88/27

Toward X-12-ARIMA

David F. Findley, Brian Monsell, Mark Otto, William Bell, and Marian Pugh
Bureau of the Census

ABSTRACT

A substantial revision of the Census Bureau's X–11 seasonal adjustment program is underway. The partially completed version of the new program presently being evaluated contains important new diagnostics and the capability of fitting ARIMA models or regressions with ARIMA errors by "nearly" exact maximum likelihood estimation. The "sliding spans" diagnostics represent a particularly valuable addition to the program's diagnostics. This paper illustrates the application of these to questions of seasonal and trading day adjustability, seasonal filter length choice, direct versus indirect adjustment, and seasonal adjustments versus trends. Some of the inadequacies in X–11's F–statistics revealed by these analyses are confirmed by using the minimum AIC procedure to compare regression–with–ARIMA–errors models. Such comparisons, which the program facilitates, also reveal the importance of outlier models. The use of outlier models reveals that the X–11 procedure is less outlier resistant than might have been expected.

KEYWORDS
Regression–with–ARIMA–errors, sliding spans, smoothing, adjustment.

1. SLIDING SPANS DIAGNOSTICS

The sliding spans technique, as it is called, involves the comparison of the "somewhat independent" adjustments of a given month's datum obtained by applying the adjustment procedure to a sequence of three or four overlapping spans of data, all of which contain the month. Seasonally adjusted month–to–month change estimates, year–to–year change estimates, and other related quantities can be examined similarly. Excessive variability among such estimates indicates unreliability. Conversely, if there is no evidence of residual seasonality in the adjusted series, then one can interpret stability, meaning a lack of significant variability, as an indication that the estimates are reliable. This term is not a synonym for "accurate" in an objective sense. (It does not seem possible to give a completely objective definition of seasonality, see Bell and Hillmer (1984).) Our usage of the term represents an extrapolation into a less well–defined situation from our experience that, when a series has an adjustment which is quite stable (and shows no residual seasonality), this adjustment has always turned out to be acceptable by all of the standards with which we are presently familiar. Two adjustments of a series can be reliable, in this sense, and yet different, leaving room for other criteria to be used to make a final choice. In later sections, we will give examples demonstrating that the sliding spans technique provides insights not available from the traditional diagnostics about seasonal and trading day adjustments, seasonal filter length selection, and such questions as whether aggregated series and derived series should be adjusted indirectly or directly.

To obtain sliding spans for a given series, an initial span is selected. (Its length will depend on the seasonal adjustment filters used.) 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 its final year. A third span is obtained from the second in like manner, and, data permitting, also a fourth. Figure 1.1 illustrates three consecutive spans, for X–11 adjustment with 3x5[1] seasonal 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. Usually, there is a strong interest in having the analysis be based on data which are as close to contemporary as possible, which is an incentive to limit the number of spans. The investigations and suggested interpretations described in the remainder of the paper are based on the use of four spans. The use of more spans would tend to increase the range of the observed adjustments and therefore make it necessary to modify our procedure for interpreting the sliding spans statistics. The number and length of the sliding spans chosen for a given series will depend upon the length of the series and on the length of the seasonal adjustment filter chosen by the analyst, as we will explain in section 2 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 1.1, 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 a sense of how reliable the seasonal factor estimate for $X_{1/81}$ is.

To describe how we make such a comparison for multiplicative adjustments, let

[1]We remind the reader that this refers to the simple centered average of three simple centered averages of five consecutive same–calendar–month–values. Such filters are applied to a detrended version of the series (the SI ratios).

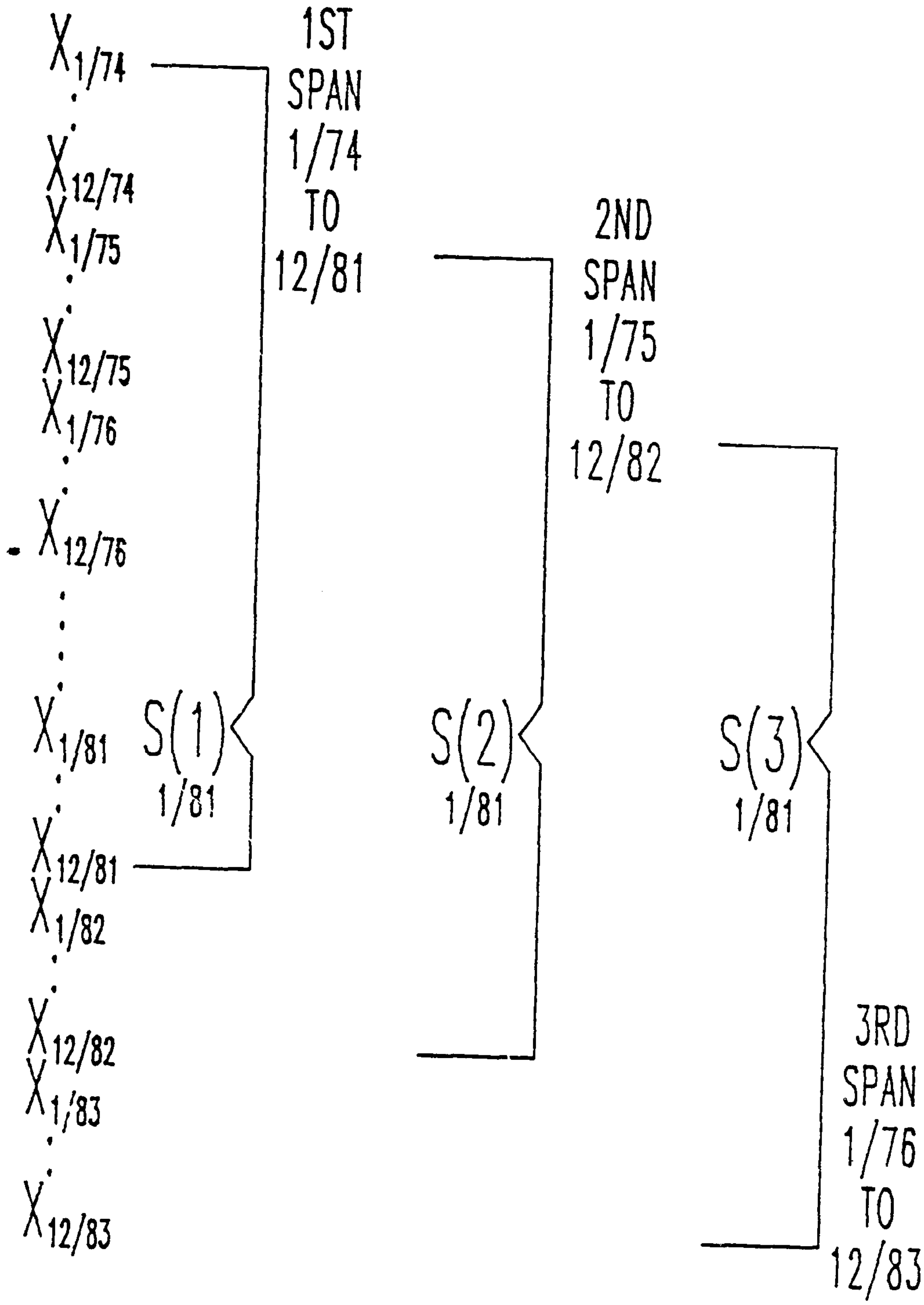


Figure 1.1 Illustration of Sliding Spans

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

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

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

$N1_t$ = {k : months t and t–1 are in the k–th span} .

The Census Bureau's X–11.2 identifies or "flags" (the time series value associated with) month t as having an unreliable seasonal factor if

$$\frac{\mathrm{Max}_{k \epsilon N_t} S_t(k) - \mathrm{Min}_{k \epsilon N_t} S_t(k)}{\mathrm{Min}_{k \epsilon N_t} S_t(k)} > 0.03. \tag{1.1}$$

An estimate of seasonally adjusted percentage change from month t–1 to t is considered unreliable if

$$\mathrm{Max}_{k \epsilon N1_t} \frac{A_t(k) - A_{t-1}(k)}{A_{t-1}(k)} - \mathrm{Min}_{k \epsilon N1_t} \frac{A_t(k) - A_{t-1}(k)}{A_{t-1}(k)} > 0.03. \tag{1.2}$$

Equation (1.1) 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 (1.2) 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 second test, we assess the reliability of the estimate of month–to–month percentage change obtained from the seasonal adjustment method employed. 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 to see if the unreliable adjustments cluster in certain calendar months or years. 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. In the sliding spans output of the Census Bureau's X–11.2 seasonal adjustment program (see the Appendix for an example) summary tables are given in which the months flagged are broken down by year, by month and by magnitude.

Since consumers of seasonally adjusted data frequently compare a month's adjusted value with the adjusted value for the same calendar month a year earlier, we also analyze the stability of seasonal adjustments from this point of view. With

$$N12_t = \{k : \text{months } t \text{ and } t\text{–}12 \text{ are in span } k\},$$

we define the estimation of year–to–year percent change for month t to be unreliable if

$$\mathrm{Max}_{k \epsilon N12_t} \frac{A_t(k) - A_t(k-12)}{A_t(k-12)} - \mathrm{Min}_{k \epsilon N12_t} \frac{A_t(k) - A_t(k-12)}{A_t(k-12)} > 0.03. \tag{1.3}$$

As we will show later by example, difficulties with trading day effect estimation can result in unreliable estimates of year–to–year percent change. In fact, our principal use for (1.3) is to help detect such difficulties.

We sometimes analyze estimated trading day factors in a similar manner. Let

$$TD_t(k) = \text{the trading day factor estimated from span } k \text{ for month } t.$$

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

$$\frac{\mathrm{Max}_{k \epsilon N_t} \; TD_t(k) \; - \; \mathrm{Min}_{k \epsilon N_t} \; TD_t(k)}{\mathrm{Min}_{k \epsilon N_t} \; TD_t(k)} > 0.02 \; . \qquad (1.4)$$

The left–hand side of (1.4) is always 0 for non–leap–year Februaries, so these Februaries are not counted in an analysis.

We have not found (1.4) to be very informative, perhaps due to the effect of the complexity of trading day patterns on estimation and the fact that trading day factors are usually in the range from 97(0.97) to 103(1.03). In our experience, a more useful indication 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 instabilities will usually be reflected in unstable estimates of month–to–month change, and, to a lesser extent, year–to–year change, in the adjusted data. See section 3.

In our investigations, the earliest of which are summarized in Findley and Monsell (1986), we have found that a series which seems to have a good seasonal adjustment according to a variety of criteria, including those of Lothian and Morry (1978), BLS (1977) and the subjective opinions of subject matter experts, usually has fewer than fifteen percent of the adjusted months flagged for erratic seasonal factors. Series with more than twenty–five percent of the months flagged almost never have acceptable adjustments. We found a "gray area" between 15 and 25 percent; a small proportion of the series whose percentage of months flagged falls in this range can probably be adequately adjusted. We recommend that seasonal adjustment not be performed, however, if more than forty percent of the estimates of month–to–month change are flagged.

The threshold 0.03 in (1.1) – (1.3) will be too large if all the seasonal factors are close to 100, but a comparison of the values themselves across the different span will probably be informative in such cases. The adjustor may also decide to change this threshold according to his or her own sense of how much variability can be tolerated in the adjustment, but we caution against raising the threshold value or upper percent limits mentioned above without careful study of the type of series being adjusted. We do not have a recommended upper limit for the acceptable percentage of unstable year–to–year changes. Values around 2% are common with good series; 10% is very high.

In our subsequent discussion, we will refer to the following summary statistics:

S(%)	=	percentage of months with unreliable seasonal factor estimates,
M–M(%)	=	percentage of months with unreliable month–to–month percent change estimates,
Y–Y(%)	=	percentage of months with unreliable year–to–year percentage change estimates
TD(%)	=	percentage of months with unreliable trading day factor estimates

The term "unreliable" is used in the sense discussed above, and "percentage" in each case is relative to the number of candidate months: This is the number of values in $\{t : N_t \text{ is nonempty}\}$ for S(%), $\{t : N1_t \text{ is nonempty}\}$ for M–M(%), etc. With this notation, our recommendations are summarized in Table 1.1.

In the examples discussed in the sections 2–6, the adjustments which are analyzed and compared are all plausible in the weak sense that the seasonally adjusted series do not evidence residual seasonality, according to the F–statistics of table D–11 of X–11.2 and X–11–ARIMA, and to an examination of X–11.2's calculated spectrum of the differenced adjusted series. Only after such tests for residual seasonality have been performed is it appropriate to be concerned with the stability properties measured by the sliding spans statistics.

Table 1.1 Summary of Adjustment Recommendations for Series Whose Maximum and Minimum Seasonal Factors Differ by at Least 10.

S(%) and M–M(%)	ADJUSTABLE?
S(%) ≤ 15.0 ; M–M(%) < 40.0	likely
15.0(%) < S(%) ≤ 25(%) ; M–M(%) < 40.0	less likely
S(%) > 25.0 or M–M(%) > 40.0	unlikely

2. SEASONAL ADJUSTABILITY: SLIDING SPANS VERSUS Q

The X–11–ARIMA program (Dagum, 1983b) helpfully provides a large number of diagnostics to determine if a series is amenable to adjustment by the X–11 method. Among these is a measure, Q, which summarizes the value of eleven diagnostic statistics (M1–M11). Despite cautionary remarks against relying exclusively on this summary measure in the X–11–ARIMA manual and in the literature cited there, our experience is that many users base their decision on whether or not to adjust a series principally on whether or not the value of Q is less than the threshold value 1.0 utilized in X–11–ARIMA. In this section, we first present a sliding spans analysis of a series, S75VS (Value of Shipments of Building Papers), which has a low Q–value (Q = 0.68) suggestive of good adjustability. Then a series with a "failing" value (Q = 1.14) is analyzed, FUOECD (Imports from OECD countries). In both cases our analysis contradicts the suggested interpretation of Q. The Q–statistic used is the modification of X–11–ARIMA's Q described in Findley and Monsell (1986), whose values are usually slightly larger than those of the original. When 3x5 (respectively, 3x3 or 3x9) seasonal filters are used, then four eight–year (respectively six– or eleven–year) spans are used to produce the sliding spans analysis. These span lengths are close to the smallest which can be used to adjust with the associated seasonal filters (a 3x5 filter spans seven years, etc.).

Table 2.1 shows the January, 1974 – February, 1978 section of a month–by–month sliding spans analysis of the estimates of seasonal factors produced by the implementation of sliding spans analysis incorporated into the Census X–11.2 seasonal adjustment program. In addition to the seasonal factors, the maximum percent differences between factors are given, for months common to more than one span, along with symbols flagging months for which the maximum exceeds the 3% threshold selected to identify months with unreliable adjustments. The specific symbols correspond to the levels of excess which define the histogram cells of the breakdown table included in the summary sliding spans analysis by X–11.2, an example of which is given in the Appendix.

Now let us examine the values of these statistics for the two series mentioned above. (See sections 3 and 12 below for further discussion of FUOECD.)

Series	Dates	Q	S(%)	M–M(%)	Y–Y(%)
S75VS	1/75–12/84	0.68	38.9	52.3	5.2
FUOECD	1/74–12/83	1.14	7.3	24.2	0.0

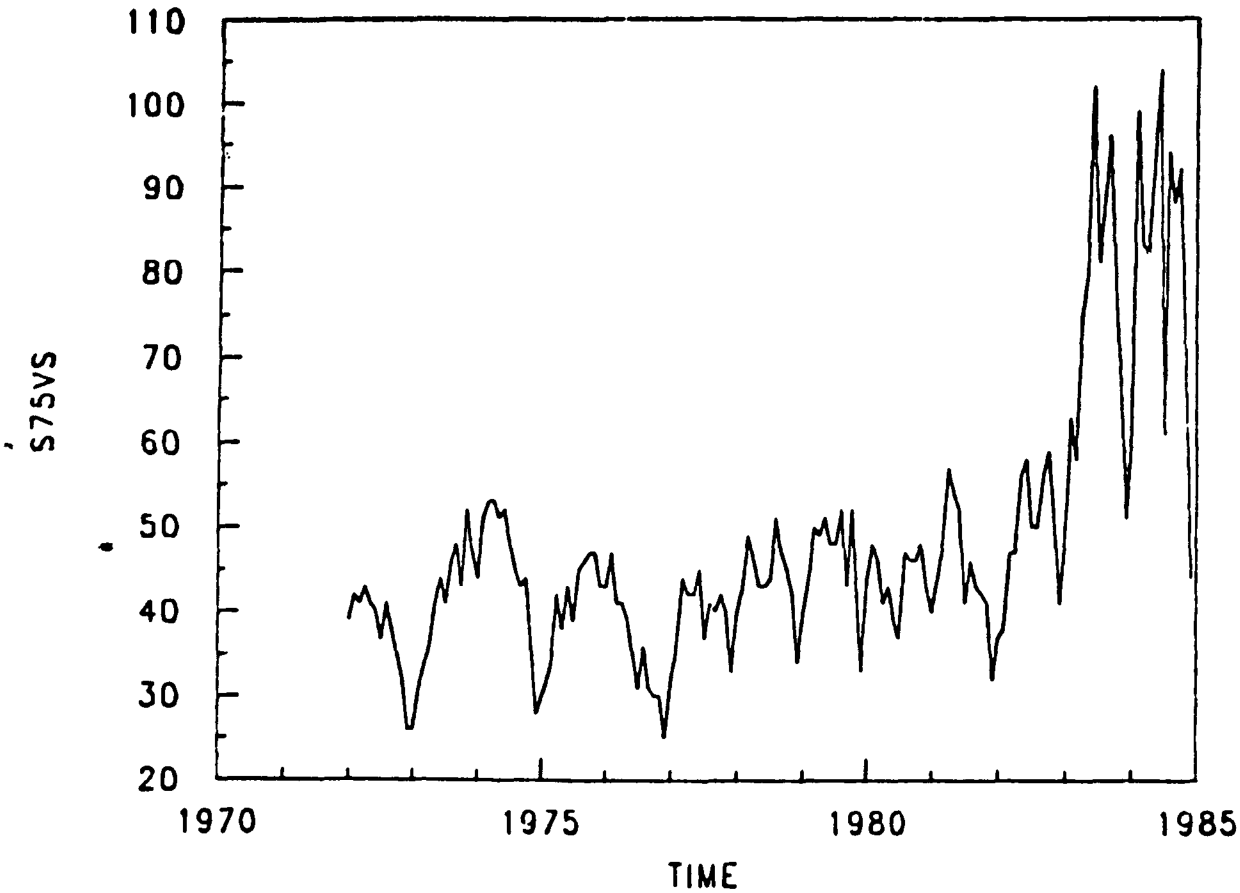
According to the criteria presented for the sliding spans statistics in section 1, the series FUOECD is reliably adjustable, whereas S75VS is not, contradicting the conclusions drawn from the Q values. Which conclusion should one accept?. The <u>sliding spans statistics have the advantage over Q and other traditional statistics of being directly interpretable</u>. Stability of the estimates of the quantities of interest is a fundamental and intelligible property. Not surprisingly, the sliding spans statistics often immediately reveal problem areas in the data. In the case of S75VS, the summary statistics of the sliding spans analysis given in the Appendix show that the more recent data for this series are especially troublesome, which a graph of the series also suggests (see Figure 2.1.) as do the M10 and M11 statistics of X–11–ARIMA, which both have the value 1.4.

Table 2.1

SLIDING SPANS ANALYSIS OF SEASONAL FACTORS FOR S75VS

	1/74 12/81	1/75 12/82	1/76 12/83	1/77 12/84	"SIGN" CHANGE	MAXIMUM % DIFF.	
1-74	91.361	**********	**********	**********		**********	
2-74	101.776	**********	**********	**********		**********	
3-74	107.530	**********	**********	**********		**********	
4-74	106.706	**********	**********	**********		**********	
5-74	104.797	**********	**********	**********		**********	
6-74	106.289	**********	**********	**********		**********	
7-74	95.269	**********	**********	**********		**********	
8-74	106.436	**********	**********	**********		**********	
9-74	101.824	**********	**********	**********		**********	
10-74	100.370	**********	**********	**********		**********	
11-74	97.885	**********	**********	**********		**********	
12-74	79.275	**********	**********	**********		**********	
1-75	91.652	96.666	**********	**********		5.471	%%%
2-75	101.626	103.891	**********	**********		2.229	
3-75	108.455	108.113	**********	**********		0.316	
4-75	106.846	110.070	**********	**********		3.017	%
5-75	104.544	104.176	**********	**********		0.353	
6-75	105.625	103.498	**********	**********		2.055	
7-75	95.364	93.362	**********	**********		2.144	
8-75	106.665	106.159	**********	**********		0.477	
9-75	101.713	99.814	**********	**********	*	1.903	
10-75	100.325	99.732	**********	**********	*	0.595	
11-75	97.447	96.786	**********	**********		0.683	
12-75	79.031	78.213	**********	**********		1.046	
1-76	92.126	96.201	94.760	**********		4.423	%%
2-76	101.299	103.409	100.593	**********		2.799	
3-76	109.690	108.576	114.190	**********		5.171	%%%
4-76	106.958	109.603	108.630	**********		2.473	
5-76	104.578	104.150	106.122	**********		1.893	
6-76	104.458	103.026	100.685	**********		3.747	%
7-76	95.728	94.199	96.277	**********		2.206	
8-76	107.485	106.951	110.095	**********		2.940	
9-76	100.938	99.718	99.209	**********	*	1.743	
10-76	100.286	100.019	99.333	**********	*	0.959	
11-76	96.538	96.514	93.854	**********		2.860	
12-76	78.398	78.012	76.533	**********		2.437	
1-77	93.154	95.773	94.821	91.104		5.125	%%%
2-77	101.848	102.687	100.531	99.346	*	3.363	%
3-77	110.864	109.574	113.762	112.203		3.822	%
4-77	106.487	108.247	108.015	105.481		2.622	
5-77	104.810	104.674	106.398	104.889		1.647	
6-77	102.588	102.006	100.585	101.216		1.991	
7-77	95.940	95.011	96.309	97.461		2.579	
8-77	108.947	108.464	110.474	110.724		2.084	
9-77	100.510	100.005	99.554	102.129	*	2.587	
10-77	100.112	100.428	99.832	102.900	*	3.073	%
11-77	95.912	96.475	94.541	96.181		2.046	
12-77	77.634	77.693	76.529	76.646		1.521	
1-78	94.019	95.087	93.874	90.834		4.682	%%
2-78	101.939	100.987	99.954	99.334	*	2.622	

Figure 2.1 PLOT OF S75VS



Two more series with low Q values but poor sliding spans statistics are discussed in section 4 (HS1F and S95NO directly adjusted); others can be found in section 5.

The high value of Q for FUOECD is caused mainly by M1, which measures the relative contribution of the irregular component, and which has the value 2.5 for this series. We occasionally encounter series in which a large irregular component does not seem to compromise the stability of the adjustments.

3. SLIDING SPANS VERSUS X–11's TRADING DAY REGRESSION F–STATISTIC

Among X–11 and X–11–ARIMA users, the most commonly applied criterion for deciding whether or not to adjust a series for trading day variation is the F–statistic from the program's trading day regression. The "irregular" values to which this regression is applied are the output of a filtering operation, which leads to some problems, see Cleveland and Devlin (1980). One problem is that the "error" terms in the regression are correlated. As a consequence of this and perhaps other factors, too, the distribution of the regression "F"–statistic will differ from the F–distribution, and the program's use of critical values from an F–distribution will sometimes result in misleading conclusions concerning the statistical significance of a trading day component. The sliding spans statistics for the eight series presented in Table 3.1 show that when the trading component of a series is poorly defined, the cost of applying X–11's trading day adjustment, as measured by the percentage of unreliable estimates of seasonally adjusted month–to–month and year–to–year change, can be high. The series analyzed are:

FUANEC	–	Imports from Certain Asian Countries
FUASIA	–	Imports from Asia
FUOECD	–	Imports from OECD Countries
FUOEEC	–	Imports from European Common Market Countries except the United Kingdom and West Germany
FUWEUR	–	Imports from Western Europe
FUWGER	–	Imports from West Germany
FUWH	–	Imports from the Western Hemisphere
XUOECD	–	Exports from OECD Countries.

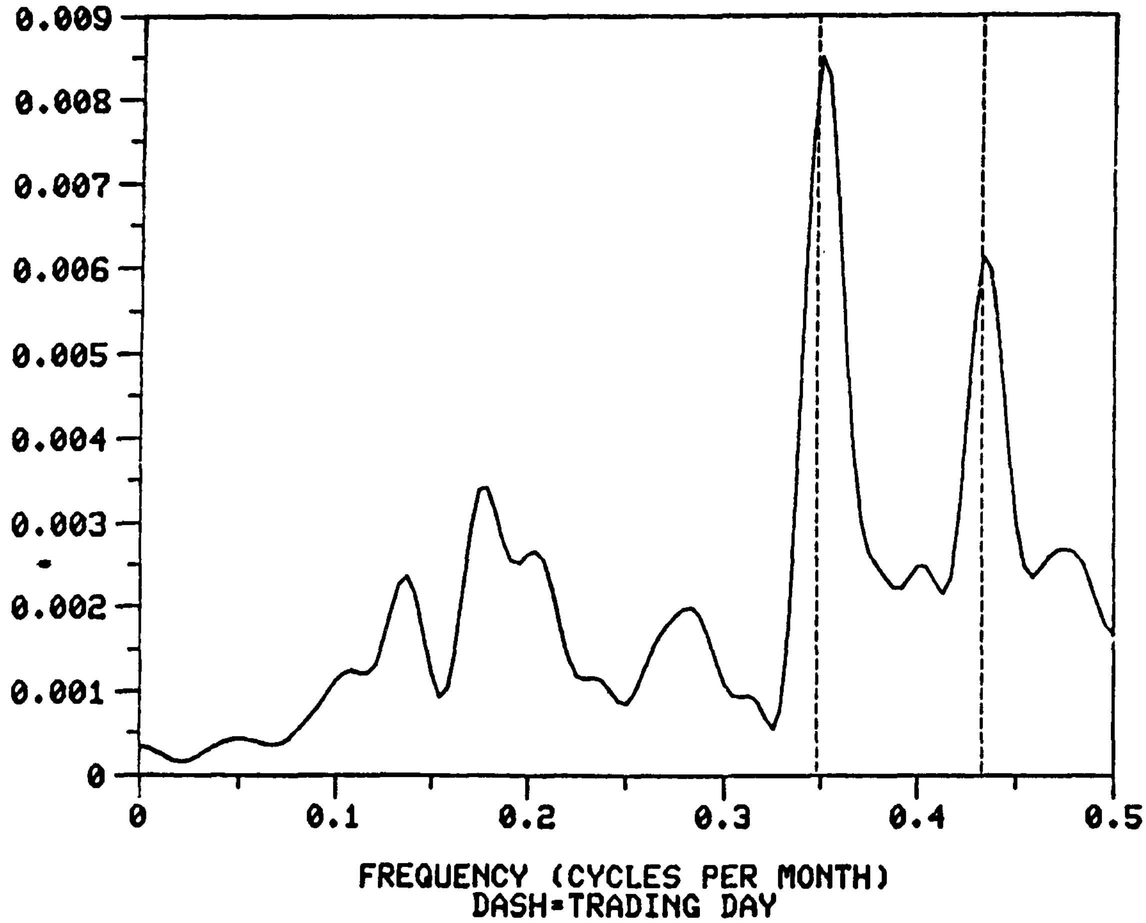
Table 3.1 TRADING DAY ADJUSTMENT/NO ADJUSTMENT: SLIDING SPANS ANALYSIS

Series	S(%)	M–M(%)	Y–Y(%)	TD(%)	F–TD*
FUANEC	22.9/10.4	45.3/29.5	19.0/0.0	0/–	4.7
FUASIA	9.4/11.5	37.9/29.5	21.4/0.0	11.1/–	5.0
FUOECD	12.5/7.3	42.1/24.2	31.0/0.0	31.1/–	6.0
FUOEEC	13.5/9.4	36.8/25.3	17.9/0.0	4.4/–	3.8
FUWEUR	12.5/9.4	35.8/24.2	26.2/0.0	22.2/–	5.4
FUWGER	6.3/0.0	29.5/12.6	25.0/0.0	15.6/–	6.2
FUWH	6.3/2.1	40.0/9.5	38.1/0.0	25.6/–	3.4
XUOECD	8.3/2.8	28.0/10.5	2.5/0.0	0.0/–	5.9

*F–TD: The trading day regression F–statistic from X–11.2 Table C15.
The 5% critical values of the F–distribution with either (6,124) degrees of freedom (for the Imports series) for (6,152) degrees of freedom (for the Exports series) is approximately 2.15.

For the Imports series, despite "significant" F–statistics favoring trading day adjustment, X–11.2's spectrum plots of the outlier–modified irregulars of Table E3, from the run in which no trading day adjustment was performed, did not show peaks at the frequencies associated with trading day effects, see Cleveland and Devlin (1980) for a discussion of spectrum diagnostics. This affirms the conclusion, suggested by the sliding spans

Figure 3.1 SPECTRUM PLOT OF MODIFIED IRREGULAR OF XUOECD



analysis, that these series should not be trading day adjusted. (After this analysis was completed, it was determined that a much higher-than-expected percentage of Customs forms were arriving at the Census Bureau one or more months late and were being assigned, incorrectly, to the month of their arrival. Such errors would mask any trading day effects in the Imports series.)

For the Exports series XUOECD, by contrast, there are well-defined spectral peaks at the trading day frequencies, see Figure 3.1, and the sliding spans statistics for the trading day adjusted series are acceptable, although somewhat worse than are obtained without trading day adjustment.

One source of more reliable alternatives to X-11's F-tests are regression models with ARIMA errors and the associated likelihood-ratio-based model comparison procedures, as we demonstrate in section 12.

4. Direct Versus Indirect Adjustment: Sliding Spans Versus Smoothness

Suppose the series to be seasonally adjusted, X_t, is the sum of component series which are also seasonally adjusted. Then, in addition to the adjusted series obtained by direct adjustment of the X_t, a second seasonally adjusted version of the series can be obtained by summing the adjusted component series. This second approach is called indirect adjustment and will, in general, yield a different series from the one obtained by direct adjustment, because of nonlinearities in the adjustment procedure arising from outlier adjustment and from multiplicative seasonal adjustment. An even larger discrepancy between direct and indirect adjustment is likely when the series to be adjusted is a more complicated function of other series, see subsection 4.2 below.

How does one choose between plausible adjustments of a series obtained from different adjustment procedures? In statistical agencies, the decision concerning which procedure to use is commonly based on one or more properties of the adjusted series which can be expected to be found desirable by a substantial number of data users. For comparing direct and indirect adjustments, the property most often employed is "smoothness", as measured by one or more "smoothness" measures. The smoothness measures calculated by X-11-ARIMA to facilitate the comparison of direct and indirect adjustments, R1 and R2, are defined as follows. If A_t, $1 \leq t \leq N$ is the seasonally adjusted series (direct or indirect) and H_t, $1 \leq t \leq N$ is the associated trend obtained via the Henderson weights, then

$$R1 = (N-1)^{-1} \Sigma_{t=2}^{N} (A_t - A_{t-1})^2$$

and

$$R2 = N^{-1} \Sigma_{t=1}^{N} (A_t - H_t)^2$$

Analogous quantities are also calculated for just the last three years of adjustments. There do not appear to be any theoretical models of seasonality for which the ideal seasonal adjustment minimizes a quantity estimated by R1 or R2, so that the use of such measures to compare adjustments ("smoother" is "better") is unsatisfactory in a fundamental way. It seems considerably more appealing to compare the reliability of the adjustments as measured by directly interpretable sliding spans statistics such as S(%) (calculated from implied seasonal factors $S_t^{ind} = X_t / A_t^{ind}$ for the indirect adjustment), M-M(%) and Y-Y(%). In fact, the smoothness and reliability criteria often agree, as the examples of subsection 4.2 below suggest, but we will begin in subsection 4.1 with an example where indirect adjustments are more reliable but less "smooth" in the sense of R1 and R2. Rather than give individual R-values, we will follow X-11-ARIMA in giving percentage difference values,

$$\Delta = 100 \cdot (R^{direct} - R^{indirect}) / R^{direct} ,$$

so that, according to the traditional use of these statistics, negative values of $\Delta 1$ or $\Delta 2$ favor direct adjustment.

4.1 An Aggregate Series

The series HS1F of total U.S. single family house construction starts is the sum of four regional series associated with totals from the northeastern, north central, southern and western states. Each regional series is seasonally adjusted so that an indirect adjustment of HS1F is available, as well as a direct adjustment. According to the standard diagnostic statistics, the adjustments of the regional series are of good quality. The estimated seasonal patterns differ substantially among the regions, as would be expected from the differences in climate. These facts suggest that seasonality is better removed at the regional level; that is, an indirect adjustment should be more satisfactory, which is the conclusion suggested by the sliding spans diagnostics. However, the smoothness statistics favor direct adjustment rather strongly. The Δ- and sliding spans statistics for the direct and indirect adjustments are given below, along with the Q-statistics calculated by X-11-ARIMA.

HS1F

	Full Series		Last 3 years	
	Δ1	Δ2	Δ1	Δ2
	−43.9	−32.4	−55.1	−68.4

Adjustment	S(%)	M−M(%)	Y−Y(%)	Q
direct	24.5	41.8	0.0	0.27
indirect	13.6	26.0	0.0	0.37

Indirect adjustment of HS1F is also favored by the statistics CONRAT and CPREV of Findley and Monsell (1986), which measure the total absolute revision experienced by each month's seasonal adjustment, from its initial to final seasonal adjustments, and the rate of convergence to the final adjustment, respectively. In our experience, this is strong evidence in favor of indirect adjustment. Moreover, the sliding spans results suggest that direct adjustment is unacceptable. Thus this example reveals the inadequacy of the R– and Q–statistics for determining the choice between direct and indirect adjustment. (Lothian and Morry (1978) warn against such a use of Q–statistics.)

4.2 Derived Series

The various New Orders series published by the M3 Branch of the Census Bureau's Industry Division are not measured directly but are obtained as the sum of the reported Value–of–Shipments series and the monthly change in the reported Unfilled Orders,

$$NO_t = VS_t + (UO_t - UO_{t-1}) ,$$

in a self–explanatory notation. The observed series VS_t and UO_t are seasonally adjusted, so both direct and indirect adjustments of NO_t can be considered.

We present below the Δ–, Q– and sliding spans statistics for three New Orders series,

S13NO – Nonferrous and Other Primary Metals
S67NO – Broadwoven Fabrics and Other Textiles
S95NO – Other Leather Products

	Full Series		Last 3 yrs.	
Series	Δ1	Δ2	Δ1	Δ2
S13NO	−10.1	− 1.1	− 1.4	− 3.5
S67NO	− 5.5	+ 0.4	− 1.5	− 2.1
S95NO	−13.3	−19.1	−16.8	−21.0

Series	Adjustment	S(%)	M−M(%)	Y−Y(%)	Q
S13NO	direct	14.2	32.4	0.0	0.03
	indirect	17.0	36.2	1.1	1.00
S67NO	direct	10.4	22.9	0.0	0.73
	indirect	11.3	29.5	0.0	0.86
S95NO	direct	23.6	33.3	0.0	0.64
	indirect	27.4	44.8	3.2	0.81

Thus, the indirect adjustment of S95NO seems unacceptable and caution seems called for with its direct adjustment. For the other series, direct adjustment seems preferable.

4.3 Raking

Sometimes, for reasons of consistency, the seasonally adjusted component series are modified to force them to have the same annual totals as the direct adjustment or some other adjustment of the aggregate. This is usually done by proportionally redistributing the difference between the indirect and the other adjustments, a procedure known as raking, see Ireland and Kullback (1968) and Fagan and Greenberg (1985) for example. Although we will not give an illustrative example here, it is worth mentioning that we have found sliding spans analysis to be a useful way to assess the effect of these modifications on the quality of the seasonal adjustments of the benchmarked components.

5. ON THE USE OF SLIDING SPANS DIAGNOSTICS TO ASSIST IN THE SELECTION OF SEASONAL FILTERS IN X-11(-ARIMA)

How sensitive is the stability/reliability of an X-11 seasonal adjustment to the choice of seasonal filter? We will now give examples to show that different choices of filters can lead to dramatically different values of the sliding spans statistics. The results for two industrial series from the M3 survey and for the total U.S. imports series, including freight and insurance, will be presented.

EDMISC – Total Consumption of Non-categorized Edible Products,
S49UO – Unfilled Orders of Ophthalmic Goods, Watches and Watch Cases.
CUT – Total U.S. Imports, including Freight and Insurance.

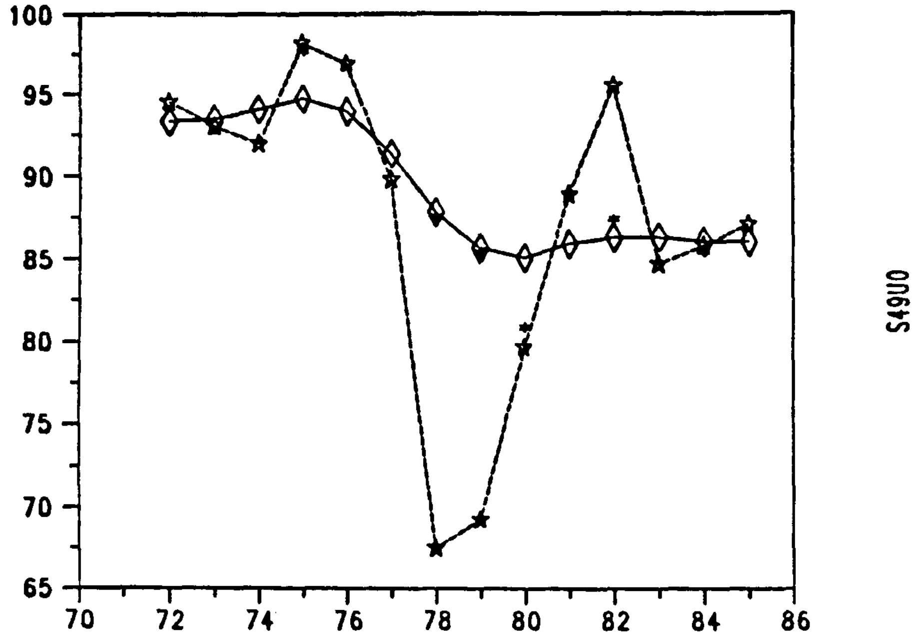
Series	Seasonal Filter	S(%)	M-M(%)	Y-Y(%)
EDMISC	3x5	35.3	42.6	0.0
	3x9	15.1	29.8	0.0
S49UO	3x3	41.7	32.5	8.3
	3x5	37.0	28.1	4.2
	3x9	10.4	9.1	0.0
CUT	3x3	8.5	28.6	10.6
	3x5	11.0	41.9	21.7

These results strongly suggest the use of 3x9 seasonal filters for EDMISC and S49UO. The most carefully analyzed X-11-seasonal filter length selection criterion with which we are familiar is the Global Moving Seasonality Ratio (GMSR, also called I/S-ratio) analyzed by Lothian (1984), which is printed out in Table D9A of X-11.2 and Table F2.H of X-11-ARIMA. For series of the length we are considering, Lothian's recommendation is that 3x3, 3x5 or 3x9 filters can be used, depending on whether the value of GMSR is between 2.3 and 4.1, 4.1 and 5.2, or 5.2 and 6.5. For EDMISC, the value of GMSR is 6.47, favoring the same filter as the sliding spans statistics. For S49UO, GMSR's value is 3.75, so there is disagreement. The graphical methods used in Lothian (1984) in support of GMSR do not clearly favor either filter. Their interpretation is made difficult by the fact that the different filters lead to rather different outlier modification factors which exaggerate the differences between the SI ratios to which the seasonal filters are applied. We refer the reader to Figures 5.1 and 5.2 for the December SI ratios (from X-11 Table D8), their modifications (via D9) and the seasonal factors (from D10), from adjustments with 3x3 and 3x9 filters, respectively, for the series from 1972-1985. December was chosen for display because it is the month whose 3x3 seasonal adjustments are most unstable as measured by the sliding spans statistics: its average value of S(%) is 5.2, and all Decembers were flagged. By contrast, with 3x9 filters only 3 out of 12 Decembers were flagged and the average value of S(%) is 2.5. It seems reasonable to prefer the adjustment with 3x9 seasonal filters, because of its much greater stability.

For CUT, the value of GMSR is 5.1, suggesting the 3x5 filter, but the M-M(%) value of 41.9 for this filter is not satisfactory, whereas the sliding spans statistics associated with the 3x3 filter seem acceptable. Only 13 years of data were available for this series, so only three 11 year spans could be calculated for adjustments with 3x9 filters. Sliding spans statistics from only three spans tend to be smaller than those obtained with four spans, see Table 7.1, making comparisons difficult. However, for this series, the "F"-statistics for the presence of stable seasonality from X-11.2's (and X-11-ARIMA's) Table D8, which are printed for each span by the sliding spans program, offer evidence that shorter than 3x9 filters, and even a shortened series, should be used, as we will now explain. Because the SI ratios in Table D8 are correlated, being detrended values of

Figure 5.1

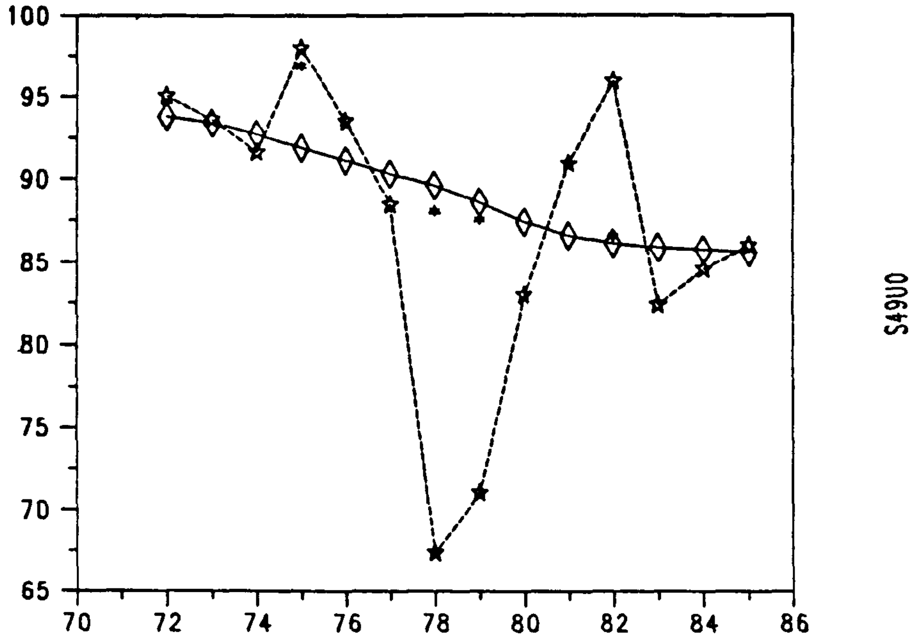
DECEMBER VALUES FOR SI RATIOS AND SEASONAL FACTORS



STAR = SI RATIO, ♦ = MODIFIED SI RATIO
DIAMOND = SEASONAL FACTOR
3 X 3 SEASONAL FILTERS

Figure 5.2

DECEMBER VALUES FOR SI RATIOS AND SEASONAL FACTORS



STAR = SI RATIO, ♦ = MODIFIED SI RATIO
DIAMOND = SEASONAL FACTOR
3 X 9 SEASONAL FILTERS

"F"–Statistics for Stable Seasonality for CUT

		Span			
		1	2	3	4
Filter	3x3	4.1	5.8	6.8	7.2
	3x5	4.3	4.8	6.9	7.0

the observed series, the associated "F"–statistics for the presence of stable seasonality do not follow an F–distribution precisely. The tradition at the Census Bureau and at Statistics Canada has been to interpret an F–statistic value less than 7.0 as indicating a seasonal pattern which is too weak to permit adjustment. By this criterion, only the last 6–8 years of CUT are adjustable.

To summarize, if the seasonal filter length suggested by Lothian's GMSR statistic does not yield an acceptable adjustment, the sliding spans statistics can sometimes reveal a more suitable filter length. Usually (see also Table 7.1 below), but not always, the sliding spans statistics S(%), M–M(%), etc. are smaller for longer filters. The "F"–statistic from Table D8 for the different spans can reveal evolution in the seasonal movements which the other sliding spans statistics do not reflect clearly, and this possibility should also be considered when selecting the length of the seasonal filter.

6. SEASONAL ADJUSTMENTS VERSUS TRENDS (X–11)

"Trends" receive much attention. The concept, however, is use–dependent rather than fixed. For example, analysts seeking short–term trends expect more wiggly graphs than investigators of long–term trends. There are statistical model based long–term forecasting procedures whose trend forecast curves differ from those of the corresponding short–term forecasting procedures in just this way, see Gersch and Kitagawa (1983). The X–11 trends are intended to reveal seasonal movements and are produced by 9–, 13– or 23–term Henderson filters (moving averages), with the 13–term filter being the most frequently used. The "final" trends are obtained by applying one of these filters to the seasonally adjusted data. Because of the filter lengths, the resulting trends are short–term trends and might be expected to have more stable month–to–month changes than the seasonally adjusted data. It is less clear what to expect for year–to–year changes. Surprisingly, perhaps, in every case we have observed, they are less stable. The examples in Table 6.1 are typical. Similar results have been observed with BAYSEA.

Table 6.1 Seasonal Adjustment/Trend Analysis of Some Foreign Trade Series

	S(%)/T(%)*	M–M(%)	Y–Y(%)
FUANEC	10.4/6.2	29.5/6.3	0.0/11.9
FUASIA	11.5/6.2	29.5/1.1	0.0/11.9
FUOECD	7.3/4.2	24.2/0.0	0.0/6.0
FUOEEC	9.4/3.1	25.3/1.1	0.0/4.8
FUWEUR	9.4/2.1	24.2/0.0	0.0/0.0
FUWGER	0.0/0.0	12.6/0.0	0.0/0.0
FUWH	2.1/5.2	9.5/0.0	0.0/9.5
XUOECD	2.8/3.5	10.5/0.0	0.0/8.3

*T(%) is the analogue of S(%) with trend values used instead of seasonal factors.

Seeking theoretical confirmation of these phenomena, through an analysis of the linear filters representing the entire seasonal adjustment and trend estimation procedures of additive X–11, William Bell, in unpublished work, calculated the coefficients of the filters which produce the revisions, from initial to final estimates, of the month–to–month and year–to–year changes, both for seasonal adjustments and for trends. He observed that the filter coefficients associated with month–to–month changes are much larger for seasonal adjustments than for trends, whereas for year–to–year changes, the situation is reversed. Although the coefficient patterns are complex and involve both positive and negative coefficients, this calculation suggests the phenomena we observed. It seems, therefore, that a data user who chooses to use X–11 trends instead of seasonal adjustments must be prepared to accept less stable estimates of year–to–year change.

7. BIAS IN THE SLIDING SPANS PROCEDURE

In a sliding spans analysis with four spans, some months are common to only two spans ("2"–months), some only to three spans ("3"–months), and some occur in all four spans ("4"–months"). Are months for which more comparisons are available more likely to be flagged? Or, to the contrary, are "4"–months the least likely to have unstable adjustments, because their adjustments are produced by closer–to–symmetric filters, see Dagum (1982, 1983a), Wallis (1982) and Pierce and McKenzie (1987)? To address these questions, Marian Pugh investigated the percentages of months identified as having excessively variable factors (S(%)>3.0) for 59 series, using spans of length 7, 8, 11 and 13 years according to whether 3x3, 3x5, 3x9 or "stable" filters were used. (The "stable" filter calculates a single seasonal factor to be applied to all data associated with a given calendar month. This factor is the average (the sample mean) of all the SI ratios associated with the calendar month.) The results presented in Table 7.1 show that, on average, "2"–months will be flagged less than "3"–months, and "3"–months less than "4"–months. The table shows that this conclusion continues to apply when only series are considered which are <u>adjustable</u> for a given filter length, according to the criteria of Table 1.1. These are summary statements, however, and need not apply to an individual series. We have not attempted to refine the threshold values to remove this bias. There are somewhat complicated, practical considerations. For example, "2"–months, which are less numerous and are least likely to be flagged, are either the most recent (and therefore most interesting) or the oldest (and usually therefore least interesting) of the months for which comparisons are available.

Table 7.1
Percentages of Months with Unstable Adjustments,
Classified by Position–In–Span and Filter Length Used.

For each series/filter combination, the count of the number of poorly adjusted months and the ratio to the total number of months is tabulated for months common to 2, 3, and 4 spans, respeetively, and for all months.

SEASONALLY ADJUSTABLE SERIES/FILTER COMBINATIONS

	3x3 FILTER	3x5 FILTER	3x9 FILTER	STABLE	ALL FILTER	
"2"–MONTHS	9/ 672		11/ 768	13/ 936	3/ 984	36/3360
	1%	1%	1%	0%	1%	
"3"–MONTHS	32/ 672		34/ 768	26/ 936	17/ 984	109/3360
	4%	4%	2%	1%	3%	
"4"–MONTHS	42/1008		78/1920	101/3744	171/4428	392/11100
	4%	4%	2%	3%	3%	
ALL MONTHS	83/2352		123/3456	140/5616	191/6396	537/17820
	3%	3%	2%	2%	3%	

ALL SERIES/FILTER LENGTH COMBINATIONS

	3x3 FILTER	3x5 FILTER	3x9 FILTER	STABLE	ALL FILTER	
"2"–MONTHS	273/1464		201/1464	118/1464	76/1464	668/ 5856
	18%	13%	8%	5%	11%	
"3"–MONTHS	420/1464		357/1464	235/1464	187/1464	1199/ 5856
	28%	24%	16%	12%	20%	
"4"–MONTHS	710/2196		879/3660	899/5856	1296/6588	3784/18300
	32%	24%	15%	19%	20%	
ALL MONTHS	1403/5124		1437/6588	1252/8784	1559/9516	5651/30012
	27%	21%	14%	16%	18%	

8. REGRESSION MODELS WITH ARMA ERRORS

Now we turn to the modeling and model estimation capabilities which have been incorporated into our developmental version of X–12–ARIMA. The models are (possibly) nonstationary autoregressive moving average models, with coefficient gaps, which frequently appear as the error process for a linear regression mean function. What follows in Sections 8–11 is taken from Otto, Bell and Burman (1987).

Regression models with autoregressive–moving average (ARMA) errors for equally spaced data can be written,

$$\phi(B)(w_t - \mathbf{x}_t'\beta) = \theta(B)\Theta(B^s)a_t, \quad t=1 \text{ to } n. \tag{8.1}$$

Here w_t is the time series to be modeled, possibly a transformed or differenced version of the original series; $\mathbf{x}_t'$ is a row vector from $\mathbf{X}$, the n x k matrix of similarly differenced regression variables; a_t is an innovation error—assumed to be approximately iid $N(0,\sigma^2)$. $\phi(B)$ is a p–order autoregressive operator,

$$\phi(B)=1-\phi_1B-\phi_2B^2-\phi_3B^3-\cdots-\phi_pB^p,$$

where B is the backshift or lag operator ($B^iy_t=y_{t-i}$). $\phi(B)$ need not include all the lags $1,\cdots,p$ and so can allow for seasonality with operators of the form $1-\phi_{12}B^{12}$ or $1-\phi_1B-\phi_{12}B^{12}$, etc.. We do not consider multiplicative AR operators, $\phi(B)\Phi(B^s)$ because of their nonlinear product coefficient–constraints.

$\theta(B)$ is a q–order non–seasonal moving average operator,

$$\theta(B)=1-\theta_1B-\theta_2B^2-\theta_3B^3-\cdots-\theta_qB^q.$$

$\Theta(B^s)$ is a Q–order seasonal moving average operator with seasonal period s,

$$\Theta(B^s)=1-\Theta_1B^s-\Theta_2B^{2s}-\Theta_3B^{3s}-\cdots-\Theta_QB^{Qs}.$$

The roots of both the non–seasonal and seasonal moving average operators are required to lie on or outside the unit circle and have no common roots with the autoregressive operator. The roots of the autoregressive operator are unconstrained.

Ljung and Box (1976) and Hillmer and Tiao (1979), hereafter HT, document the importance of using exact likelihood methods for the estimation of moving average parameters only. We follow HT's approach of conditioning on the first observations for estimating AR parameters.

9. EXACT MA LIKELIHOOD EVALUATION

In this section we briefly review exact likelihood evaluation for pure MA models. A detailed derivation of the exact form of the likelihood is given by HT and Ljung and Box (1976). For simplicity in this review we consider models with only nonseasonal MA terms, $\theta(B)$, not the full MA operator, $\theta(B)\Theta(B)$. We refer the reader to HT for details on how to handle multiplicative seasonal models efficiently.

The exact form of the likelihood (density of the data) is obtained by relating the data to a set of iid innovations through a linear transformation. Let $\mathbf{a}=(a_1,\cdots,a_n)'$ be the vector of innovations shown in equation (2.1), $\mathbf{a}_*=(a_{1-q},a_{2-q},\cdots,a_{-1},a_0)'$ be q_* initial innovations, assumed to be from the same stochastic process as the a's, so $\mathbf{a}_* \sim N(0,\sigma^2\mathbf{I})$. Let $\mathbf{w}=(w_1,\cdots,w_n)'$ be the data vector of observations satisfying (8.1), and let $\mathbf{w}_*=(w_{1-q},w_{2-q},\cdots,w_{-1},w_0)'$, be q_* artificial initial values prior to the period of observations. As described in Ljung and Box (1976), we can define $\mathbf{w}_*$ by linearly relating it to $\mathbf{a}_*$ by an arbitrary lower triangular system such that $|\mathbf{J}|$, the Jacobian of the transformation between $[\mathbf{a}_*', \mathbf{a}']'$ and $[\mathbf{w}_*', \mathbf{w}']'$ is 1. The transformation from $\mathbf{a}$ to $\mathbf{w}$ and our choice of triangular system will be described below. This transformation allows us to rewrite the exact likelihood in terms of $\mathbf{w}$ and $\mathbf{w}_*$,

$$p(\mathbf{w}_*,\mathbf{w}) = p(\mathbf{a}_*,\mathbf{a})|J| = p(\mathbf{a}_*,\mathbf{a}). \tag{9.1}$$

Now, the joint density, $p(\mathbf{w}_*,\mathbf{w})$, can be factored as

$$p(\mathbf{w})p(\mathbf{w}_*|\mathbf{w}) = p(\mathbf{w}_*,\mathbf{w}). \tag{9.2}$$

We obtain the desired unconditional density, $p(\mathbf{w})$, (the exact likelihood) by obtaining an expression for $p(\mathbf{w}_*,\mathbf{w})$ and identifying $p(\mathbf{w})$ and $p(\mathbf{w}_*|\mathbf{w})$ in (9.2).

The pure MA model is defined as follows:

$$w_t = \theta(B)a_t,\ t=1,\ldots,n,\ .$$

or

$$w_t = -\theta_q a_{t-q} - \theta_{q-1}a_{t-q-1} - \ldots - \theta_1 a_{t-1} + a_t \tag{9.3}$$

Notice that the equations for w_1 to w_q require $\mathbf{a}_*$ so we include q_* more equations for the initial conditions. We use Tunnicliffe–Wilson's (1983) choice of triangular system relating $\mathbf{w}_*$ to $\mathbf{a}_*$, which is

$$\begin{aligned}
w_{1-q} &= a_{1-q} \\
w_{2-q} &= -\theta_1 a_{1-q} + a_{2-q} \\
w_{3-q} &= -\theta_2 a_{1-q} - \theta_1 a_{2-q} + a_{3-q} \\
&\vdots \\
w_0 &= -\theta_{q-1}a_{1-q} - \theta_{q-2}a_{2-q} - \cdots - \theta_1 a_{-1} + a_0
\end{aligned} \tag{9.4}$$

Equations 9.3 and 9.4 in matrix form are,

$$\begin{bmatrix} w_{1-q} \\ \vdots \\ w_0 \\ \hline \vdots \\ w_n \end{bmatrix} = \begin{bmatrix} 1 & & & & & & \\ \vdots & \ddots & & & & & \\ -\theta_q & -\theta_{q-1} & \cdots & -\theta_1 & 1 & & \\ \vdots & \ddots & & & \ddots & \ddots & \\ 0 & \cdots & -\theta_q & \cdots & & -\theta_1 & 1 \end{bmatrix} \begin{bmatrix} a_{1-q} \\ \vdots \\ a_0 \\ \hline \vdots \\ a_n \end{bmatrix}$$

and so in an obvious notation

$$\begin{bmatrix} \mathbf{w}_* \\ \hline \mathbf{w} \end{bmatrix} = A \begin{bmatrix} \mathbf{a}_* \\ \hline \mathbf{a} \end{bmatrix}$$

The innovations and the initial innovations are related to the data and initial values by the inverse transformation,

$$\begin{bmatrix} \mathbf{a}_* \\ \hline \mathbf{a} \end{bmatrix} = A^{-1} \begin{bmatrix} \mathbf{w}_* \\ \hline \mathbf{w} \end{bmatrix} \tag{9.5}$$

where A^{-1} is a matrix of a finite set of π–weights from the expansion $\pi(B) = \sum_{j=0}^{\infty} \pi_j B^j = \theta(B)^{-1}$, where $\pi_0 = 1$ (Tunnicliffe–Wilson 1983)

$$A^{-1} = \begin{bmatrix} 1 & & & \\ \pi_1 & 1 & & \\ \vdots & & \ddots & \\ \pi_{n+q-1} & \cdots & \pi_1 & 1 \end{bmatrix}. \tag{9.6}$$

The π–weights are obtained by equating coefficients in $\theta(B)\pi(B) = 1$.

To separate the equations relating to the initial conditions from those relating to the data, we partition A^{-1} with the first q columns labelled G and the remaining n columns labelled **H**. (9.5) can be rewritten,

$$[G \vdots H]\begin{bmatrix} \mathbf{w}_* \\ \hline \mathbf{w} \end{bmatrix} = \begin{bmatrix} \mathbf{a}_* \\ \hline \mathbf{a} \end{bmatrix}$$

or

$$\mathbf{G}\mathbf{w}_* + \mathbf{H}\mathbf{w} = \begin{bmatrix} \mathbf{a}_* \\ \hline \mathbf{a} \end{bmatrix}$$

If we regress **Hw** on **G** we get an estimate of the initial values given the data, $\hat{\mathbf{w}}_* = -(G'G)^{-1}G'\mathbf{H}\mathbf{w}$. In fact, $\hat{\mathbf{w}}_*$ is the conditional mean and $(G'G)^{-1}\sigma^2$ is the conditional covariance of $\mathbf{w}_*$ given $\mathbf{w}$. Now we can define the unconditional density, p(**w**) (see Ljung and Box 1976),

$$p(\mathbf{w}) = (2\pi\sigma^2)^{-n/2}|G'G|^{-1/2} \times \exp(-(\mathbf{G}\hat{\mathbf{w}}_*+\mathbf{H}\mathbf{w})'(\mathbf{G}\hat{\mathbf{w}}_*+\mathbf{H}\mathbf{w})/2\sigma^2) \quad (9.7)$$

Note in (9.7) that G and H are functions of θ, through the π–weights.

The joint sum of squares, $(\mathbf{G}\hat{\mathbf{w}}_*+\mathbf{H}\mathbf{w})'(\mathbf{G}\hat{\mathbf{w}}_*+\mathbf{H}\mathbf{w})$, is a quadratic form that can be rewritten in terms of **w**, using

$$(\mathbf{G}\hat{\mathbf{w}}_*+\mathbf{H}\mathbf{w}) = \mathbf{H}\mathbf{w} - G(G'G)^{-1}G'\mathbf{H}\mathbf{w} = (I-G(G'G)^{-1}G')\mathbf{H}\mathbf{w}.$$

Let $\mathbf{C} = (I-G(G'G)^{-1}G')\mathbf{H}$ where $I-G(G'G)^{-1}G'$ is idempotent, thus

$$C'C = H'(I-G(G'G)^{-1}G')H$$

and $(C'C)^{-1}\sigma^2$ is the covariance matrix of **w**. Note that **C** is the linear transformation of the data to the exact likelihood residuals, ELR's,

$$\begin{bmatrix} \hat{\mathbf{a}}_* \\ \hline \hat{\mathbf{a}} \end{bmatrix} = \mathbf{C}\mathbf{w}\,. \quad (9.8)$$

Finally, the unconditional density, (9.7), can be written

$$p(\mathbf{w}) = (2\pi\sigma^2)^{-n/2}|G'G|^{-1/2}e^{-\mathbf{w}'C'C\mathbf{w}/2\sigma^2}. \quad (9.9)$$

This is the likelihood, L, that is maximized for exact likelihood estimation of pure MA models. Alternatively, we can minimize the deviance, a monotonically decreasing function of $L = p(\mathbf{w})$ defined as follows. Take the derivative of L with respect to σ^2, set it equal to zero, and solve. The result is

$$\sigma^2 = \mathbf{w}'C'C\mathbf{w}/n.$$

Substitute this back into the likelihood to get

$$L(\theta) = (2\pi(\mathbf{w}'C'C\mathbf{w}/n))^{-n/2}|G'G|^{-1/2}e^{-n/2}.$$

Then to get a transformation of the likelihood that is as free as possible from unnecessary constants, we define the

$$\text{deviance} = (n/2\pi e)L(\theta)^{-2/n} = (\mathbf{w}'C'C\mathbf{w})|G'G|^{1/n},$$

Now, $L(\theta)$ is a monotonically decreasing function of the deviance, so minimizing the deviance will maximize the likelihood.

10. IGLS ESTIMATION OF REGRESSION MODELS WITH ARMA ERRORS

Oberhofer and Kmenta (1974) prove a theorem regarding iterative generalized least squares (IGLS) procedures for obtaining maximum likelihood estimates when direct maximization with respect to all the parameters is difficult. The theorem applies to regression models with ARMA errors and the result shows that jointly maximizing the Gaussian likelihood over β, ϕ, and θ can be done by iteratively maximizing it over β given ϕ and θ and visa–versa.

10.1. Regression Models With AR Errors

Conditional least squares estimation of regression models with autoregressive errors provides a simple example of IGLS estimation. Jointly estimating β and ϕ is a nonlinear problem but estimating each separately is two linear problems. First, let $w_t^f = \phi(B)w_t$, and $x_t^{f\prime} = \phi(B)x_t^{\prime}$, t=p+1, ⋯ ,n, where f denotes AR filtering. The likelihood in terms of β given ϕ is,

$$L(\beta|\phi) = (2\pi\sigma^2)^{-(n-p)/2} e^{-(w^f - X^f\beta)'(w^f - X^f\beta)/2\sigma^2} \tag{10.1}$$

Given ϕ, this is maximized over β by

$$\hat{\beta}^{(i)} = (X^{f(i)\prime}X^{f(i)})^{-1}\, X^{f(i)\prime} w^{f(i)} \tag{10.2}$$

where i indicates the iteration. The likelihood in terms of ϕ given β is

$$L(\phi|\beta) = (2\pi\sigma^2)^{-(n-p)/2} e^{\,-(z-Z\phi)'(z-Z\phi)/2\sigma^2} \tag{10.3}$$

where $z_t = w_t - x_t^{\prime}\beta$ for t = 1, ⋯ ,n, and the columns of Z are the lagged values of $\mathbf{z} = (z_{p+1}, \cdots, z_n)'$. Given β, (10.3) is maximized over ϕ by

$$\hat{\phi}^{(i)} = (Z^{(i)\prime}Z^{(i)})^{-1}\, Z^{(i)\prime} z^{(i)} \tag{10.4}$$

The scheme for maximizing $L(\beta,\phi)$ is to iterate between (10.2) and (10.4). Note that in both cases the regression is calculated easily by doing an OLS regression on the transformed variables. In (10.2), the transformation is the AR filter, and in (10.4) transformation is to the regression residuals. So a difficult nonlinear problem is reduced to a procedure that iterates between two simple linear regressions. Also, since the regression and AR parameter estimates are asymptotically independent (Pierce 1971) their asymptotic variances are obtained from the regressions directly after estimates have converged,

$$\text{var}(\hat{\beta}) = (X^{f\prime}X^f)^{-1}\sigma^2$$

and

$$\text{var}(\hat{\phi}) = (Z'Z)^{-1}\sigma^2.$$

10.2. Regression Models With MA Errors

For regression models with MA errors we modify (9.9), the exact density for pure MA models, to include regression effects,

$$p(w) = (2\pi\sigma^2)^{-n/2}|G'G|^{-1/2} \times \exp(-(w-X\beta)'C'C(w-X\beta)/2\sigma^2) \tag{10.5}$$

For given θ this is maximized over β at

$$\hat{\beta} = (X'C'C\,X)^{-1}X'C'Cw$$

the GLS estimate. At each iteration, a new value of θ is obtained by fixing β, thus updating the inverse covariance matrix $\Sigma^{-1} = C'C\sigma^{-2}$; then a new β is obtained with updated covariance matrix. These iterations continue until convergence.

Note, as was shown in 9.8, that C linearly transforms both $\mathbf{w}$ and X into exact likelihood residuals, ELR's, $\tilde{\mathbf{w}} = C\mathbf{w}$ and $\tilde{X} = CX$. After the transformation to ELR's, the $\hat{\beta}$ can be obtained by an OLS regression of $\tilde{\mathbf{w}}$ on $\tilde{X}$,

$$\hat{\beta} = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{w}.$$

$\tilde{w}$ and $\tilde{X}$ can be obtained recursively as shown in Otto, Bell, and Burman (1987). Finally, assuming the model is correct, the regression estimates are asymptotically independent from the MA parameter estimates, so the q x q covariance matrix of θ is obtained from the inverse of the negative Hessian of the likelihood function (Pierce 1971) and the k x k covariance matrix of β is obtained from the regression,

$$\mathrm{var}(\hat{\beta}) = (\tilde{X}'\tilde{X})^{-1}\sigma^2.$$

10.3. Regression Models With ARMA Errors

We can consider three possible approaches to estimating regression models with ARMA errors: (1), maximize $L(\beta,\phi,\theta)$ jointly over β, ϕ, and θ by nonlinear least squares, (2), use a two stage IGLS where ϕ and θ are estimated jointly by nonlinear least squares and β is estimated by iterative GLS regression, and (3), use a three stage process where only θ is estimated by nonlinear least squares and both β and ϕ are estimated by separate GLS regressions.

The two stage procedure is similar to the regression with MA errors except now both the AR and MA parameters are jointly estimated by nonlinear least squares. By letting the n–p x n matrix, L, represent the linear AR filter,

$$\mathbf{L} = \begin{bmatrix} -\phi_p & -\phi_{p-1} & \cdots & -\phi_1 & 1 & & & \\ & -\phi_p & \cdots & -\phi_2 & -\phi_1 & 1 & & \\ & & \ddots & & & \ddots & \ddots & \\ 0 & & \cdots & -\phi_p & \cdots & & -\phi_1 & 1 \end{bmatrix},$$

the joint likelihood (10.5) can be modified to include AR terms,

$$p(\mathbf{w}) = (2\pi\sigma^2)^{-(n-p)/2}|G'G|^{-1/2} \times \exp(-(\mathbf{w}-X\beta)'L'C'CL(\mathbf{w}-X\beta)/2\sigma^2).$$

The GLS regression for β is

$$\hat{\beta} = (X'L'C'CLX)^{-1}X'L'C'CL\mathbf{w},$$

and the covariance matrix of $\hat{\beta}$ is

$$\mathrm{var}(\hat{\beta}) = (X'L'C'CLX)^{-1}\sigma^2. \qquad (10.6)$$

Notice that these are regression results with the data **w** and regression variables **X** filtered by both the AR filter ($\phi(B)\sim L$) and MA–ELR filter (C).

The three stage method is similar to the regression models with AR errors but at the beginning of each iteration, estimates of the θ's are obtained by nonlinear least squares, and the ELR's are taken as part of the transformation for each GLS. For each iteration's β GLS step the ELR's are taken after the AR filtering and for each iteration's ϕ–GLS step the ELR's are taken after the regression residual transformation. The step–by–step procedure for this method is fairly involved; it is described in Otto, Bell, and Burman (1987).

11. COMMONLY USED REGRESSOR VARIABLES

Our current developmental version of X–12–ARIMA incorporates four kinds of regression variables. These model stable seasonality, trading day variation, additive outliers and level–shift outliers. Other variables, for modeling Easter effects, for example, will be added later. Also, the program will accept as input the design matrix of any linear regression variables the user wishes to include in the model, such as the ramp function defined below.

Stable Seasonality. The conceptual regression model for a stable (perfectly repetitive), additive, seasonal component S_t for monthly data has a coefficient for each calender month, and these twelve coefficients are constrained to sum to zero, leaving eleven unconstrained coefficients. This can be handled with a regression model of the form

$$S_t = \sum_{i=1}^{11} \alpha_i(M_{it} - M_{12t})$$

where

$$M_{it} = \begin{cases} 1, & \text{if month } t \text{ is the i-th calendar month} \\ 0, & \text{otherwise} \end{cases}$$

for i=1,...,12 corresponding to (January,...,December).

<u>Trading Day Effects</u>. A regression model for additive trading day effects (Bell and Hillmer 1983, Bell 1983a) is

$$TD_t = \sum_{i=1}^{6} \beta_i (D_{it} - D_{7t}) + \beta_7 LYF_t$$

where D_{it} is the number of i–th weekdays in month t, i=1,...,7 (Monday,...,Sunday) and

$$LYF_t = \begin{cases} 0.75, & \text{if month } t \text{ is a leap-year February} \\ -0.25, & \text{if month } t \text{ is a non-leap-year February} \\ 0, & \text{if month } t \text{ is not a February.} \end{cases}$$

<u>Outliers</u>. We consider two possible ways in which the observed datum for month t_0 can be an outlier (see Bell 1983b). The additive outlier regression variable is

$$AO_t^{(t_0)} = \begin{cases} 1, & \text{if } t = t_0 \\ 0, & \text{if } t \neq t_0, \end{cases}$$

and the level shift outlier variable is

$$LS_t^{(t_0)} = \begin{cases} 0, & \text{if } t \geq t_0, \\ -1, & \text{if } t < t_0. \end{cases}$$

An additional regression variable which will be considered for the series examined in the next section is the "ramp" function at time t_0 defined by

$$R_t^{(t_0)} = \begin{cases} 0, & t \leq t_0 \\ (t-t_0), & t > t_0. \end{cases}$$

The examination of various combinations of such regressors, perhaps with different ARIMA models for the regression errors, will often give rise to comparisons of models which are non–nested. That is, the models with fewer parameters will frequently not be special cases of the models with more parameters. In this situation, the familiar F–tests and chi–square tests are inapplicable and the familiar "Box–Jenkins" time series methodology lacks an objective procedure for making such comparisons. Findley (1984, 1987) and Findley and Wei (1988) provide a large–sample theory supporting the use of Akaike's minimum AIC criterion (MAIC) for such comparisons. For a given, estimated model, AIC is defined by

$$AIC = (-2) \log \text{maximized likelihood} + 2 \text{ (number of estimated parameters)}.$$

The model with the smallest AIC is preferred.

When Akaike's procedure is applied to compare two nested models, it reduces to a familiar log–likelihood–ratio test with a large–sample chi–square test statistic having degrees of freedom equal to the difference in the number of parameters. In this situation, differences of AIC values less than 1.0 will often be insignificant unless the difference in the number of parameters estimated is large. The difference between two AIC values can have greater variance when non–nested comparisons are made, and differences a little larger than 1.0 may then not indicate a substantial preference for the model with smallest AIC.

12. REGRESSION AS AN AID TO SEASONAL ADJUSTMENT

Our analysis of the series CUT in section 5 suggested that only the later years of data should be considered for seasonal adjustment. We will consider the data from January, 1980 through October, 1987. An "airline" model for the logarithm of the data with appropriate regression terms,

$$(1-B)(1-B^{12})(\log y_t - x_t'\beta) = (1-\theta_1 B)(1-\theta_{12} B^{12})a_t \tag{12.1}$$

seems to fit the data well. The regression terms utilized were the trading day expression TD_t described in section 11; additive outlier variables $AO_t^{(t_0)}$ for times t_0 corresponding to April, 1982, December, 1986 and January, 1987; and a level shift outlier variable $LS_t^{(t_0)}$, with t_0 corresponding to March, 1984, all of which were favored by the MAIC procedure and also had nominal F– and t– statistics which were significant at close to the $\alpha = .01$ level. An examination of the graph of log CUT given as Figure 12.1 suggests that a ramp function $R_t^{(t_0)}$ with t_0 corresponding to March, 1984 might fit better than a level–shift expression. When this was tried, the estimated slope was small ($\hat{\beta}_R$=.0018) and the AIC value of –221. for the model with the ramp was larger than the AIC value of –228. for the model with the level shift, so the model with level–shift was retained. (Note that this comparison concerns non–nested models and that statistical procedures for making such comparisons other than the minimum AIC procedure are not readily available.) Figure 12.2 displays the fitted outlier regression component. Subject–matter experts from the Census Bureau's Foreign Trade Division later told us that the data values for December, 1986 and January, 1987 contained substantial errors (which were corrected subsequently) thus providing an independent confirmation of this part of our model.

Removing from the data the outlier effects estimated with the model above has a substantial impact on both the diagnostic F–statistics and the seasonal factors produced by X–11. For the original data, the Table D8 F–statistic for the hypothesis of no stable seasonality had the value 7.1 (barely significant according to the discussion of section 5) which changed to 9.8 when the series preadjusted for outliers was used. The value of the F–statistic from X–11's Table C15 for the null hypothesis of no trading day effects was 6.7 for the original data and 5.3 for the preadjusted data. As was discussed in section 5, the correlation in the errors of X–11's regressions makes it necessary to have empirical criteria for interpreting the significance of these statistics. If a well–fitting ARIMA–with–regression–model is available for the data, then both the likelihood ratio test (which has a chi–square limiting distribution under the null hypothesis of no effect) and also the likelihood ratio–based minimum AIC procedure represent more objective alternatives to these empirical criteria.

The AIC's for the model (12.1) are –215. without TD_t and –228. with, suggesting the presence of a statistically significant trading day component.

For testing the existence of a significant stable seasonal component, a regression– with–IMA(1,1)–error model was fit,

$$(1-B)(\log y_t - \mathbf{x}_t'\beta) = (1 - \theta_1 B)a_t \ , \tag{12.2}$$

which included a seasonal regression component S_t in addition to the trading day and AO– and LS– outlier effects described above. This is the model that would result if $\theta_{12} = 1$ in (12.1) (Bell 1987). (The estimates of θ_{12} obtained for (12.2) for the different regressions considered were all close to 0.7.) A comparison of the AIC values of the maximum likelihood estimates of (12.2) with (AIC = –280.6) and without (AIC = –238.6) an S_t component, favors the assumption that CUT has such a component.

The Table 12.1 below gives the X–11 seasonal factors first for the original CUT data, then for the outlier preadjusted data, followed by the minimum variance seasonal factors obtained from the Hillmer–Tiao (1982) signal extraction procedure obtained from the fitted version of (12.1). (Hillmer, Bell, and Tiao (1983) discuss application of the procedure to models with calendar and outlier variables.) The latter factors are similar to those obtained by X–11 from the preadjusted data. They differ markedly in December and January from those obtained by X–11 from the original data, revealing a lack of robustness in X–11's procedure. (We have encountered a quarterly series in which three outliers strongly influenced the seasonal factors of three of the four quarters over a five year stretch.) Scott (1987) has made similar observations and has also identified certain instabilities associated with identifying and estimating outliers.

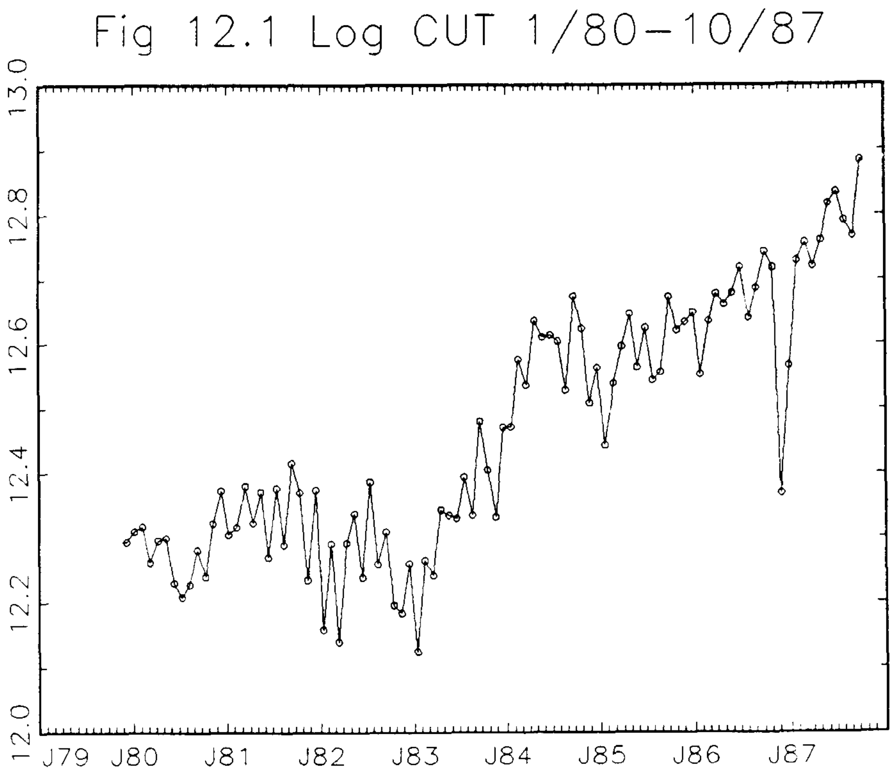


Fig 12.1 Log CUT 1/80-10/87
13.0
12.8
12.6
12.4
12.2
12.0
J79 J80 J81 J82 J83 J84 J85 J86 J87

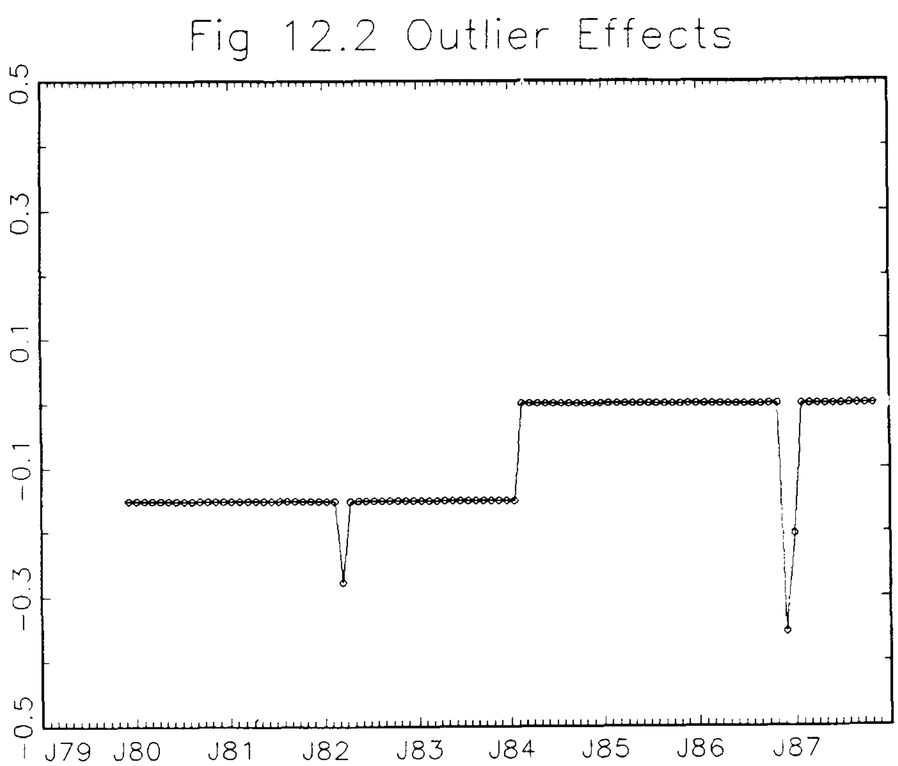


Fig 12.2 Outlier Effects
0.5
0.3
0.1
-0.1
-0.3
-0.5
J79 J80 J81 J82 J83 J84 J85 J86 J87

Table 12.1. Effects of Outliers on Seasonal Factor Estimates

The first panel presents the seasonal factors obtained from X–11 for CUT. The second was obtained by X–11 from the preadjusted series from which the outlier effects identified by the estimated model had been removed. The third panel displays the seasonal factor obtained from the Hillmer–Tiao model–based adjustment procedure applied to the outlier preadjusted series.

FINAL SEASONAL FACTORS FOR IMPORT SERIES

3X5 MOVING AVERAGE SELECTED.

YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
1980	102.01	93.57	100.13	98.72	102.70	103.14	98.27	102.41	96.87	107.94	100.81	93.64
1981	101.71	93.21	100.44	98.47	103.04	103.11	98.81	102.27	96.98	107.85	100.91	93.41
1982	101.40	92.31	100.69	98.47	103.69	103.14	99.58	102.01	97.04	107.97	101.09	92.92
1983	100.80	91.42	100.87	98.61	104.32	103.25	100.74	101.53	97.30	107.77	101.11	92.45
1984	100.16	90.64	101.14	98.98	104.87	103.07	102.16	100.91	97.34	107.80	100.97	91.91
1985	99.40	90.54	101.56	99.23	105.17	102.89	103.36	100.29	97.36	107.43	100.88	91.70
1986	99.00	90.63	101.84	99.62	105.25	102.59	103.93	99.87	97.12	107.43	101.07	91.66
1987	98.87	90.68	101.74	99.94	105.19	102.55	104.06	99.71	97.07	107.41	********	********

FINAL SEASONAL FACTORS FOR SERIES MODIFIED FOR OUTLIERS

3X5 MOVING AVERAGE SELECTED.

YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
1980	105.15	94.34	98.76	98.96	101.88	103.04	97.55	101.72	97.27	107.70	98.78	95.13
1981	104.84	93.97	98.91	98.91	102.12	102.91	97.99	101.72	97.28	107.80	98.92	94.99
1982	104.43	93.15	99.02	99.17	102.58	102.71	98.63	101.60	97.13	108.10	99.22	94.84
1983	103.80	92.25	99.10	99.57	103.05	102.43	99.76	101.05	97.07	107.86	99.46	95.09
1984	103.16	91.65	99.33	99.85	103.46	101.99	101.13	100.20	96.68	107.53	99.95	95.35
1985	102.65	91.69	99.46	100.01	103.63	101.57	102.41	99.25	96.33	106.81	100.34	95.82
1986	102.50	92.05	99.47	100.16	103.64	101.25	103.10	98.58	95.85	106.50	100.77	96.05
1987	102.54	92.24	99.34	100.39	103.57	101.16	103.37	98.23	95.72	106.30	********	********

HILMER-TIAO CANONICAL SEASONAL FACTORS

YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
1980	104.82	95.36	100.28	99.52	102.94	102.12	98.23	101.11	95.45	106.02	99.45	95.58
1981	104.78	95.09	99.94	99.40	102.86	102.19	98.54	101.55	95.79	106.49	99.56	95.35
1982	104.42	94.55	99.58	98.99	102.85	102.25	99.18	101.79	96.11	106.82	99.69	95.24
1983	103.88	94.26	99.48	98.56	102.97	102.10	99.83	101.41	96.16	106.93	100.19	95.43
1984	103.51	94.38	99.47	98.37	103.13	101.84	100.39	100.76	96.01	106.79	100.67	95.68
1985	103.23	94.62	99.38	98.40	103.13	101.51	100.83	100.10	95.84	106.57	100.93	96.04
1986	103.26	95.10	99.33	98.39	102.91	101.25	101.20	99.65	95.71	106.39	100.97	96.23
1987	103.40	95.55	99.33	98.25	102.72	101.14	101.41	99.44	95.62	106.34		

For a final application of the use of these modeling and model comparison procedures, we return to the discussion of section 3 concerning the question of trading day adjusting some foreign trade series. The AIC values associated with the trading day and outlier regression models with seasonal ARIMA errors given in Table 12.2 below affirm the conclusions of section 3: Only for the export series XUOECD is there some justification for doing a trading day adjustment.

Table 12.2. AIC ANALYSIS FOR TRADING ADJUSTMENT OF FOREIGN TRADE SERIES

		Outliers		AIC's
Series	Model Type	# AO's	#LS's	TD/No TD
FUANEC	$(0\ 1\ 1)(1\ 0\ 1)_{12}$	2	2	–188.3/–192.4
FUASIA	$(0\ 1\ 2)(1\ 0\ 1)_{12}$	0	2	–202.5/–210.2
FUOECD	$(0\ 1\ 1)(0\ 1\ 1)_{12}$	1	1	–230.9/–240.0
FUOEEC	$(0\ 1\ 1)(1\ 0\ 1)_{12}$	1	0	–173.4/–175.1
FUWEUR	$(0\ 1\ 1)(0\ 1\ 1)_{12}$	2	0	–232.1/–242.3
FUWGER	S_t+constant+ (0 1 1) error	2	0	–227.7/–237.9
FUWH	$(0\ 1\ 1)(0\ 1\ 1)_{12}$	0	1	–259.9/–261.8
XUOECD	S_t+constant+ (0,1,1) error	4	2	–560.8/–539.5

13. OTHER POSSIBLE ENHANCEMENTS FOR X–12–ARIMA

Time Series Diagnostics

We plan to add some ARIMA modeling and outlier diagnostics, so that the user has the option of utilizing X–12–ARIMA for model and outlier identification purposes. Bell (1983) describes an outlier identification program which is a prototype of what we are currently using. We are also seeking diagnostics to help the user select the forecast horizon for extending the series prior to seasonal adjustment, when a regression or seasonal ARIMA model has been fitted.

Seasonal Adjustment Diagnostics

Some of X–11's diagnostics, such as the "F"–statistics and the F2B table, are based on sums of squares, which can be strongly influenced by outliers. Several approaches are being considered to robustify these diagnostics in ways that do not substantially affect the program's execution time. The simplest approach, currently under investigation, is to replace the irregular component of the series with the modified irregular from Table E3. The graphical diagnostics of X–11.2, including spectra, have been incorporated into the program.

Filters

With the exception of the Henderson trend filter, the symmetric filters of X–11 have an attractively simple form. This is less true of the asymmetric filters used for the adjustment of the most recent observations, which are usually the data most of interest. It may be worth abandoning simplicity altogether in favor of filters with good frequency response function characteristics designed by the techniques in current use in electrical engineering. Another alternative, when a seasonal ARIMA model has been fitted to the series, is to use the filters determined by a model–based signal extraction procedure.

User Interface

The Namelist input methodology of X–11.2 will be used to reduce the amount of information required from the user and to facilitate interactive processing.

14. SUMMARY

The seasonal adjustment program under development at the Census Bureau combines features of the new X–11.2 version of the Census Bureau Method II program with enhancements to Statistics Canada's X–11–ARIMA methodology and diagnostics. With reference to X–11 and X–11–ARIMA, its most important new features are the sliding spans diagnostics and the possibility of using regression models with seasonal ARIMA errors. The sliding spans diagnostics offer more insight than previous diagnostics into the reliability of the adjustments produced by the programs many options. The new regression option makes it possible to address a variety of special characterisitcs of an individual series which compromise the performance of the X–11.2 and X–11–ARIMA programs. The program is being developed on an IBM PC/AT with math coprocessor and is intended to be portable enough to run on any computer for which a sufficiently modern FORTRAN is available (FORTRAN77 + Namelist) and which has enough memory (presumably 640 KB).

REFERENCES

BELL, WILLIAM R. (1983a). "Seasonal Decomposition of Deterministic Effects," Research Report No. 84/01, Statistical Research Division, Bureau of the Census, Wasington, D.C.

BELL, WILLIAM R. (1983b). "A Computer Program for Detecting Outliers in Time Series." Proceedings of the Business and Economics Section of the American Statistical Association, 634–639.

BELL, WILLIAM R. and HILLMER, STEVEN C. (1983). "Modeling Time Series With Calendar Variation," Journal of the American Statistical Association, 78, 526–534.

BELL, WILLIAM R. and HILLMER, STEVEN C. (1984). "Issues Involved with the Seasonal Adjustment of Time Series." JBES 2, 291–320.

BELL, WILLIAM R. (1987). "A Note on Overdifferencing and the Equivalence of Seasonal Time Series Models With Monthly Means and Models With $(0,1,1)_{12}$ Seasonal Par When $\theta = 1$," Journal of Business and Economic Statistics, 5, 383–387.

BLS (1977). "Criteria for Judging the Adequacy of a Seasonal Adjustment." Bureau of Labor Statistics Technical Report (Committee on Seasonal Adjustment Methods).

CLEVELAND, WILLIAM and DEVLIN, S. J. (1980). "Calendar Effects in Monthly Time Series: Detection by Spectrum Analysis and Graphical Methods." JASA 75, 487–496.

DAGUM, ESTELLA (1982). "The Effects of Asymmetric Filters on Seasonal Factor Revisions," JASA 77, 732–738.

DAGUM, ESTELLA B. (1983b). The X–11–ARIMA Seasonal Adjustment Method. Ottawa: Canada.

DAGUM, ESTELLA B. (1983a). "Spectral Properties of the Concurrent and Forecasting Seasonal Linear Filters of the X–11–ARIMA Method." The Canadian Journal of Statistics, 11, 73–90.

FAGAN, JAMES and GREENBERG, BRIAN (1985). Algorithms for Making Tables Additive: Raking, Maximum Likelihood and Minimum Chi–Square. Research Report No. 85/12, Statistical Research Division, Bureau of the Census, Washington, D.C.

FINDLEY, DAVID F. (1984). "On the Unbiasedness Property of AIC for Exact or Approximating Linear Stochastic Time Series Models," J. Time Ser. Anal. 6, 229–252.

FINDLEY, DAVID F. and MONSELL, BRIAN C. (1986). New Techniques for Determining If a Time Series Can Be Seasonally Adjusted Reliably, and Their Application to U.S.Foreign Trade Series. Regional Econometric Modeling. (Eds. M. R. Perryman and J. R. Schmidt) Amsterdam: Kluwer–Nijhoff, 195–228.

FINDLEY, DAVID F. (1987). "An Analysis of AIC and the Entropy Maximization Principle for Comparing Exact or Approximate Models for Random Samples Including Non–Nested and Conditional Models." submitted for publication.

FINDLEY, DAVID F. and WEI, CHING ZONG (1988). "Results on Kullback–Leibler Entropy, AIC and the Comparison of Not Necessarily Nested Time Series Regression Models," in preparation.

HILLMER, STEVEN C., BELL, WILLIAM R., and TIAO, GEORGE C. (1983). "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.

HILLMER, STEVEN C. and TIAO, GEORGE C. (1982). "An ARIMA Model–Based Approach to Seasonal Adjustment," Journal of the American Statistical Association, 77, 63–70.

HILLMER, STEVEN H. and TIAO, GEORGE C. (1979). "Likelihood Function of Stationary Multiple Autoregressive Moving Average Models," Journal of the American Statistical Association, Vol. 74, No. 367, 652–660.

IRELAND, C. K. and KULLBACK, S. (1968). "Contingency Tables with Given Marginals." Biometrika 55, 1968.

LOTHIAN, J. (1984). "The Identification and Treatment of Moving Seasonality in the X–11–ARIMA Seasonal Adjustment Method." Research paper. Business Finance Division, Statistics Canada.

LOTHIAN, J. and MORRY, M. (1978). "A Set of Quality Control Statistics for the X–11–ARIMA Seasonal Adjustment Program." Research paper. Seasonal Adjustment and Time Series Staff, Statistics Canada.

LJUNG, GRETA M. and BOX, G.E.P. (1976). "Studies in the Modelling of Discrete Time Series," Technical Report No. 476, Department of Statistics, University of Wisconsin–Madison.

OBERHOFER, W. and KMENTA, J. (1974). "A General Procedure for Obtaining Maximum Likelihood Estimates in Generalizing Regression Models," Econometrica, Vol. 42, No. 3, 579–590.

OTTO, MARK C., BELL, WILLIAM R., and BURMAN, J. Peter (1987). "An Iterative GLS Approach to Maximum Likelihood Estimation of Regression Models with ARIMA Errors," Research Report No. 87/34, Statistical Research Division, Bureau of the Census, Washington, D.C.

PIERCE, DAVID A. (1971). "Least Squares Estimation in the Regression Model with Autoregressive–Moving Average Errors," Biometrika, Vol. 58, 299–312.

PIERCE, DAVID A. and McKENZIE, SANDRA K. (1987). "On Concurrent Seasonal Adjustment." JASA 82, 720–732.

SCOTT, STUART (1987). "On the Impact of Outliers on Seasonal Adjustment," Proceedings of the Business and Economics Section of the American Statistical Association.

TUNNICLIFFE–WILSON, G. (1983). "The Estimation for Time Series Models Part I. Yet Another Algorithm for the Exact Likelihood of ARIMA Models," Technical Report No. 2528, Mathematics Research Center, University of Wisconsin–Madison.

WALLIS, K. F. (1982). "Seasonal Adjustment and Revision of Current Data: Linear Filters for the X–11–Method," JRSS Ser. A 145, 74–85.

SLIDING SPANS ANALYSIS

S 0. SUMMARY OF OPTIONS SELECTED FOR THIS RUN

PRINT ONLY BREAKDOWN TABLES
TRADING DAY FACTORS NOT ANALYZED
YEAR-TO-YEAR CHANGES ANALYZED
NUMBER OF SPANS : 3
LENGTH OF SPANS : 96
MONTH OF FIRST OBSERVATION IN FIRST SPAN : 1
YEAR OF FIRST OBSERVATION IN FIRST SPAN : 74
NAME OF SERIES BEING ADJUSTED : FUOECD

SUMMARY OF F-TESTS FOR STABLE AND RESIDUAL SEASONALITY FOR EACH SPAN

	SPAN 1	SPAN 2	SPAN 3
STABLE SEASONALITY (D 8)	8.16 ***	7.66 ***	7.54 ***
RESIDUAL SEASONALITY (D11) FULL SERIES	0.51	0.55	0.48
RESIDUAL SEASONALITY (D11) LAST THREE YEARS ONLY	0.31	0.61	0.73
RESIDUAL SEASONALITY (E 2) FULL SERIES	0.17	0.20	0.14
RESIDUAL SEASONALITY (E 2) LAST THREE YEARS ONLY	0.31	0.67	1.18

*** : F-TEST SIGNIFICANT AT THE 0.1 PERCENT LEVEL
** : F-TEST SIGNIFICANT AT THE 1.0 PERCENT LEVEL
* : F-TEST SIGNIFICANT AT THE 5.0 PERCENT LEVEL

NOTE: SUDDEN LARGE CHANGES IN THE LEVEL OF THE SEASONALLY ADJUSTED SERIES WILL INVALIDATE THE RESULTS OF THE RESIDUAL SEASONALITY F-TEST FOR THE LAST THREE YEAR PERIOD.

S 1.A BREAKDOWN OF 3.0 PERCENT OR MORE DIFFERENCES IN SEASONAL FACTORS FOR FUOECD

TOTAL : 9 OUT OF 96 (9.4 %)
(USUALLY, 15% IS TOO HIGH, AND 25% IS MUCH TOO HIGH)

JANUARY	:	0	(AMPD = 1.4)
FEBRUARY	:	4	(AMPD = 2.8)
MARCH	:	2	(AMPD = 2.1)
APRIL	:	0	(AMPD = 1.1)
MAY	:	0	(AMPD = 1.1)
JUNE	:	0	(AMPD = 1.6)
JULY	:	0	(AMPD = 0.9)
AUGUST	:	1	(AMPD = 2.2)
SEPTEMBER	:	0	(AMPD = 0.7)
OCTOBER	:	2	(AMPD = 2.5)
NOVEMBER	:	0	(AMPD = 0.9)
DECEMBER	:	0	(AMPD = 1.6)

1975	:	2	(AMPD = 1.6)
1976	:	1	(AMPD = 2.0)
1977	:	0	(AMPD = 1.4)
1978	:	0	(AMPD = 0.8)
1979	:	1	(AMPD = 0.8)
1980	:	1	(AMPD = 1.6)
1981	:	3	(AMPD = 2.5)
1982	:	1	(AMPD = 1.8)

AMPD = AVERAGE MAXIMUM PERCENTAGE DIFFERENCE

7 OUT OF 96 (7.3 %) OF THE MONTH(S) TESTED HAD A CHANGE OF DIRECTION.
2 OUT OF 7 (28.6 %) OF THE MONTH(S) WITH CHANGES OF DIRECTION WERE FLAGGED.

S.1.B HISTOGRAM OF MAXIMUM PERCENTAGE DIFFERENCE IN SEASONAL FACTORS FOR FLAGGED MONTHS

%	: GREATER THAN OR EQUAL TO 3.0% BUT LESS THAN 4.0% :	5
%%	: GREATER THAN OR EQUAL TO 4.0% BUT LESS THAN 5.0% :	3
%%%	: GREATER THAN OR EQUAL TO 5.0% BUT LESS THAN 6.0% :	1
%%%%	: GREATER THAN OR EQUAL TO 6.0% :	0

S.1.C STATISTICS FOR MAXIMUM PERCENTAGE DIFFERENCE OF THE SEASONAL FACTORS

MINIMUM	:	0.020
25TH PERCENTILE	:	0.773
MEDIAN	:	1.345
->65TH PERCENTILE	:	1.780<-
75TH PERCENTILE	:	2.128
->85TH PERCENTILE	:	2.829<-
MAXIMUM	:	5.804

S 4.A BREAKDOWN OF 3.0 PERCENT OR MORE DIFFERENCES IN MONTH-TO-MONTH CHANGES IN S. A. DATA FOR FUOECD

TOTAL : 22 OUT OF 95 (23.2 %)
(USUALLY, 35% IS TOO HIGH, AND 40% IS MUCH TOO HIGH)

JANUARY	:	2	(AMPD = 2.4)
FEBRUARY	:	3	(AMPD = 3.5)
MARCH	:	5	(AMPD = 3.0)
APRIL	:	1	(AMPD = 1.5)
MAY	:	2	(AMPD = 1.4)
JUNE	:	1	(AMPD = 1.6)
JULY	:	1	(AMPD = 1.6)
AUGUST	:	3	(AMPD = 3.1)
SEPTEMBER	:	0	(AMPD = 1.7)
OCTOBER	:	2	(AMPD = 2.2)
NOVEMBER	:	1	(AMPD = 1.7)
DECEMBER	:	1	(AMPD = 1.4)

1975	:	4	(AMPD = 1.9)
1976	:	1	(AMPD = 1.9)
1977	:	1	(AMPD = 1.4)
1978	:	0	(AMPD = 1.1)
1979	:	1	(AMPD = 1.5)
1980	:	3	(AMPD = 2.5)
1981	:	7	(AMPD = 3.7)
1982	:	5	(AMPD = 2.7)

AMPD = AVERAGE MAXIMUM PERCENTAGE DIFFERENCE

11 OUT OF 95 (11.6 %) OF THE MONTH(S) TESTED HAD A CHANGE OF DIRECTION.
4 OUT OF 11 (36.4 %) OF THE MONTH(S) WITH CHANGES OF DIRECTION WERE FLAGGED.

S.4.B HISTOGRAM OF MAXIMUM PERCENTAGE DIFFERENCE IN MONTH-TO-MONTH CHANGES IN S. A. DATA FOR FLAGGED MONTHS

$	: GREATER THAN OR EQUAL TO 3.0% BUT LESS THAN 5.0% :	17
$$	: GREATER THAN OR EQUAL TO 5.0% BUT LESS THAN 7.0% :	4
$$$	: GREATER THAN OR EQUAL TO 7.0% BUT LESS THAN 10.0% :	1
$$$$	: GREATER THAN OR EQUAL TO 10.0% :	0

S.4.C STATISTICS FOR MAXIMUM PERCENTAGE DIFFERENCE OF THE MONTH-TO-MONTH CHANGES IN S. A. DATA

MINIMUM	:	0.070
25TH PERCENTILE	:	1.036
MEDIAN	:	1.786
->65TH PERCENTILE	:	2.421<-
75TH PERCENTILE	:	2.928
->85TH PERCENTILE	:	3.465<-
MAXIMUM	:	8.685

S 5.A BREAKDOWN OF 3.0 PERCENT OR MORE DIFFERENCES IN YEAR-TO-YEAR CHANGES IN S. A. DATA FOR FUOECD

TOTAL : 0 OUT OF 84 (0.0 %)
(USUALLY, 10% IS TOO HIGH)

JANUARY	:	0	(AMPD = 0.8)
FEBRUARY	:	0	(AMPD = 1.1)
MARCH	:	0	(AMPD = 1.2)
APRIL	:	0	(AMPD = 0.7)
MAY	:	0	(AMPD = 0.4)
JUNE	:	0	(AMPD = 0.7)
JULY	:	0	(AMPD = 0.6)
AUGUST	:	0	(AMPD = 1.1)
SEPTEMBER	:	0	(AMPD = 0.5)
OCTOBER	:	0	(AMPD = 1.6)
NOVEMBER	:	0	(AMPD = 0.6)
DECEMBER	:	0	(AMPD = 1.0)

1976	:	0	(AMPD = 0.6)
1977	:	0	(AMPD = 0.8)
1978	:	0	(AMPD = 1.1)
1979	:	0	(AMPD = 0.8)
1980	:	0	(AMPD = 1.0)
1981	:	0	(AMPD = 1.1)
1982	:	0	(AMPD = 0.6)

AMPD = AVERAGE MAXIMUM PERCENTAGE DIFFERENCE

0 OUT OF 84 (0.0 %) OF THE MONTH(S) TESTED HAD A CHANGE OF DIRECTION.
0 OUT OF 0 (0.0 %) OF THE MONTH(S) WITH CHANGES OF DIRECTION WERE FLAGGED.

S.5.B HISTOGRAM OF MAXIMUM PERCENTAGE DIFFERENCE IN YEAR-TO-YEAR CHANGES IN S. A. DATA FOR FLAGGED MONTHS

@	: GREATER THAN OR EQUAL TO 3.0% BUT LESS THAN 4.0% :	0
@@	: GREATER THAN OR EQUAL TO 4.0% BUT LESS THAN 5.0% :	0
@@@	: GREATER THAN OR EQUAL TO 5.0% BUT LESS THAN 6.0% :	0
@@@@	: GREATER THAN OR EQUAL TO 6.0% :	0

S.5.C STATISTICS FOR MAXIMUM PERCENTAGE DIFFERENCE OF THE YEAR-TO-YEAR CHANGES IN S. A. DATA

MINIMUM	:	0.011
25TH PERCENTILE	:	0.518
MEDIAN	:	0.784
->65TH PERCENTILE	:	0.965<-
75TH PERCENTILE	:	1.141
->85TH PERCENTILE	:	1.286<-
MAXIMUM	:	2.522

S 7. RANGE ANALYSIS OF SEASONAL FACTORS FOR FUOECD

MEANS OF SEASONAL FACTORS FOR EACH MONTH
(MOVEMENTS WITHIN A MONTH SHOULD BE SMALL)

	SPAN 1		SPAN 2		SPAN 3		MPD	TOTAL	
JANUARY	99.05		99.48		100.63		1.60	99.72	
FEBRUARY	89.50	MIN	88.67	MIN	87.59	MIN	2.19	88.59	MIN
MARCH	109.89	MAX	109.47	MAX	108.10	MAX	1.65	109.15	MAX
APRIL	104.83		104.75		104.95		0.20	104.84	
MAY	99.79		100.30		102.05		2.27	100.71	
JUNE	103.46		104.58		104.17		1.08	104.07	
JULY	102.59		103.15		102.05		1.07	102.60	
AUGUST	97.95		98.94		100.84		2.96	99.24	
SEPTEMBER	90.30		90.61		90.26		0.39	90.39	
OCTOBER	96.71		96.95		98.51		1.87	97.39	
NOVEMBER	100.68		99.52		99.04		1.65	99.75	
DECEMBER	104.38		103.54		101.74		2.59	103.22	

MPD = MAXIMUM PERCENT DIFFERENCE = (MAX - MIN)/ MIN
TOTAL = AVERAGE TAKEN OVER ALL 3 SPANS

SUMMARY OF RANGE MEASURES

	RANGE MEANS	R-R MEANS	MIN SF	MAX SF	RANGE SF	R-R SF
SPAN 1	20.39	1.2278	87.91	111.32	23.41	1.2664
SPAN 2	20.80	1.2346	87.92	113.57	25.65	1.2918
SPAN 3	20.52	1.2342	86.66	113.88	27.22	1.3141
ALL SPANS	20.57	1.2322	86.66	113.88	27.22	1.3141

R-R = RANGE RATIO = MAX / MIN
SF = SEASONAL FACTORS

Discussion

William P. Cleveland
Federal Reserve Board

These two papers present valuable new tools for seasonally adjusting time series. Dagum and Quenneville offer a way to handle changing trading-day patterns, a feature which has been needed for some time. The creators of X-11 and their predecessors realized that seasonal patterns were not the same from year to year. Trading-day factors are subject to the same influences which cause seasonal factor changes. The necessity of using regression as an estimation technique for calendar related effects made it much harder to compute moving patterns. Findley et. al. provide a very nice companion paper presenting sliding spans as an evaluation tool and a way to compare specifications of models including seasonal ARIMA and regression components using Akaike's minimum AIC criterion. One could apply these ideas in addition to the tests provided in Dagum and Quenneville to select between no trading-day, fixed trading-day, or moving trading-day adjustments.

Both of these papers deserve careful reading. They reflect much careful and thoughtful work, offering valuable insights on estimation procedures and evaluation of results.

The technique selected by Dagum and Quenneville for handling moving trading-day adjustments is a direct descendent of work by several authors on stochastic coefficient models and model-based seasonal adjustment, including Gersch and Kitagawa; Havenner, Swamy, and Tinsley; and Spivey and Machak. The long memory models for seasonality explored by Dempster, Jonas, and Carlin also carry the same spirit of random walk changes to initial fixed effects. Dagum and Quenneville have given us a study of the estimation properties of their models and a smoothing technique which leads to more believable and interpretable paths for trading day effects. My experience with trading-day coefficients is that their estimates are less stable than seasonal dummy estimates, particularly over spans of five years or less. This is partly due to the time it takes the calendar to cycle through its different weekday patterns. Thus, I would favor a result not permitting trading-day coefficients to change very fast. The backward smoother employed always starts from the last point, so the coefficient paths will be sensitive to the most recent data. It might be well to devise an alternative starting point or use a forward smoother. I would also expect trading-day coefficients to be estimated along with a complete model for the series as a general practice, rather than using X-11 residuals. This prevents undesirable interactions of trading-day effects with the moving averages in X-11.

Evaluation of the quality of seasonal adjustment is a continuing problem, and we should be grateful to David Findley and his colleagues at the Census Bureau for developing a very helpful procedure in this regard. A simple test for remaining seasonality in a seasonally adjusted series does not reveal much, as an adjustment is rarely deficient in this way. One might find holiday or trading-day effects if these were needed and not included

Smoothness is also a less than satisfactory criterion. One could use the trend-cycle instead of the seasonally adjusted series if smoothness was really the ultimate goal. It seems rather that the focus is on removing the predictable seasonal movements of the series. If a sliding spans analysis shows unstable seasonal factors or seasonally adjusted values, then we should not be comfortable with estimates of the seasonal part of the series. Since this paper uses fixed trading-day coefficients, one might need to observe the stability of the coefficients relative to the length of the span used.

One of the best things about sliding spans is the variety of questions to which it speaks. This is well illustrated in the paper using several specific situations.

The other major contribution of X-12 software is the ability to estimate regression models with ARIMA error terms. This will permit better estimation of trading-day effects and provide test statistics which are not contaminated by correlated errors. One should also be able to generate forecasts and, thus, do X-11 ARIMA.

Floor Discussion

Marian Pugh
Bureau of the Census

The first paper, "Deterministic and Stochastic Models for the Estimation of Trading Day Variations," was presented by Benoit Quenneville of Statistics Canada. He defined trading day variation and presented two stochastic models - a random walk model, and a random walk with random drift model, for trading day variation. The second paper, "Toward X-12-ARIMA," was presented by David Findley of the Census Bureau. He described features that will be incorporated in this enhanced version of X-11-ARIMA. He focused on the implementation of sliding spans diagnostics and the estimation of SARIMA models with regression variables.

The discussant was William Cleveland of the Federal Reserve Board. Cleveland commended Dagum and Quenneville on their work with stochastic trading day components, a project "waiting to be done". He commented that their efforts have made previous work on time varying models more accessible and usable. He noted that Kalman filter estimation of trading day coefficients produces a ragged path of coefficients, but that the smoothing algorithm alleviates this problem. He approved of the greater weight given by the random walk model to the later years, and commented on the tradeoff between using enough data to get decent estimates and capturing short term behavior. Cleveland cautioned Quenneville and Dagum on running their program on residuals from X-11, explaining that using prefiltered data may smear out trading day effects in the estimated spectra. He suggested that they estimate trading day externally, possibly with a model based approach.

Cleveland began his discussion of "Toward X-12-ARIMA" by joking that he hoped that Findley doesn't wait until he is happy with the program before releasing it. He noted that the paper concentrates more on sliding spans than on the modeling material, whereas the presentation stressed the ARIMA modeling capabilities. He summarized that sliding spans can be used to assess the quality of a seasonal adjustment performed by X-11, and to decide issues such as the selection of seasonal filter lengths and choosing between indirect versus direct adjustment of aggregate series. He supposed that sliding spans diagnostics could also be implemented for SABL. He commented that previous to sliding spans, it was hard to get a criterion for the quality of a seasonal adjustment.

Cleveland concluded his discussion by presenting diagrams of the decision making process which he follows when he seasonally adjusts data using X-11. He remarked that both papers present helpful tools for evaluating seasonal adjustments using statistical tests and diagnostics. He asked whether artificial intelligence type expert systems could be used to automate the feedback loop he had described, and to help the novice make some of the decisions regarding seasonal adjustment.

Quenneville and Findley then responded to Cleveland's discussion. Quenneville explained that he and Dr. Dagum had started with the X-11 residual because they wanted to add the stochastic trading day as an option to the X-11 program. Findley stated that we have a responsibility to provide a reliable and simple decision making procedure for the naive user of X-11.

The floor was then opened to questions from the audience. Bob McIntire of the Bureau of Labor Statistics asked what the relationship between the developmental work at the Census Bureau and at Statistics Canada is, and whether their programs will be compatible with each other, and with previous versions of X-11. McIntire further asked whether there were plans to incorporate methodology for estimating discontinuous or moving holiday effects.

Findley responded that the newer X-11-ARIMA versions are compatible with X-11, because the X-11 program is nested within the ARIMA estimation. Findley said that the Census Bureau has not looked at discontinuous holiday models, but he hopes that the new program will enable users to model their own holiday effects. Findley added that he would like to cooperate with Statistics Canada to produce a single seasonal adjustment package, and that he would be amenable to compromise on what features should be included.

Cleveland asked what holiday effects are currently available in X-11, and Findley responded that the X-11.1 program includes the Easter, Labor Day, and Thanksgiving models used by the Census Bureau's Business Division.

Stuart Scott of the Bureau of Labor Statistics asked for advice on identifying outlier effects and interventions. Findley stated that his staff is unsure how to proceed. For the X-11 forecasting and seasonal adjustment procedure, work by Peter Burman and Mark Otto indicates that it is only important to model gross outliers. Findley's group hasn't defined "gross" yet. They have been using Bill Bell's sequential t-test procedure for identifying outliers, but have not settled on a decision criterion yet.