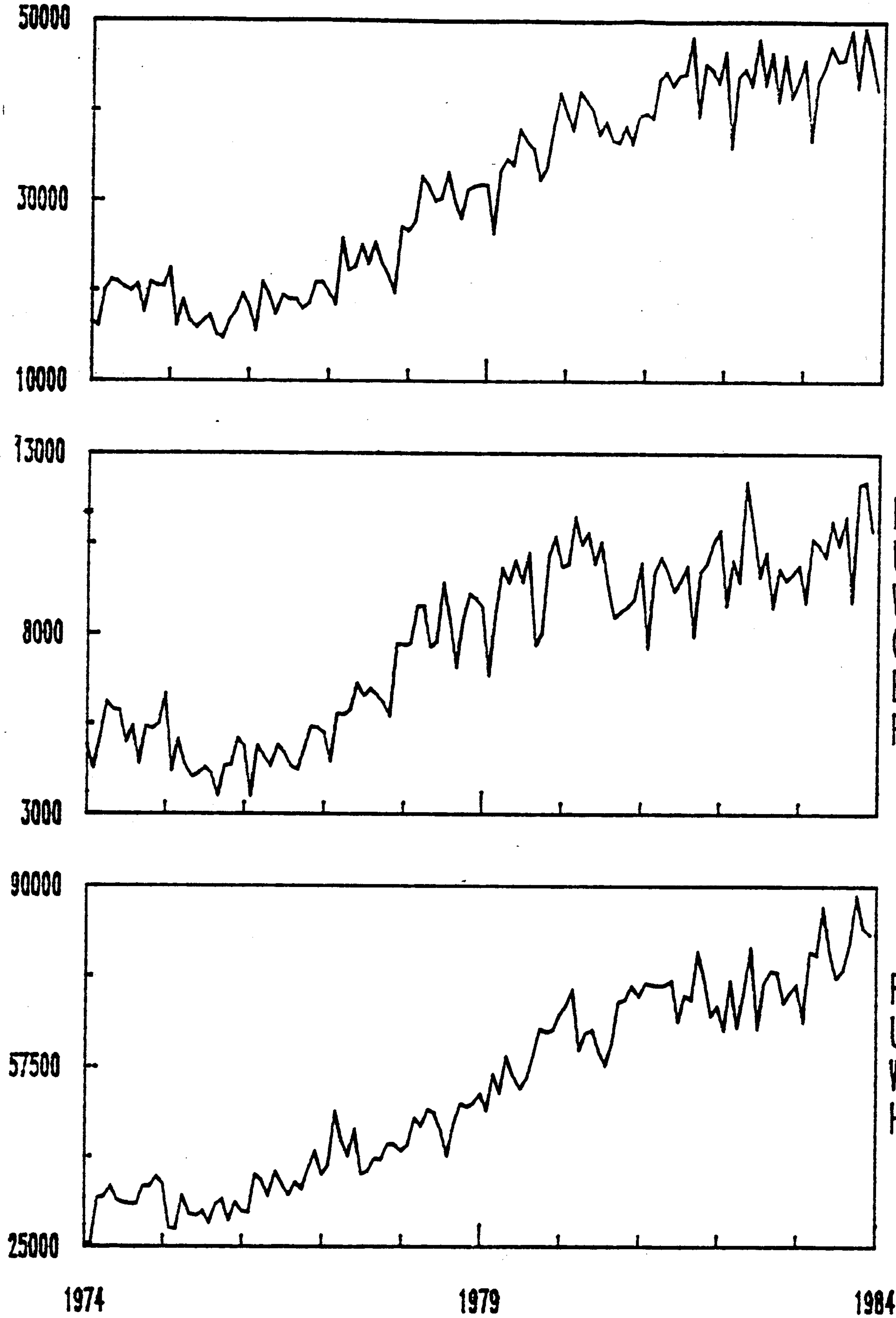

4000
2000
0
XUSASIA
67500
36750
6000
XUWEUR
77500
42250
7000
XUWH
1966
1972
1978
1984


35000
18500
2000
FUAFR
42000
23500
5000
FUANEC
2500
1250
0
FUCACM
1974
1979
1984


2500
1250
0
FUCOME
41000
24250
7500
FUJAP
50000
30000
10000
FUOECD
1974
1979
1984


110000
65000
20000
FUDEVC
37500
20000
2500
FULAR
100000
55000
10000
FUASIA
1970
1977
1984


17500
10500
3500
FUOEEC
3000
1500
0
FUSA
15500
8750
2000
FUUK
1974
1979
1984


50000
30000
10000
FUWEUR
13000
8000
3000
FUWGER
90000
57500
25000
FUWH
1974
1979
1984