A comparison of methods for seasonal adjustment of the monetary aggregates

Marco Bianchi

The views expressed are those of the author and do not necessarily reflect those of the Bank of England. The author thanks Creon Butler and Spencer Dale (Monetary Analysis, BOE), Stephen Balchin and Iain MacLeay (CSO), Andrew Harvey and Siem Koopman (LSE), Tim Young (Foreign Exchange Division, BOE), Martin Boyle, Mike Clements, John Thorp and Philip Turnbull (Monetary and Financial Statistics Division, BOE) for useful comments and suggestions.

Issued by the Monetary Analysis Division, Bank of England, London, EC2R 8AH to which requests for individual copies should be addressed: envelopes should be marked for the attention of the Publications Group. (Telephone: 0171-601-4030).

Bank of England 1996
ISSN 0142-6753

Contents

Abstract

In this interim study, we compare a number of methods for seasonal adjustment of the monetary aggregates published by the Bank. The methods considered are GLAS, STL, X-11 ARIMA and STAMP. The performance of the different methods is evaluated on the monthly series of M4 and its five counter parts. We also present some evidence from simulation experiments. The aim of this interim work was to provide an initial comparison of the four methods against agreed criteria. It has been published to stimulate discussion. Follow up work will seek to reach a conclusion about the most appropriate seasonal adjustment method for the Bank of England to apply to these series.

1 Background and Scope

Over the period September 1991- May 1992 a Working Party on Seasonal Adjustment was convened by the Bank's Chief Economist, Mr M A King, with the aim of providing "*an analysis of the seasonal adjustment procedure currently in use in the Bank of England, and to compare its properties with other practical methods*" (see Working Party, 1992, page 5). It was the aim of the Working Party to provide illustrations of the properties of the different procedures, as well as to summarise the main findings in a non-technical report.

As in the Working Party (1992), it is the purpose of this study to evaluate different approaches to seasonal adjustment, using similar criteria and guidelines to judge the performance of the different methods. However, our study differs from the Working Party (1992) in the following respects:

- We concentrate our attention on the seasonal adjustment of monthly, rather than quarterly data.
- We consider four methods for seasonal adjustment, GLAS, STAMP, STL and X-11 ARIMA, the latter two of which were not considered in the Working Party. We also test a more recent release of STAMP (version 5.0), which allows the joint seasonal adjustment of multiple time series.
- We evaluate the performance of the different methods within a "live-test", by monitoring the results on a month-by-month basis, over the period October 1994 - June 1995.
- We present some evidence from a small simulation study.

The goal of our study is to put forward a framework for seasonal adjustment that can constitute a basis for further work and discussion. Thus, it is *not* the objective of the present study to make a final recommendation about which seasonal adjustment method should be used by the Bank of England. This will involve a deeper investigation, the results of which we hope to be able to report in due course.

2 Introduction and Motivation

Seasonal adjustment removes regular effects from a time series in order that more fundamental components such as the trend and the business cycle can be identified more easily. The aim of seasonal adjustment procedures is to decompose the actual series y_t into a seasonal and a nonseasonal part. We often have the decomposition

$$y_t = T_t + S_t + \epsilon_t, \qquad t = 1, \ldots, T \tag{1}$$

where T_t is the trend (or trend–cycle), S_t is the seasonal and ϵ_t is the irregular component or remainder.

It is recognised in applications that the series in the original scale, y_t, may not satisfy the additivity assumption of equation (1). A scale transformation of y_t, say $y_t^{(p)}$, is considered leading to the decomposition

$$y_t^{(p)} = T_t + S_t + \epsilon_t. \tag{2}$$

A popular class of scale transformations is the class of *power transformations*

$$y_t^{(p)} = \begin{cases} y_t^p & \text{if} \quad p > 0 \\ \ln y_t & \text{if} \quad p = 0 \\ -y_t^p & \text{if} \quad p < 0 \end{cases} . \tag{3}$$

With $p = 1$, the decomposition is purely additive, that is no transformation is actually needed to achieve additivity; with $p = 0$, the decomposition is purely multiplicative, i.e. logs are required to ensure the additivity property.[1] After the series in the transformed scale has been seasonally adjusted, the estimated unobserved components are transformed back to the original scale by the inverse transformation

$$v_t^{[p]} = \begin{cases} v_t^{1/p} & \text{if} \quad p > 0 \\ e^{v_t} & \text{if} \quad p = 0 \\ (-v_t)^{1/p} & \text{if} \quad p < 0 \end{cases} , \tag{4}$$

[1] In applications, the value of p is selected which minimizes the interaction between the trend and the seasonal, to prevent the situation in which the seasonal component has amplitudes that depend on the trend. This is done by computing a t-statistic for the interaction term in a non–additive model and choosing the power that makes the t-statistic as close to zero as possible. See Cleveland, Devlin and Terpenning (1981) for details.

where v_t can be any of the components above (trend, seasonal, irregular, etc.). Sometimes, an additional component is also included in the decomposition, called the calendar component, C_t, designed to capture the composition-of-the-month (or trading day) effects. This leads to the representation

$$y_t^{(p)} = T_t + S_t + C_t + \epsilon_t, \tag{5}$$

where the calendar component is estimated by the linear regression model

$$C_t = \sum_{k=1}^{7} \alpha_k d_k(t) \tag{6}$$

where d_k is a matrix with seven columns. For monthly series, the first column reports the number of days in the t month, so that α_1 represents the month-length effect; the remaining six columns represent the number of Saturdays, Sundays, Mondays, Tuesdays, Wednesdays and Thursdays minus the number of Fridays. The coefficients $\alpha_2, \ldots, \alpha_7$ capture the composition-of-the-month, or trading day effects.[2]

The task of seasonal adjustment is to get an estimate of the seasonal S_t in equation (2), or the seasonal and the calendar S_t and C_t in equation (5). In order to remove the variation in the data due to the seasonal component, a definition of seasonality is required; we like the definition given by Cleveland (1983), page 106:

> *When seasonally adjusting a series, in isolation, without ascribing the seasonal variation to some specific cause, the best we can do is simply say that we want to remove variation at and near the seasonal frequency (1 cycle per 12 months for*

[2] The α_k's are estimated by robust regression techniques (M-estimators) and plugged into equation (6) to obtain an estimate of the calendar component $\hat{C}_t$. Estimates of the α's are obtained from a regression where the dependent variable, given by the irregular component series obtained from an initial seasonal adjustment decomposition, is regressed on the set of explanatory variables given by the number of days in a month, the number of Saturdays, Sundays, Mondays, Tuesdays, Wednesdays and Thursdays normalized by the number of Fridays. Once the calendar component has been estimated, it is removed from the original series, and a second decomposition is run on the original series minus the calendar effects (thus, removing trading day effects and the seasonal adjustment cannot be done simultaneously, but it is rather a two-stage process).

monthly data with a yearly seasonal component) and its harmonics; just what "near" means is and must be somewhat arbritary.[3]

The outline of the paper is as follows. In Section 3 we give a brief overview of the approaches to seasonal adjustment considered. In Section 4 we put forward the criteria used for the evaluation of the different approaches, while in Section 5 we present the data set. In Sections 6-7 we present the main results. Section 8 summarises and draws some interim conclusions pending further work.

3 Overview of the Methods

The methods considered in this study are GLAS (a method developed by the Bank of England in 1992 - see Young, 1992 - based on a smoothing procedure by Lane, 1972), X-11 ARIMA (Dagum, 1988), STAMP (Harvey, Koopman, Doornik, and Shephard, 1995) and STL (Cleveland *et.al.* 1990). An overview of the methods is given in the following paragraphs. For details, readers are referred to the above mentioned papers.

GLAS: GLAS stands for "General Linear Abstraction of Seasonality". It represents the package currently used at the Bank of England for seasonal adjustment of the monetary series. Seasonal adjustment is performed by GLAS in two steps. In the first step, the trend of the series is constructed using a moving-average of data with a triangular shaped weighting pattern covering approximately two years (23 months or 7 quarters). In the second step, the trend of the series is subtracted from the actual values, and the seasonal pattern is built up from the residual series by smoothing like-month or like-quarter observations — i.e. all January values, all February values etc.[4] As for the trend component,

[3] The concept of variation is defined in terms of power spectrum at different frequencies and by harmonics it is meant seasonal wavelengths of 6, 4, 3, 2.4, 2; see Cleveland (1983) for details.

[4] These are called the '*monthly sub-series*', the series composed of all values observed in the same month, over different years.

smoothing is achieved by means of a moving–average with triangular weighting pattern with a seasonal window width of approximately half a year (5 months or 3 quarters).

The number of points used in the moving average estimation of the trend and the seasonal components, called respectively *seasonal and trend window widths*, are the only two parameters to be selected in GLAS. These parameters, denoted by n_s and n_t respectively, govern the 'degree of smoothness' of the implied trend and seasonal estimates. In particular, increasing n_t makes the trend smoother, whereas increasing n_s makes the curve fitted to the monthly sub–series smoother. By construction, n_s and n_t are odd integer numbers.

It is important to understand how an estimate of the trend at a given point in time, say t_0, is obtained in GLAS. Given the trend window width n_t, the set of n_t nearest neighbour points in time to t_0 (including t_0) is identified. Let's call this set $N(t_0)$. Define the triangular weighting function

$$k_{glas}(u) = \begin{cases} (1-|u|) & \text{for } t \in N(t_0) \\ 0 & \text{otherwise,} \end{cases} \tag{7}$$

where $u = (t_0 - t)/n_t$, then observations $y_t \in N(t_0)$ are assigned neighbourhood weights $w_t = k_{glas}(u)/\sum k_{glas}(u)$. The estimated value of the trend at time t_0 is simply calculated as the weighted average

$$T(t_0) = \sum_{t \in N(t_0)} w_t y_t. \tag{8}$$

As an example, suppose we want to estimate the value of the trend at time $t_0 = 4$ by averaging the nearest $n_t = 7$ points (including the observation at time t_0). The set of nearest neighbour points to t_0 is $N(t_0 = 4) = \{1, 2, 3, \boxed{4}, 5, 6, 7\}$. Calculating the distances $u_t = (4-t)/7$ for $t = 1, \ldots, 7$, we derive the set of weights $k_{glas}(u) = 1 - |u_t|$ that, normalised to sum to one, give:

Time Index (t)	1	2	3	4	5	6	7
$k_{glas}(u)$	0.570	0.710	0.860	1	0.860	0.710	0.570
w_t	0.108	0.135	0.162	0.189	0.162	0.135	0.108

The biggest weight is given to the observation at the evaluation point t_0, whereas weights proportionally decrease as we move away in time from the evaluation point, in either directions. The trend estimate is given by $T(t_0 = 4) = 0.108y_1 + 0.135y_2 + 0.162y_3 + 0.189y_4 + 0.162y_5 + 0.135y_6 + 0.108y_7$.

An appealing feature of GLAS is its simplicity, which is clearly underlined by the example above. Another strength of GLAS is that an algorithm is employed to determine the weights at the end points of the series, based on the theoretical work by Lane (1972). At the end points of the series, progressively more assymetric versions of the triangular weighting pattern are used. The model developed by Lane (1972) describes how to derive the weights such that *the amount of revisions to the trend and the seasonal estimates for each period as later data become available is minimised*. This "Lane minimum revision algorithm" is applied in GLAS.

STL: STL stands for "Seasonal-Trend decomposition based on Lowess", where Lowess (sometimes called Loess) stands for "**LO**cally **WE**ighted Scatterplot **S**moother". STL shares the same basic principle (or philosophy) as GLAS, namely the idea of nonparametric regression, or locally weighted averaging of the data. However, STL has a number of differences.

First of all, STL exploits a different function to calculate the local weights in the series. The *tri-cube* weight function

$$k_{stl}(u) = \begin{cases} (1 - |u|^3)^3 & \text{for } t \in N(t_0) \\ 0 & \text{otherwise,} \end{cases} \tag{9}$$

is used in STL as a replacement for equation (7). With reference to the example in the previous paragraph, the weights obtained by the tri-cube function are

Time Index (t)	1	2	3	4	5	6	7
w_t(GLAS)	0.108	0.135	0.162	0.189	0.162	0.135	0.108
w_t(STL)	0.122	0.145	0.155	0.156	0.155	0.145	0.122

The weights w_t employed by GLAS and STL are shown in Figure 1. It is evident that STL attaches weights which decay much more slowly as we move away from the evaluation point.

Secondly, an estimate of the trend (or the seasonal) at t_0 is not obtained, in STL, by simply plugging the weights into equation (8); rather, it is obtained by fitting a polynomial of degree d by *weighted least squares*, using the weights derived from equation (9). Three possibilities are available in practice, called local *constant* fitting ($d = 0$), local *linear* fitting ($d = 1$) and local *quadratic* fitting ($d = 2$); for trend estimation we can either have local linear or quadratic fitting (i.e. $d_t = 1$ or $d_t = 2$), whereas for the seasonal we can have either local constant or local linear fitting (i.e. $d_s = 0$ or $d_s = 1$). If we want to have a trend estimation with local quadratic fitting, for example, STL calculates the values of a, b, and c — $\hat{a}$, $\hat{b}$, $\hat{c}$ respectively — which minimise the weighted residual sum of squares

$$\mathrm{RSS}(t_0) = \sum_{t \in N(t_0)} w_t \cdot (y_t - a - bt - ct^2)^2, \qquad (10)$$

and then it derives the trend estimate at t_0 as

$$T(t_0) = \hat{a} + \hat{b} \cdot t_0 + \hat{c} \cdot t_0^2.$$

Thirdly, an *iterative* estimating procedure (called *backfitting algorithm*) is employed in STL. A preliminary estimate of the seasonal is derived by smoothing the monthly sub–series by a Lowess smoother with a seasonal window n_s. Then, the 12 separate smoothed sub–series are recomposed, and the long–term trend still present in these series is removed by a Lowess smoother with a trend window n_t. The backfitting algorithm alternates between the estimation of the seasonal and the trend until convergence is reached, usually after a small number of passes.[5]

[5] Such an iterative process can, however, be incompatible with the balancing constraint. In other words, this advantage of STL over GLAS may be lost in the presence of *direct* (or *ex–ante*) balancing. In fact, a solution to this problem has been suggested by Cleveland *et.al.* (1990), section 5.2, page 69, but implementing it would require considerable amount of programming.

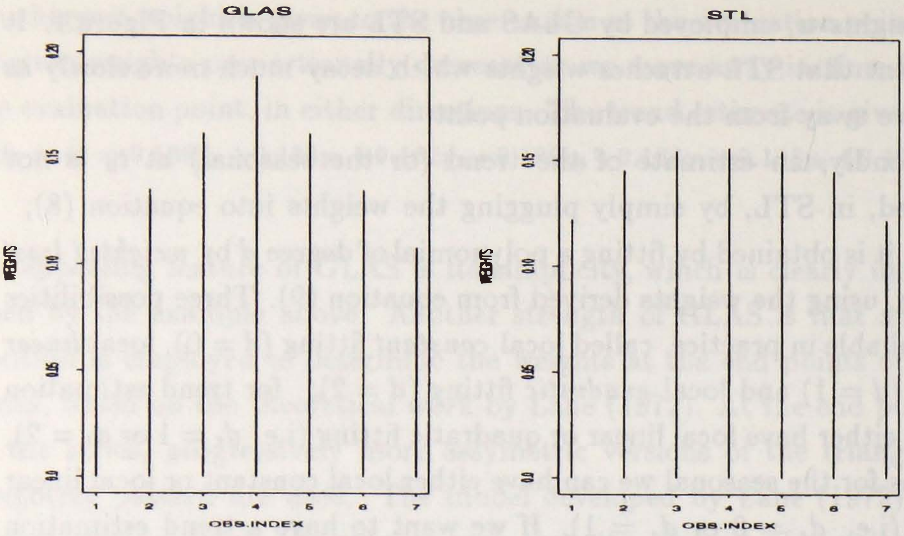

Figure 1: Neighbourhood weights w_t given to observations $y_t \in N_t$ for the fit at $t_0 = 4$ with $n_t = 7$, using the triangular (GLAS) or the tri–cube (STL) weighting functions.

The use of *iterative* weighted least squares ensures the *robustness* of the procedure; this means that outliers do not have the effect of distorting the estimate of the trend and the seasonal components in STL. The robustness property also provides the basis for the estimation of trading day effects (the calendar component C_t in equation 5) by means of standard robust regression methods (see Cleveland and Devlin, 1982, for details).

In summary, STL shares with GLAS the nonparametric approach to seasonal adjustment. Four parameters are selected in STL, including the seasonal smoothing window n_s, the trend smoothing window n_t, the degree of polynomial fitting for the seasonal and the trend, d_s and d_t respectively. A diagnostic plot, called the seasonal–irregular plot, is generated with the STL decomposition which can be used as a guide to select a 'suitable' value of n_s.[6] For a given value of n_s, n_t is *automati-*

[6] The plot helps to decide how much of the variation in the series should go into the seasonal and how much should go into the irregular component. See Cleveland *et.al.* (1990) for details.

cally chosen in a way that *the trend and the seasonal do not compete for the same variation in the data.*[7] The selection of the degree of the polynomial for the trend estimate depends, in many cases, on the underlying pattern of the data. If the trend of the series has a gentle curvature with few local maxima and minima, local linear fitting ($d_t = 1$) is appropriate. Quadratic fitting can be used otherwise. Local linear fitting is generally used for the estimate of the seasonal — that is $d_s = 1$.

X-11 ARIMA: X-11 ARIMA is a package for seasonal adjustment which represents an improvement over the X11 method developed by the Bureau of Census in the sixties. The X-11 method carries out the decomposition into trend, seasonal, and irregular by applying a series of weighted moving averages. In this respect, the method falls into the same family of nonparametric methods as GLAS and STL. Moreover, a *back-fitting* algorithm analogous to the one employed in STL is also implementated in X-11. A key difference exists however in the weighting function adopted in X-11; neither the triangular nor the tri-cube function is used, but the so called "Henderson filter" with $n_t = 43$ points for the trend estimation and a moving trimmed mean of length $n_s = 9$ for the monthly sub-series.[8]

The following steps are implemented in X-11 to obtain the seasonal adjustment (see Working Party, 1992, page 22):

(a) A preliminary estimate of trend is obtained as a moving average of the original series.

(b) A preliminary estimate of the seasonal-irregular component is obtained by subtracting the preliminary trend in step (a) from the original series.

---

[7] This result is achieved by a frequency domain analysis called by Cleveland *et.al.* 'Eigenvalue Analysis' — see Cleveland *et.al.* (1990) for technical details. As a result of such a consistent smoothing, the estimate of the trend component obtained in STL is not an output of the procedure. The purpose of the trend estimate is simply to facilitate the extraction of the seasonal component. This implies that getting a desirable trend-cycle interpretation requires a trend post-smoothing, that is a Lowess fit to the adjusted series, as discussed in Cleveland et.al. (1990).

[8] These default values can of course be changed by the user.

(c) Preliminary seasonal factors are obtained by smoothing the seasonal - irregular values for corresponding months with a moving average.

(d) The preliminary seasonal factors are centred, that is they are made to sum to zero over any 12 successive months.

(e) The preliminary seasonally adjusted series is obtained by subtracting the centred seasonal from the original series.

(f) A revised estimate of trend is obtained as a moving average of the preliminary seasonally adjusted series calculated in step (e).

(g) A revised seasonal–irregular is obtained by subtracting the revised trend estimate in step (f) from the original series.

(h) Revised seasonal factors are calculated from the revised seasonal–irregular as in step (c) and centred as in step (d).

(i) The final seasonal adjusted series is obtained by subtracting the seasonal factors in step (h) from the original series.

As in GLAS and STL, the strength of the local average smoothers in the X-11 method is in their local nature, ensuring that observations in the distant past would be adjusted slightly or not at all when a new observation is made available. Their weakness is in the high variability associated with the adjustments to the most recent observations, due to the use of *asymmetric* moving averages to obtain an adjustment at the end points of the series. This has led to the development of the X-11 ARIMA method (Dagum, 1988) whereby the series is forecasted using a seasonal ARIMA model; the X-11 smoothing operations are then applied to the extended series. In this way, symmetric filters can also be used at the end points, and smaller revisions can usually be obtained. X11–ARIMA is the recommended package of the Government Statistical Service.

STAMP: STAMP employs the idea of structural time series modelling where the unobserved components in equation (2) are assumed to follow well-defined stochastic processes. A fairly general form for the trend component is given by

$$T_t = T_{t-1} + \tau_{t-1} + v_t \tag{11}$$

with

$$\tau_t = \tau_{t-1} + u_t \tag{12}$$

where v and u are zero mean white noise processes with variances σ_T^2 and σ_τ^2. If $\sigma_\tau^2 = 0$, no stochastic slope is specified for the trend. For the seasonal component, the stochastic process is given by

$$S_t = -S_{t-1} - \cdots - S_{t-s-1} + h_t \tag{13}$$

where s is the frequency of the data (for example, $s = 12$ for monthly data) and h_t is a white noise process with variance σ_S^2. A more sophisticated model for the seasonal is obtained by a set of trigonometric terms at the seasonal frequencies (see Harvey, 1989, section 2.3.4).

The variances of the error terms, σ_ϵ^2, σ_T^2, σ_τ^2, σ_S^2, are called the *hyperparameters* of the model. The hyperparameters play in STAMP the same role that the seasonal and the trend window widths (n_s, n_t) play in GLAS, STL and X-11 ARIMA; they govern the amount of smoothing in the construction of the trend and the seasonal estimates. The bigger the value of the variance σ_T^2 relative to σ_ϵ^2 (the variance σ_S^2 relative to σ_ϵ^2), the more are past observations discounted in constructing the trend (seasonal) pattern for the forecast function. In this sense, we can define in STAMP $n_t \sim q_T = \sigma_T/\sigma_\epsilon$ and $n_s \sim q_S = \sigma_S/\sigma_\epsilon$, where q_T and q_S are called the q-ratios for the trend and the seasonal components. The weighting pattern in STAMP, for given values of the hyperparameters, is shown in Figure 2. It emerges from the figure that weights are still non-zero (though close to zero) for observations far away from the evaluation point ($t = 37$ in our plots); weights attached to some observations are negative.

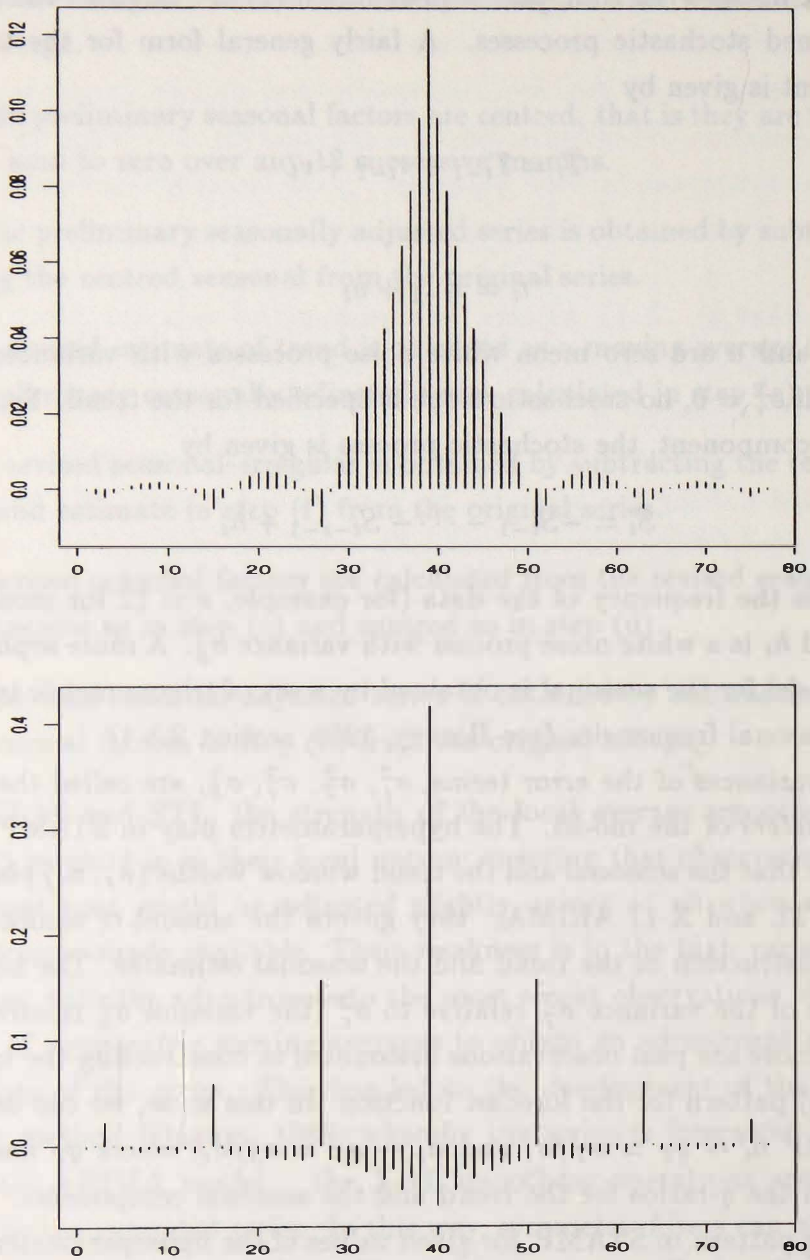

Figure 2: Weights (on the y-axis) employed by STAMP for the trend (left panel) and the seasonal (right panel) components of a monthly time series at time $t = 39$.

The appealing feature of STAMP is that the smoothing parameters (n_t and n_s) are estimated by maximum likelihood. The smoothing parameters are thus endogenously determined, rather than fixed *a priori*; moreover, these hyperparameters can (and usually will) vary over time, as long as more observations are entered into the series.[9] Given the hyperparameters, the estimates of the components, which are obtained by the Kalman filter, are optimal in the sense that they minimise the mean square error of *one-step ahead prediction*. Statistical tests on the significance of the different components (for example the time-varying slope for the trend component) are available which can lead to a simplification of the general model. Finally, *multivariate* estimation is performed in STAMP in the presence of the balancing constraint, thus ensuring efficient estimates (due to the effective use of *all* available information, given by *all* the component series).

Other relevant features in STAMP are appreciated by economists and policy-makers. These include the possibility of incorporating information on structural breaks by means of intervention dummy variables, the ability to produce forecasts and the ease of interpretation of the stochastic disturbances in the model as representing *economic shocks*.

[9] This is an example of what, in the terminology of nonparametric statistics, might be called "time varying window width".

4 The Evaluation Criteria

There are a number of desirable features that could be incorporated in a seasonal adjustment software. These can vary depending on the aim of the final users and also on the particular application at hand. For our purposes (that is the publication of a large number of seasonally adjusted series, which satisfy accounting constraints, by the Monetary and Financial Statistics Division of the Bank of England), it is considered that the following requirements are most important to the methods (see also the Seasonal Adjustment Working Party, 1992):

1. a) ability to extract fully the seasonal component over the full length of the series and b) to obtain the "best" estimate of the *current* value of the seasonal component, for policy making purposes;

2. ability to deal with balancing (or adding–up) constraints;

3. ability to estimate calendar and trading day effects;

4. a) consistency and b) rapid convergence of successive estimates to the final estimate in the presence of current updating;

5. low maintenance costs, and reliable external support if needed;

6. a) ease–of–use of the methods, ie user understands how to run the program and how to interpret the results, and b) ability to apply them to a large number of series;

7. ability to produce diagnostic statistics and graphical displays for the interpretation of the results;

8. ability to produce "reliable" trend estimates; and

9. standardisation.

The above criteria, which are in rough order of importance, are hereafter briefly discussed.

Balancing or Adding–Up Constraint: The problem of balancing is encountered in applications where the time series to be seasonally adjusted represents the aggregate of a number of seasonally adjusted components. The balancing (or adding–up) constraint means that the seasonal obtained when adjusting the aggregate series must equal the sum of the seasonals of the components, on the grounds that the same logical relationships should be preserved in the seasonally adjusted world as hold in the unadjusted world.[10] Balancing, which is considered a *crucial* requirement for the seasonal adjustment of most of the Bank's series, can be achieved in two main ways: *ex–ante* (direct balancing) by constraining the amount of smoothing applied for the estimation of the trend and the seasonal to be the same for every series; or *ex–post* (indirect balancing) by seasonally adjusting one of the series in a residual manner or by spreading the adjustment across all series, or by adjusting all the series and by nominating a row and column of the accounting matrix to act as residuals (see Read, 1991, page 19). While it is not totally clear which of the two alternatives is most advantageous, we focus in this work on *ex–ante* balancing (a brief discussion of *ex–post* balancing is presented in section 6.5).

A conceptual case on the benefits and the costs of *ex–ante* balancing has to be made.[11] Among the advantages, we have that potential problems with residual seasonality implied in the *ex–post* balancing procedure are avoided. According to a previous study on seasonal adjustment of money aggregates by Read (1991), "*the balancing adjustments sometimes proved to be so large that a few of the seasonally adjusted series contained patterns which were apparently seasonal, albeit usually of a different form from that present in the unadjusted data*" (pages 1-2). Among the costs of *ex–ante* balancing, we have to mention the difficulties in making it compatible with:

[10] There is also another type of balancing, which requires the seasonal adjustment of monthly and quarterly series to balance (see, for example, the Working Party 1992). However, this was not considered in our study.

[11] Indeed the present Bank method (GLAS) was explicitly written to avoid the need for *ex-post* balancing.

1) the use of power transformations;

2) the estimation of trading day effects; and

3) the robust estimation of the seasonal and trend components in the presence of outliers.

Since the relative importance of the costs–benefits associated with *ex-ante* versus *ex-post* balancing can vary considerably depending on the application at hand, it is important to have some guidelines indicating in which form balancing should be applied. These are discussed in more detail in the Appendix.

Extraction of the Seasonal and Policy Making: The purpose of seasonal adjustment is to remove seasonality from a series. This can be formally assessed by comparing the spectra of the seasonally adjusted and unadjusted series. If the seasonal factor has been removed, the peaks appearing at the seasonal frequencies in the spectrum of the unadjusted series should no longer appear in the spectrum of the adjusted series. In applications, it may well happen that *any* sensible seasonal adjustment method can virtually achieve the task of removing the seasonal factor. Thus, this criterion alone may not be sufficient to choose between the different seasonal adjustment methods.

A second, rather important, criterion is the ability of a method to produce the "best" estimate of what the *current* estimate of the seasonally adjusted series is for policy making purposes. Users of seasonally adjusted data are likely to be most concerned with the current behaviour of a series (i.e., its current trend in a policy context), rather than its past behaviour. As the estimation of the current trend and current seasonal implies some idea of forecasting future movements in the series, methods incorporating a forecasting aspect may provide a better framework to produce the best seasonally adjusted estimates for currrent observations.

Holidays and Trading Days: "Moving" holidays and trading day effects may be important in monthly, weekly and daily time series. Holiday effects refer to the effects associated with "moving" holidays such as Easter; [12] by trading day effects we mean the effects associated with the composition of the month (for example, the fraction of Saturdays and Sundays in each month). The value of a series can be higher or smaller in a particular month simply because the fraction of Saturdays and Sundays in that month is higher or smaller than average. Trading day effects reflect a variation in the series due to the *weekly cycle* in monthly or weekly series. Such a variation can be an important source of the total variation in the data, which should be incorporated in the calendar component (C_t in equation 5).

Revisions: Revisions are defined as the arithmetic difference between the initially published seasonally adjusted figure and the figure produced with additional data. Seasonally adjusted series subject to frequent and large revisions are generally disliked by policy makers. This is particularly true if the revisions indicate changes in the direction of the trend or cyclical movements. The relative size of revisions are primarily related to (i) differences in the smoothing linear filters applied to the same observation as it changes its position in time and to (ii) the fact that new information enters into the series when a further observation is made available.

Other things being equal, methods delivering small revisions would be preferred. But the speed of revisions is also considered to be important; if a series is to be revised, end-users should prefer that revisions are quickly completed, if possible within the first few months. A trade-off may be observed in practice between size and speed of revisions, which can make comparisons more difficult.

[12]The effects of public holidays, if these fall on the same days every year, will be part of the seasonal. Thus, no specific adjustment is needed to capture the effects of fixed public holidays (unless some systematic and material effect can be identified from the day of the week on which these occur.

Maintenance costs and external support. Maintenance costs and external support (the latter particularly important if there is likely to be a need for periodic updating of the program) very much depend on the reputation and reliability of the institution developing a software for seasonal adjustment. It must be considered here that, although within our technical competence, producing and maintaining a sophisticated package for seasonal adjustment is clearly not one of the core functions of the Bank of England. Other things being equal, we tend to favour packages developed and supported by well-known institutions or universities, as the result of a long-term and team-oriented research. Maintenance costs also depend on the complexity of the method. This argues in favour of a simple method (other things being equal). This issue is not considered further in this paper.

Ease–of–Use: At the Bank of England, as well as in several other institutions, seasonal adjustment is *routinely* applied to a large number of series, with regular weekly, monthly and quarterly up-dating (according to the frequency of the series). Therefore, a seasonal adjustment method which is transparent, easy to use and user friendly is ideally favoured. An important requirement is also the capability to run the seasonal adjustment program quickly for a large number of series (e.g. over night, or even within the working day), in a completely automated way, as part of a suite of data aggregation procedures.

Graphics and Diagnostics: Different seasonal adjustment methods may sometimes produce similar results in practice. Nonetheless, some packages may be more developed than others in terms of graphics and regression diagnostics. Graphics, for example, provide a powerful tool for judging the adequacy of the decomposition and for gaining a better understanding of the behaviour of the trend, seasonal and irregular components. Numerical statistics are also useful to judge the success of the seasonal adjustment procedure.

The following plots and diagnostic plots can be considered desirable:[13]

0. Plot of the actual series in the original scale (y_t) and its power transformation ($y_t^{(p)}$).

1. Plot of the components (actual, trend, seasonal, calendar, irregular).

2. Plot of the seasonally adjusted and unadjusted series.

3. Seasonal related plots, for the interpretation of the results of seasonal adjustment.

4. Irregular related plots, for diagnostic checking (for example, detection of trading day effects).

5. Plot of the trend with actual values and confidence bands.

The following test statistics are also desirable:

a) Test for the presence of seasonality in the series.

b) Roughness of the seasonally adjusted series ("roughness coefficients").

The former test is provided by Kendall's test statistic of *no* seasonality, given by

$$K = \frac{12}{c \cdot r \cdot (r+1)} \sum_{i=1}^{r} \left[M_i - \frac{c(r+1)}{2} \right]^2 \quad \sim \quad \chi^2_{r-1} \qquad (14)$$

where c is the number of years, r is the number of periods (for example $r = 12$ for monthly data) and M_i's are the so called "period scores".[14]

[13] A discussion of these plots can be found in the Appendix.

[14] The period scores are defined by Kendall and Ord (1993), page 24. For a monthly series, for example, monthly scores are obtained in the following way: in every year, observations are ranked from the smallest (with an assigned rank-score of 1), to the second smallest (with an assigned rank-score of 2), etc., up to the largest (with an assigned rank-score of 12). The monthly scores are obtained by summing the rank-scores for the same months across the different years.

Smoothness of the seasonally adjusted series is measured by the roughness coefficients

$$R_1 = T^{-1}\sum_{t=1}^{T}(\hat{y}_t - \hat{y}_{t-1})^2, \quad \text{or} \quad R_2 = T^{-1}\sum_{t=1}^{T}|\hat{y}_t - \hat{y}_{t-1}|, \qquad (15)$$

where $\hat{y}_t$ is the seasonally adjusted series.

Trend estimation. Policy makers often seek to extrapolate the trend of a series from the seasonally adjusted values. It is desirable that trend estimation is performed on the seasonally adjusted series, that is after having estimated and removed from the unadjusted series the seasonal component.[15] However, it is likely that the final estimate of the trend will be more sensitive to the choice of the trend smoothing procedure than to the choice of the seasonal adjustment method. The following features (listed in order of importance) are considered most relevant for internal use at the Bank of England:

1) Speed of picking up turning points in the business cycle;

2) minimal revisions; and

3) trend unaffected by outliers.

Clearly, there is a conflict between some of these criteria; it is therefore significant to list the criteria in order of importance.

Standardisation. Seasonal adjustment is performed by a number of UK governmental organisations and departments. Other things being equal, a method would be preferred which is common to most departments.

[15] This procedure, which is called *ex-post trend smoothing*, see Cleveland *et.al.* (1990), is particularly important for the nonparametric methods. In a model based procedure the trend is estimated at the same time as the seasonals. Of course, policy makers may wish in general to make their own trends based on the seasonally adjusted series.

5 The Data Set

The different seasonal adjustment methods were tested on a data set comprising of M4 and its five counterparts (see also the Seasonal Adjustment Working Party, 1992). For all the counterparts, the flow rather than the stock series were considered for seasonal adjustment. The seasonally adjusted data must satisfy the balancing constraint equation:

Change in M4	=	Public Sector Borrowing Requirement (PSBR)
	+	Public Sector Debt Sales to M4 Private Sector (PSDS) (increase minus)
	+	Sterling Lending by Banks and Building Societies to M4 Private Sector (SLPS)
	+	Sterling Net Non–Deposit Liabilities of Banks and Building Societies (SLND) (increase minus)
	+	Total External and Foreign Currency Transactions (TEXT).

Figure 3 displays the above series over the sample period 1987:1–1994:9 (93 observations).[16] It is worth noticing that some of the counterpart series display rather different patterns over time (compare for example PSBR with Sterling Lending to the M4 Private Sector, or Total External Transactions); in particular, it is apparent from the plots that seasonal patterns are more pronounced for the Public Sector Borrowing Requirment series (PSBR). This is confirmed by the Kendall's tests for the absence of seasonality reported below:

Series	M4	PSBR	PSDS	SLPS	SLND	TEXT
K-statistic	43.3	58.0	25.0	40.7	48.5	20.0
p-value	0.00	0.00	0.01	0.00	0.00	0.04

[16] The series do not incorporate prior adjustments for outliers. Similar results, however, were obtained on the prior adjusted series.

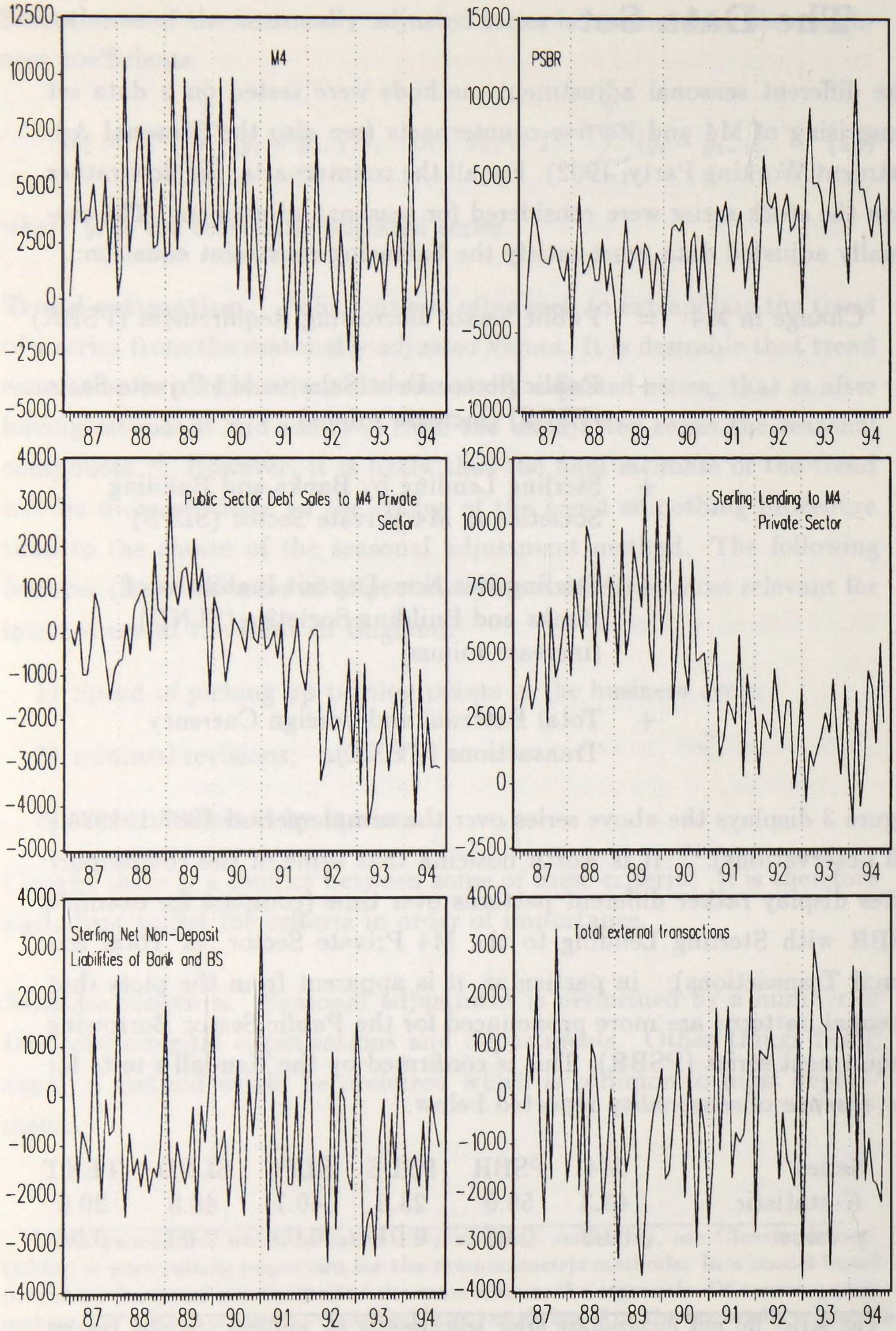

Figure 3: Series of M4 and its five counterparts. Monthly data from 1987:1 to 1994:9.

6 Empirical Results

6.1 Structure of the "Live–Test"

The live–test on the M4 series and its five counter parts is structured in the following way. For each method, all the series are seasonally adjusted using data from January 1987 to September 1994 (the latter is the date when we started our investigation). The parameters of the decomposition are selected once and for all at this stage. Although the selected parameters are then kept fixed in most cases,[17] the seasonal adjustment is re-estimated whenever a new observation is made available, i.e. there is current up–dating.

For GLAS, STL, X-11 ARIMA and STAMP, we have used the following parameters:[18]

GLAS: The seasonal window width is $n_s = 5$ and the trend window width is $n_t = 23$. These values are given as default values within GLAS.

STL: In STL, the power transformation minimising the interaction between the trend and the seasonal components is $p = 1$; the value of the seasonal smoothing window is $n_s = 9$, which leads to a value of the trend smoothing window of $n_t = 23$. Local linear fitting is used to construct the trend and the seasonal estimates, i.e. $d_s = d_t = 1$. No robust option is adopted in the decomposition, due to the need to satisfy the balancing constraint *ex–ante*. The value of the seasonal window is selected such that no significant autocorrelation is detected at the seasonal frequency (i.e. lag 12).

[17] In one of two STAMP runs they are allowed to vary — see the STAMP paragraph below.

[18] Since some of the methods do not (yet) have the trading day option, this was not used on *any* of the methods so as to compare like with like.

X-11: The default options are selected in X-11 ARIMA (but without incorporating trading day effects). The linear version was used without ARIMA forecasting, so as to preserve the balancing constraint (but for some series, such as M4, the ARIMA forecasting option would have been dropped automatically by the program, due to the poor accuracy of the forecast).

STAMP: The hyperparameters are obtained in STAMP by multivariate maximum likelihood estimation with the homogeneity restriction, applied to the five counterpart series simultaneously. In one case, these parameters are kept fixed during the live experiment; in another case, these parameters are re-estimated by the program at each point in time. In both cases we have used the trigonometric seasonal and the trend-slope component is not included in the model, since it was not found to be statistically significant.

6.2 Seasonal Adjustment of M4

For the different methods, the results of the decomposition `data = trend + seasonal + irregular` for the M4 series are reported in Figure 4. The roughness statistics obtained from the seasonally adjusted series using the different methods are:

	GLAS	STL	X-11 ARIMA	STAMP
R_1-statistic	270×10^4	350×10^4	593×10^4	414×10^4
R_2-statistic	1262	1479	1786	1573

GLAS and STL remove more variability in the original series than STAMP and X11–ARIMA. A comparison of the seasonal factors in the sample period 1992–1994 is also shown in Figure 5. It can be seen that noticeable differences are found in June 1992, 1993 and 1994, although it is far argument how significantly they are (in this period, 0.1% If M4 was about £500m).

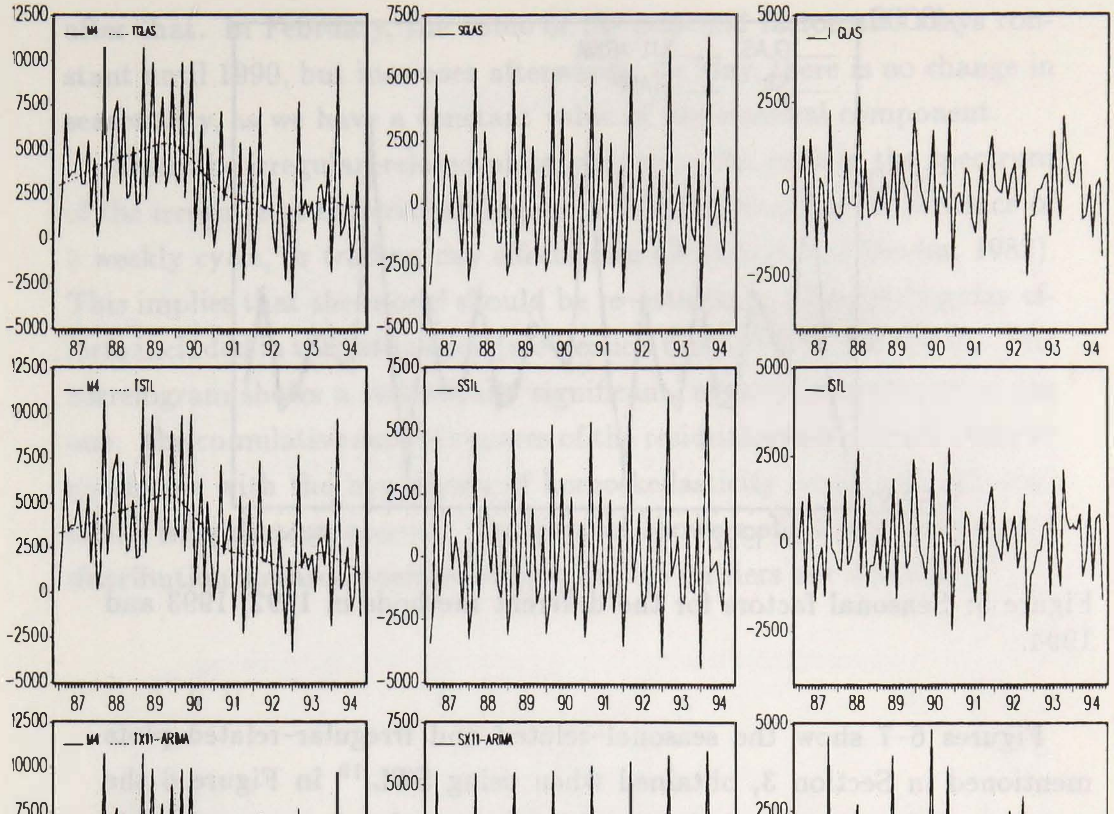

Figure 4: Decomposition of the M4 series using GLAS, STL, X-11 ARIMA and STAMP. Sample range: 1987:1–1994:9. Note: series (seasonally unadjusted trend, seasonals, irregulars) are plotted in the same scale.

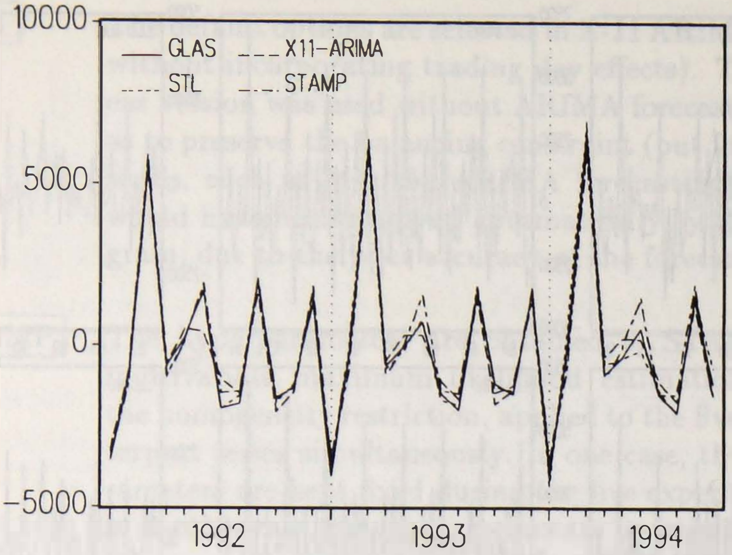

Figure 5: Seasonal factors for the different methods in 1992, 1993 and 1994.

Figures 6–7 show the seasonal-related and irregular-related plots mentioned in Section 3, obtained when using STL.[19] In Figure 6 the estimate of the spectrum indicates the presence of a three-month cycle (indicated by the vertical dotted line corresponding to the highest peak in the spectrum). This finding is confirmed by the seasonal sub-series plot: starting with the smallest level in January, the seasonal factor reaches a peak in March;[20] then, the seasonal suddenly drops in April to slowly increase once again in May and reach a new peak in June. In the last two quarters, we also have peaks in September and December but with a constant average value of the seasonal factor during the months of July, August, October and November. The three-months cycle mainly reflects the payment of interests on bank deposits, at the end of each quarter. The plot of the monthly sub-series over time reveal a slight change in seasonality patterns starting from 1990. The value of the seasonal factor in January, for example, stays constant from 1987 to 1990 but decreases

[19] For an explanation of how to interpret these plots, see the Appendix.

[20] The signs of the seasonal factors in January and March reflect particularly the seasonality of PSBR, with large tax revenue in January and end-financial year spending in March.

after that. In February, the value of the seasonal factor also stays constant until 1990, but increases afterwards. In May, there is no change in seasonality, as we have a constant value of the seasonal component.

From the irregular-related plots, we notice the peak in the spectrum of the irregular at the critical frequency 0.348, indicating the presence of a weekly cycle, or trading day effects (see Cleveland and Devlin, 1982). This implies that the model should be re-estimated with trading day effects included in the estimation (see Section 6.5 for further analysis). The correlogram shows a statistically significant, negative correlation at lag one; the cumulative sum of squares of the residuals show a linear pattern consistent with the hypothesis of homoskedasticity (constant variance) of the irregular component; the boxplot shows a slight skewness of the distribution towards positive values, but no outliers are detected.[21]

[21] The boxplot is a popular graphical display used by statisticians to summarise the distribution of the data in a visual way. The data are sorted from smallest to biggest. Then, the median is the number in the middle position, represented by the white horizontal line drawn in the box; the interquartile range, or quartile deviation, measures the distance between the first and third quartiles (the values below and above which we have 25% of the observations), represented by the height of the box; the horizontal lines called "*whiskers*" are extended from the edges of the box in both directions to at most a distance of 1.5 times the interquartile range. In summary, the size of the box represents a measure of the variability in the data; the median line within the box indicates possible asymmetry (for example skewness) of the distribution; observations falling outside the limits given by the whiskers are defined as outliers.

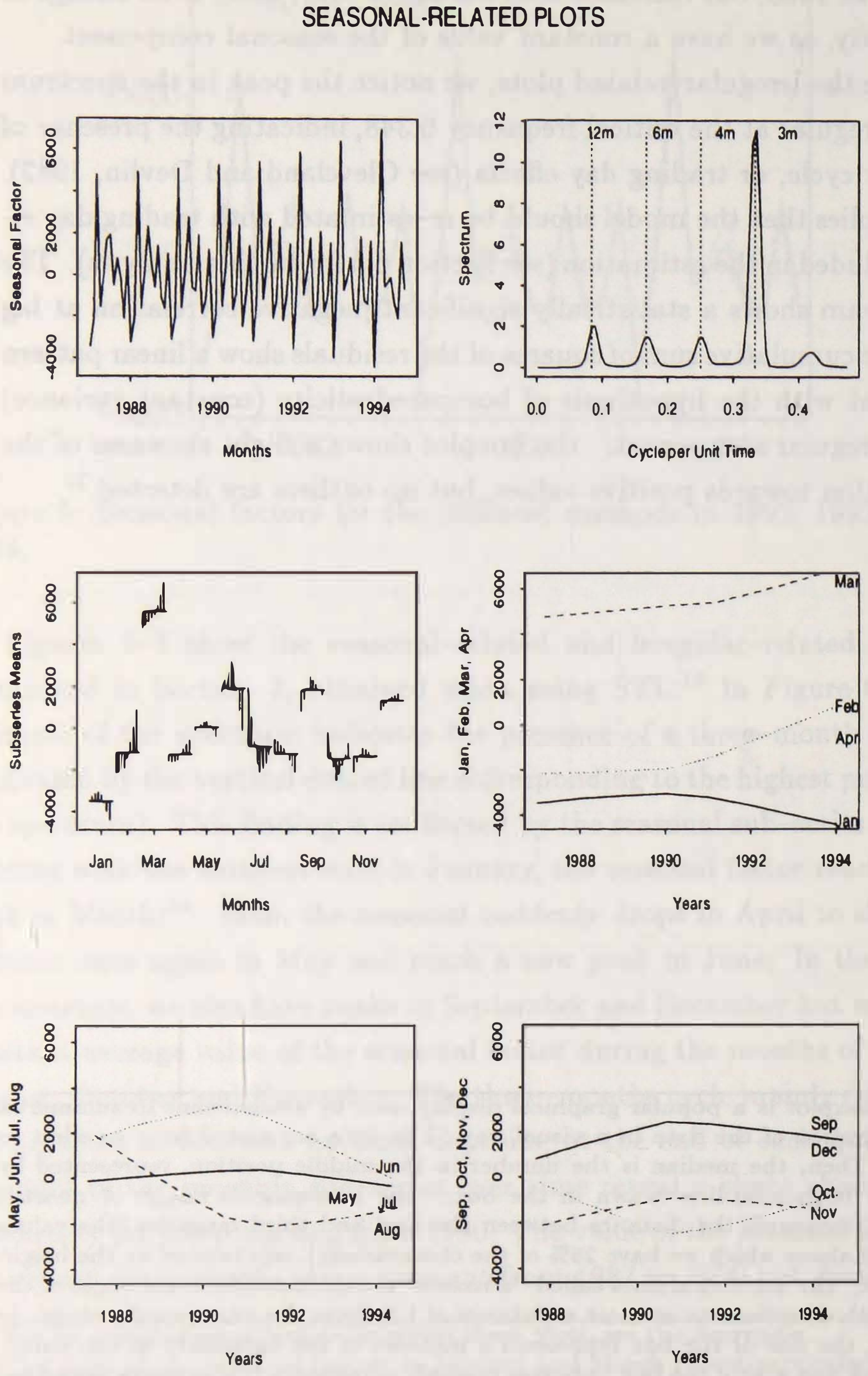

Figure 6: Seasonal–related plots.

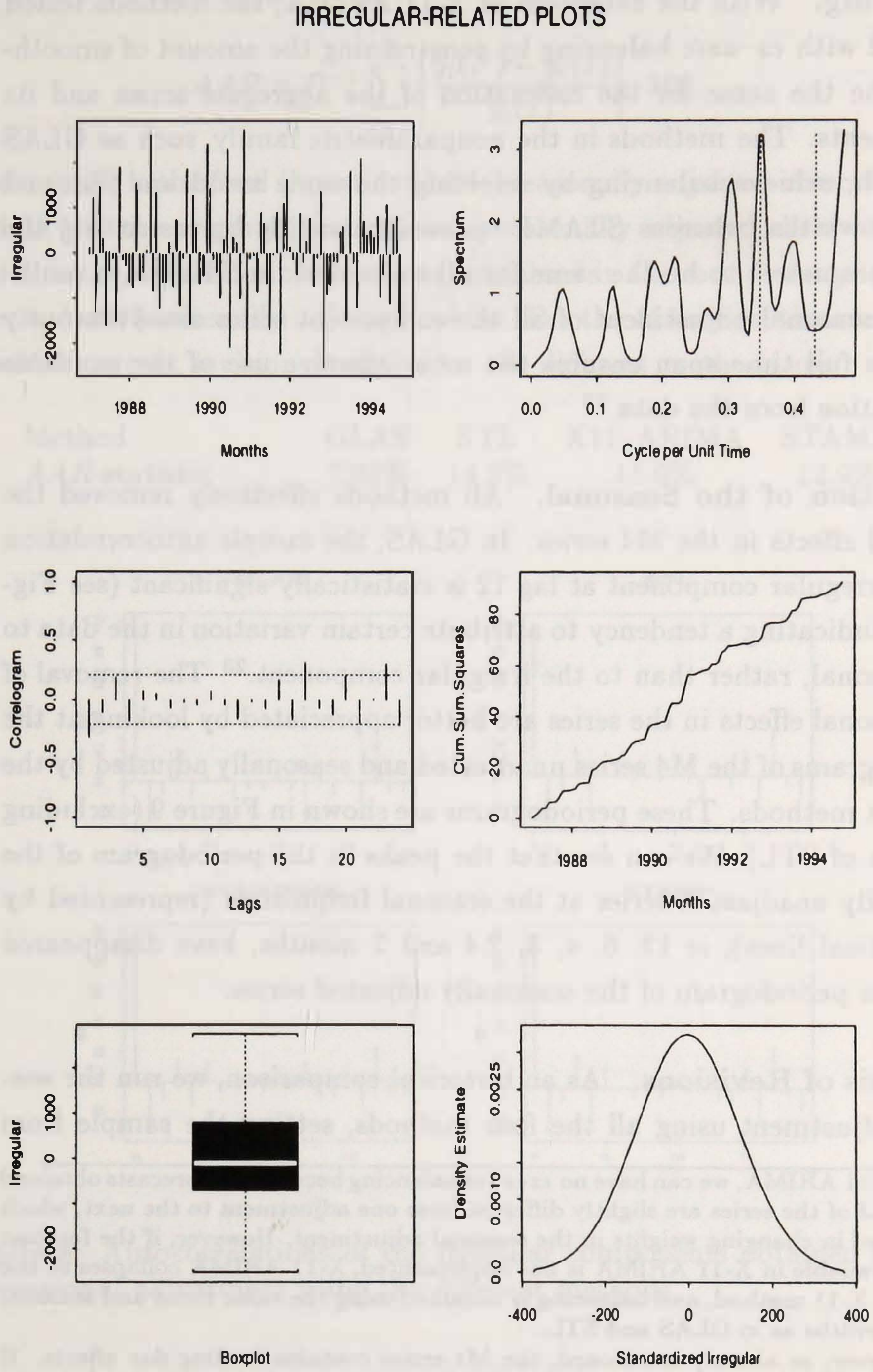

Figure 7: Irregular–related diagnostic plots. Trading–days not estimated.

6.3 Historical Comparison of the Methods

Balancing. With the exception of X-11 ARIMA, the methods tested can deal with *ex–ante* balancing by constraining the amount of smoothing to be the same for the estimation of the aggregate series and its components. The methods in the nonparametric family, such as GLAS and STL, achieve balancing by selecting the same trend and seasonal window widths, whereas STAMP ensures balancing by restricting the hyperparameters to be the same for all the series. In STAMP, a multivariate seasonal adjustment of all the component series simultaneously over the full time span ensures the most effective use of the available information from the data.[22]

Extraction of the Seasonal. All methods effectively removed the seasonal effects in the M4 series. In GLAS, the sample autocorrelation of the irregular component at lag 12 is statistically significant (see Figure 8), indicating a tendency to attribute certain variation in the data to the seasonal, rather than to the irregular component.[23] The removal of the seasonal effects in the series are better appreciated by looking at the periodograms of the M4 series unadjusted and seasonally adjusted by the different methods. These periodograms are shown in Figure 9 (excluding the case of STL). We can see that the peaks in the periodogram of the seasonally unadjasted series at the seasonal frequencies (represented by the vertical lines), ie 12, 6, 4, 3, 2.4 and 2 months, have disappeared from the periodogram of the seasonally adjusted series.

Analysis of Revisions. As an historical comparison, we run the seasonal adjustment using all the four methods, setting the sample from

[22] In X-11 ARIMA, we can have no *ex–ante* balancing because the forecasts obtained at the end of the series are slightly different from one adjustment to the next, which is reflected in changing weights in the seasonal adjustment. However, if the forecast option available in X-11 ARIMA is not implemented, X-11 ARIMA collapses to the standard X-11 method, and balancing is obtained using the same trend and seasonal window widths as in GLAS and STL.

[23] However, as already mentioned, the M4 series contains trading day effects. If these effects are estimated and removed from the series, the autocorrelation at lag 12 of the new irregular component is no longer significant in GLAS (with $n_s = 5$).

1987:1 to 1992:12 and from 1987:1 to 1994:12. A common measure of the size of revisions is given by the Average Absolute percent Revision

$$AAR = B^{-1} \sum_{t=1}^{B} \left| \frac{\hat{y}_t(F) - \hat{y}_t(I)}{\hat{y}_t(I)} \right| \cdot 100, \quad (16)$$

where $\hat{y}_t(F)$ is the final (latest available) seasonally adjusted observation and $\hat{y}_t(I)$ is the original (first estimate) seasonally adjusted observations and B is the number of observations contained in the sample range common to both (seasonally adjusted) series. For the M4 series, the average absolute percent revision statistics for the different methods are

Method	GLAS	STL	X11–ARIMA	STAMP
AAR-statistic	7.98%	14.3%	11.0%	12.9%

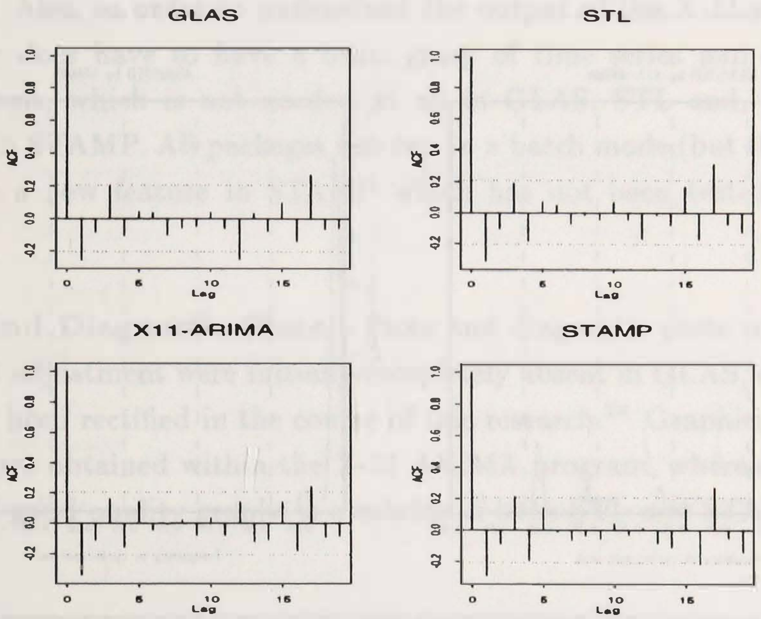

Figure 8: Autocorrelations of the irregular components obtained from the seasonal adjustment using the different methods.

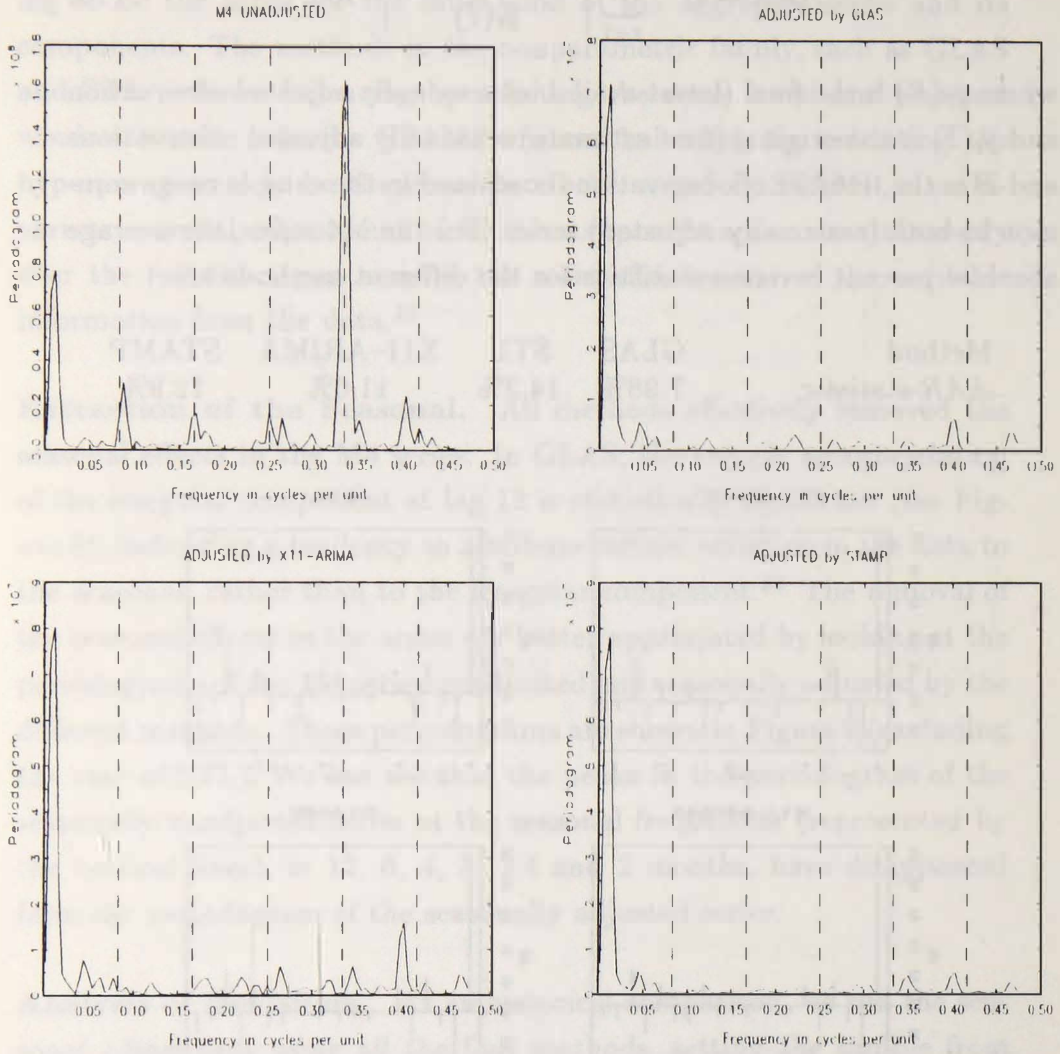

Figure 9: Periodograms of M4 seasonally unadjusted (top-left) and seasonally adjusted by GLAS (top-right), X11–ARIMA (bottom-left) and STAMP (bottom-right).

The smallest revisions are obtained by GLAS, followed by X-11 ARIMA and STAMP. The largest revisions are given by STL. These conclusions are confirmed by the plots shown in Figure 10, which shows the absolute value of revisions expressed as the difference between the seasonally adjusted series obtained in December 1992 and December 1994. The figure indicates that GLAS has relatively small and quickly converging revisions as we move far back in time. Slower convergence is observed for X-11 ARIMA and STAMP, and it fails to be satisfied in STL. Within a year, revisions are smallest for STAMP (in the bottom figure, the dotted line stands below the solid line in 1992), whereas X-11 ARIMA has frequent small revisions followed by periodical large peaks.

Ease–of–Use. All packages were easy to use with the exception of X-11 ARIMA, where it is difficult to have a full control of the available options. Also, in order to understand the output of the X-11 program, the user does have to have a basic grasp of time series and seasonal adjustment, which is not needed at all in GLAS, STL and, to some extent, in STAMP. All packages can run in a batch mode (but the batch mode is a new feature in STAMP which has not been tested in this study).

Plots and Diagnostic Plots. Plots and diagnostic plots related to seasonal adjustment were initially completely absent in GLAS, although this has been rectified in the course of this research.[24] Graphics of poor quality are obtained within the X-11 ARIMA program, whereas a wide range of good quality graphs is available in both STL and STAMP.

[24] The detection of some deficiencies of GLAS in the context of our application has led us to improve the method by generating plots and diagnostic plots as an output of the program, by estimating trading day effects and ex–post trend smoothing with confidence bands using smoothing splines. The basic GLAS seasonal adjustment procedure has been reprogrammed within the S-PLUS programming language, a most popular and widely used language in statistics (see Chambers and Hastie, 1993). By running the new GLAS program, all plots and diagnostic plots associated with the seasonal adjustment are automatically saved in a postscript file (to be subsequently printed out), whereas the seasonally adjusted series is saved in an ASCII format.

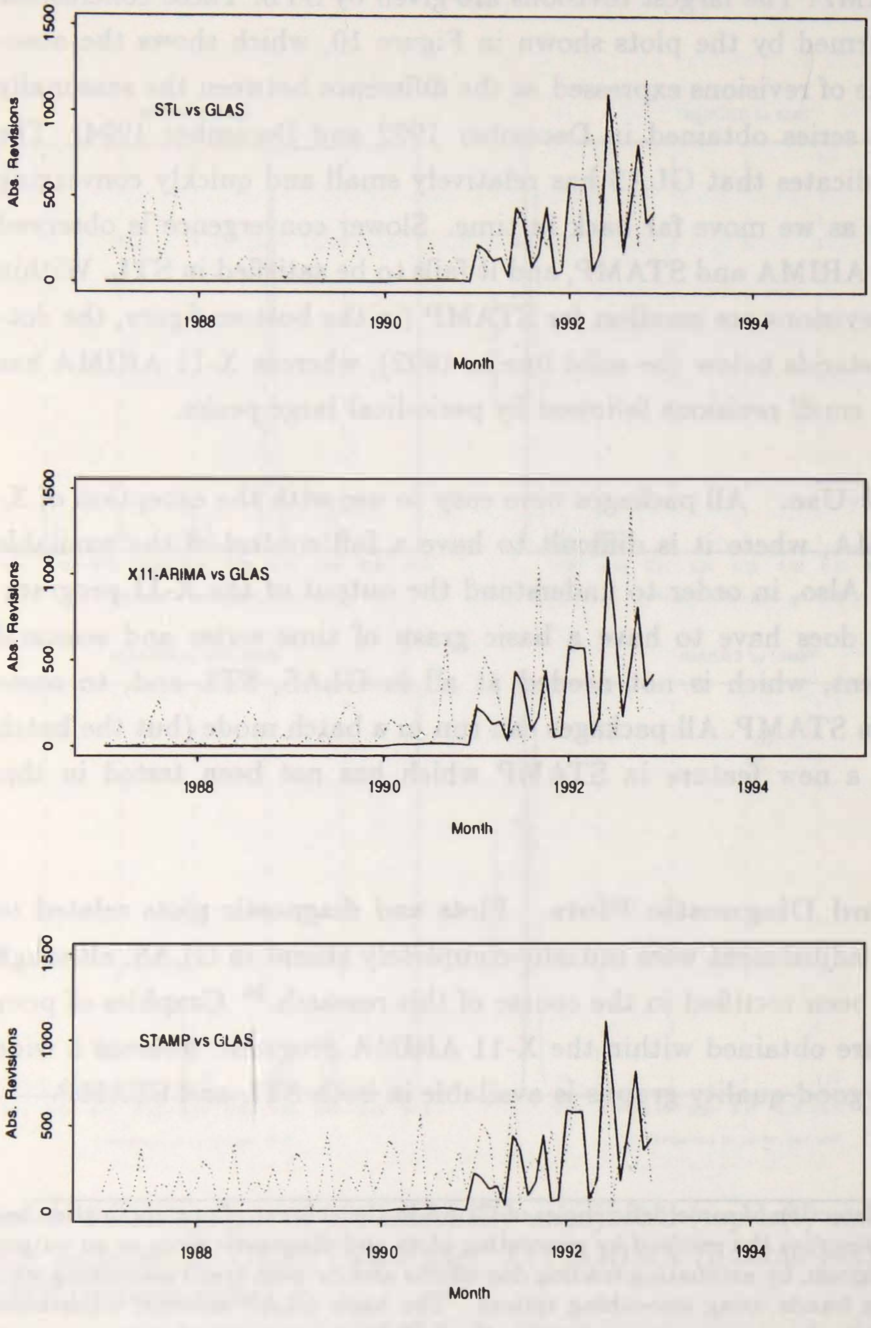

Figure 10: Revisions (in absolute value) obtained for the M4 series when subtracting the seasonally adjusted series obtained when running the seasonal adjustment in the samples 1987:1–1992:12 and 1987:1–1994:12. Note: GLAS (solid line) is taken as the reference method.

An overall comparison of the different methods against some of our evaluation criteria is shown in Table 1.

	GLAS	GLAS Improved	STAMP	X-11 ARIMA	STL
Ex-ante Balancing	Yes	Yes	Yes	Difficult	Yes
Residual Seasonality	No	No	No	No	No
Trading Days	No	Yes	Yes	Yes	Yes
Overall revisions	Small	Small	Fair	Fair	Large
Revisions within a year	Fair	Fair	Small	Fair	Large
Ease-of-Use	Good	Good	Good	Fair	Good
Batch mode	Yes	Yes	Yes	Yes	Yes
Diagnostics	Poor	Good	Good	Good	Good
Plots	None	Good	Good	Poor	Good
Trend estimates*	Fair	Good	Good	Fair	Good

Table 1: The different methods and some of our evaluation criteria. Note: * Trend estimates are in the sense of ex-post trend smoothing and our judgment refers to whether it is possible to use a variety of techniques including smoothing splines, LOWESS, and other nonparametric smoothers. These are easily available in the S-PLUS language, in which "GLAS Improved" and "STL" have been programmed.

6.4 Live–Test Comparisons

Final users of seasonal adjustment methods are frequently inclined to calculate growth rates of a seasonally adjusted series to have insights about relative movements in the series (for example the value in the current month relative to the previous month). In our study, a live test was conducted on the M4 series, starting from October 1994 until June 1995. Tables 2 and 3 report the results of the test for all the methods.

The numbers in the table represent the growth rates in M4 estimated at different points in time (what we call 'different runs'), according to the different methods.[25] The tables confirm the result already apparent in Figure 10 that STAMP has the smallest revisions within a year or so. We also have confirmation of the relatively large revisions produced by STL. For GLAS, different growth rates can be obtained in some months (in particular October 1994) depending on whether trading day effects are included or excluded from the estimation. Substantially different results are obtained when using STAMP with fixed smoothing parameters (i.e. hyperparameters) in different runs, or unconstrained smoothing parameters. These conclusions are confirmed by looking at the boxplots in Figure 11, which visually summarise, for all the methods, the distribution of the revisions in absolute value reported in the last column of Table 2.

On the basis of these findings, and if the results of seasonal adjustment are to be related to the size and the stability of revisions, STL, X-11 ARIMA and STAMP$^-$ perform relatively poorly. Interestingly enough, the best performing method appears to be STAMP with constrained hyperparameters, despite the disadvantage of not accounting for trading day effects.[26] The performance of GLAS also appears good, particularly when incorporating trading day effects.

It is also interesting to look more closely at the results obtained by

[25] The growth rates at time t are obtained by taking the ratio of the seasonally adjusted flow of M4 at time t over the value of the stock of M4 at time $t-1$.

[26] It would be nice to see here to what extent the results could be further improved by using STAMP with trading days; this option will be available in the forthcoming version of STAMP.

STAMP when estimating the hyperparameters at each point in time (STAMP$^-$), rather than fixing them at the values initially estimated in September 1994. For the trend and the seasonal components the estimated q-ratios are reported in Table 4 and plotted in Figure 12. It can be noticed that the trend q-ratios are rather stable, whereas the seasonal q-ratios are increasing over time. The latter fact accounts for the large amount of revisions associated with the method "STAMP$^-$". In fact, consider, for example, the seasonal adjustment in October 1994: at that time, we have obtained an estimate of the seasonal factor based on a q-ratio of hyperparameters of 0.039; in June 1995, however, a corresponding estimate was obtained based on a q-ratio of 0.048![27]

[27] Such a consideration might suggest that revisions in themselves should perhaps not be a very major cause of concern, as one would not get the best seasonal adjustment for October 1994 in June 1995 by using the q-ratio estimated in October 1994. A further possibility, in the case of fixed hyperparameters, would be to use the *average* value of the q_T and the q_S estimated during the period September 1993 – September 1994.

	Method	First Run	Last Run	Gap (abs.val)
Oct 94	Glas	-0.03	+0.05	0.08
	Stamp	+0.04	+0.13	0.09
	Glas*	+0.06	+0.14	0.08
	Stl*	-0.01	+0.17	0.18
	Stamp$^-$	+0.05	+0.24	0.19
	X11-Arima*	-0.14	+0.12	0.26
Nov 94	Glas	+0.82	+0.86	0.04
	Stamp	+0.83	+0.87	0.04
	Glas*	+0.85	+0.88	0.03
	Stl*	+0.67	+0.79	0.12
	Stamp$^-$	+0.79	+0.69	0.10
	X11-Arima*	+1.10	+1.09	0.01
Dec 94	Glas	+0.46	+0.52	0.06
	Stamp	+0.47	+0.52	0.05
	Glas*	+0.44	+0.50	0.06
	Stl*	+0.39	+0.51	0.12
	Stamp$^-$	+0.46	+0.52	0.06
	X11-Arima*	+0.50	+0.57	0.07
Jan 95	Glas	+0.26	+0.43	0.17
	Stamp	+0.25	+0.30	0.05
	Glas*	+0.26	+0.44	0.18
	Stl*	+0.40	+0.52	0.12
	Stamp$^-$	+0.28	+0.45	0.17
	X11-Arima*	+0.31	+0.51	0.20
Feb 95	Glas	+0.54	+0.68	0.14
	Stamp	+0.81	+0.86	0.05
	Glas*	+0.55	+0.68	0.13
	Stl*	+0.47	+0.57	0.10
	Stamp$^-$	+0.62	+0.68	0.06
	X11-Arima*	+0.73	+0.75	0.02
Mar 95	Glas	+1.00	+1.00	0.00
	Stamp	+1.22	+1.21	0.01
	Glas*	+0.93	+0.93	0.00
	Stl*	+0.73	+0.79	0.06
	Stamp$^-$	+0.77	+0.86	0.09
	X11-Arima*	+1.09	+1.00	0.09
Apr 95	Glas	+0.33	+0.40	0.07
	Stamp	+0.30	+0.32	0.02
	Glas*	+0.45	+0.52	0.07
	Stl*	+0.42	+0.53	0.09
	Stamp$^-$	+0.44	+0.50	0.06
	X11-Arima*	+0.47	+0.45	0.02
May 95	Glas	+0.90	+0.94	0.04
	Stamp	+0.95	+0.95	0.00
	Glas*	+0.89	+0.91	0.02
	Stl*	+0.86	+0.90	0.04
	Stamp$^-$	+0.79	+0.84	0.05
	X11-Arima*	+0.90	+0.85	0.05

Table 2: Rates of change (as percentage points) in M4 in the first (October 1994) and the last run (June 1995). Methods: GLAS; STAMP with balancing by homogeneity restriction in multivariate estimation, fixed hyperparameters in different runs and no trading days; GLAS*, STL* and X11-Arima* with trading days; STAMP$^-$ by univariate estimation with varying hyperparameters (that is no balancing constraint).

Runs		1	2	3	4	5	6	7	8	9
Oct 94	G	-0.03	+0.01	+0.01	-0.05	-0.02	+0.05	+0.02	+0.04	+0.05
	S	+0.04	+0.09	+0.09	+0.09	+0.10	+0.13	+0.12	+0.13	+0.13
	G^*	+0.06	+0.11	+0.10	+0.03	+0.06	+0.14	+0.11	+0.13	+0.14
	Stl^*	-0.01	+0.08	+0.08	+0.09	+0.10	+0.15	+0.14	+0.16	+0.17
	S^-	+0.05	+0.09	+0.10	+0.09	+0.13	+0.23	+0.24	+0.23	+0.24
	$X11^*$	-0.14	-0.11	-0.03	-0.03	-0.02	+0.12	+0.13	+0.12	+0.13
Nov 94		-	+0.82	+0.82	+0.74	+0.78	+0.87	+0.84	+0.85	+0.86
		-	+0.83	+0.84	+0.83	+0.86	+0.88	+0.86	+0.87	+0.87
		-	+0.85	+0.84	+0.76	+0.81	+0.89	+0.88	+0.88	+0.88
		-	+0.67	+0.69	+0.70	+0.71	+0.74	+0.74	+0.77	+0.79
		-	+0.79	+0.80	+0.78	+0.71	+0.63	+0.59	+0.65	+0.69
		-	+1.10	+1.14	+1.15	+1.16	+1.00	+1.17	+1.18	+1.09
Dec 94		-	-	+0.46	+0.37	+0.42	+0.52	+0.48	+0.50	+0.52
		-	-	+0.47	+0.46	+0.48	+0.52	+0.51	+0.52	+0.52
		-	-	+0.44	+0.34	+0.40	+0.50	+0.46	+0.49	+0.50
		-	-	+0.39	+0.41	+0.42	+0.47	+0.46	+0.50	+0.51
		-	-	+0.46	+0.45	+0.45	+0.49	+0.47	+0.51	+0.52
		-	-	+0.50	+0.50	+0.54	+0.58	+0.55	+0.54	+0.57
Jan 95		-	-	-	+0.26	+0.32	+0.44	+0.39	+0.41	+0.43
		-	-	-	+0.25	+0.28	+0.31	+0.29	+0.30	+0.30
		-	-	-	+0.26	+0.32	+0.45	+0.41	+0.43	+0.44
		-	-	-	+0.40	+0.41	+0.47	+0.46	+0.50	+0.52
		-	-	-	+0.28	+0.35	+0.44	+0.42	+0.44	+0.45
		-	-	-	+0.31	+0.35	+0.55	+0.51	+0.51	+0.51
Feb 95		-	-	-	-	+0.54	+0.68	+0.63	+0.65	+0.68
		-	-	-	-	+0.81	+0.86	+0.84	+0.86	+0.86
		-	-	-	-	+0.55	+0.68	+0.64	+0.68	+0.68
		-	-	-	-	+0.47	+0.52	+0.51	+0.55	+0.57
		-	-	-	-	+0.62	+0.61	+0.57	+0.65	+0.68
		-	-	-	-	+0.73	+0.80	+0.75	+0.76	+0.75
Mar 95		-	-	-	-	-	+1.00	+0.94	+0.97	+1.00
		-	-	-	-	-	+1.22	+1.19	+1.20	+1.21
		-	-	-	-	-	+0.93	+0.88	+0.91	+0.93
		-	-	-	-	-	+0.73	+0.71	+0.76	+0.79
		-	-	-	-	-	+0.77	+0.70	+0.81	+0.86
		-	-	-	-	-	+1.09	+0.98	+0.99	+1.00
Apr 95		-	-	-	-	-	-	+0.33	+0.37	+0.40
		-	-	-	-	-	-	+0.30	+0.32	+0.32
		-	-	-	-	-	-	+0.45	+0.50	+0.52
		-	-	-	-	-	-	+0.42	+0.50	+0.53
		-	-	-	-	-	-	+0.44	+0.49	+0.50
		-	-	-	-	-	-	+0.47	+0.49	+0.45
May 95		-	-	-	-	-	-	-	+0.90	+0.94
		-	-	-	-	-	-	-	+0.95	+0.95
		-	-	-	-	-	-	-	+0.89	+0.91
		-	-	-	-	-	-	-	+0.86	+0.90
		-	-	-	-	-	-	-	+0.79	+0.84
		-	-	-	-	-	-	-	+0.90	+0.85
Jun 95		-	-	-	-	-	-	-	-	+0.83
		-	-	-	-	-	-	-	-	+0.73
		-	-	-	-	-	-	-	-	+0.74
		-	-	-	-	-	-	-	-	+0.83
		-	-	-	-	-	-	-	-	+0.76
		-	-	-	-	-	-	-	-	+0.62

Table 3: Rates of change in M4 for all runs.

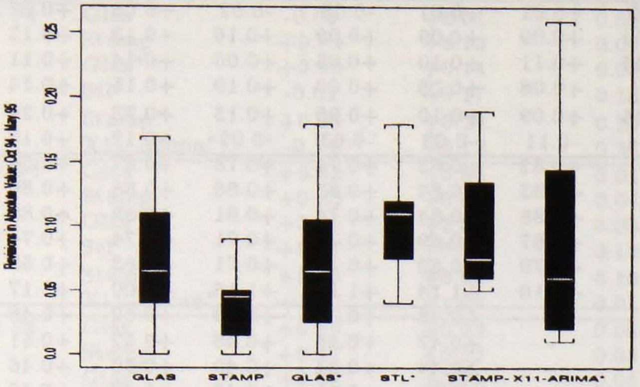

Figure 11: Boxplots of the revision gap (in absolute value) for the different methods, from October 1994 to May 1995 (i.e. numbers in Table 2, last column).

	Sep 94	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun
q_T	0.278	0.283	0.276	0.275	0.275	0.274	0.292	0.279	0.281	0.279
q_S	0.039	0.038	0.042	0.043	0.041	0.047	0.053	0.047	0.047	0.048

Table 4: q-ratios of hyperparameters estimated by STAMP for the trend and the seasonal components, from September 1994 to June 1995.

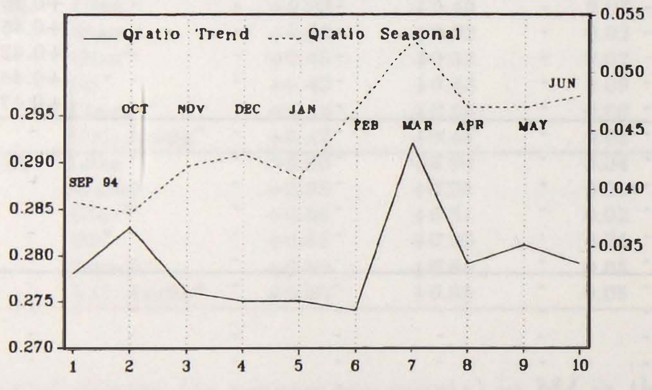

Figure 12: Plot of q-ratios of hyperparameters estimated by STAMP for the trend and the seasonal components, from September 1994 to June 1995.

It appears that a judgement on the different methods, and especially between GLAS and STAMP, very much depends on the relative importance attached to the different criteria, in particular the importance of the criterion of reasonably small and quickly converging revisions relative to reliable estimates for the very recent past. This point has been made originally by Lane (1972), who writes in the summary of his paper:

> "*Formulae are developed for moving-average weights which give the current trend estimate with the least revision on average of all filters of this type when judged against final trend estimates subsequently made on a retrospective basis with a given trend filter. This is a variety of moving regression model and assumes that the series has a certain structure locally. The basic theory is developed and the results applied to some typical series. It is shown that with schemes of this type it is difficult to get reliable estimates for the very recent past.*"

and

> "*The implication of these results is that estimates of current trend obtained by applying simple linear moving averages of this type will necessarily be rather unreliable. This is probably to be expected as estimation of the current trend implies some idea of forecasting the direction of movement of the series and we are using only very basic means to do this.*"

A method like GLAS can effectively provide reasonably small and quickly converging revisions but at a cost of less stable estimates, relative to STAMP, of most recent observations, whereas a method like STAMP has the opposite features (in STAMP, revisions are a consequence of wanting to compute optimal estimates).

In other words, if the estimation of the current trend and current seasonal implies some idea of forecasting future movements in the series, then STAMP uses much more sophisticated forecasting tools than

GLAS — thus, at the theoretical level, STAMP is more likely to produce the "best" seasonally adjusted estimates for currrent observations. By contrast, GLAS specifically seeks to minimise the period over which revisions last as well as their size. For the above reasons, STAMP appears to us more appropriate for *policy making purposes*, whereas GLAS appears more appropriate for *keeping revisions to a minimum*.

6.5 Trading Day Effects

On the basis of the results obtained for the M4 series, we have to consider the problem of balancing and trading day effects more closely. In the M4 series, the presence of trading day effects implies that a substantial variability is left in the seasonally adjusted series if a calendar component is not incorporated into the model. At the same time, the estimation of the calendar component in the M4 series poses some problems for (although, as we shall see, is not incompatible with) *ex-ante* balancing.

When the model is re-estimated including the calendar component, estimates and t-ratios of the α_k's in equation (6) obtained by robust linear regression within GLAS are reported in Table 5. For most days, the point estimates are not significantly different from zero according to standard t-tests, but the number of Wednesdays and Thursdays in a month appear to have substantially higher t-ratios. The coefficient of determination is $R^2 = 0.16$. For STL, when including the calendar component, the roughness of the seasonally adjusted series becomes smaller, as demonstrated by Figure 13, showing the seasonally adjusted series obtained by STL when excluding or including the trading day option in the estimation. A substantial change occurs following the removal of the trading day effects in the series.

We notice that the peak in the spectrum of the irregular component at the "calendar" frequency, indicating the presence of trading day effects in Figure 7,[28] has disappeared in Figure 14, thanks to the removal of trading day effects from the irregular component of the series. The correlogram (with 95% confidence bands) indicates no significant lags up to

[28] See the Appendix for details.

25 lags; the plot of the cumulative sum of squares of the (standardised) irregular shows no departure from the assumption of constant variance (homoskedasticity), whereas the boxplot and the density estimate of the irregular indicate no departure from the normality assumption (in particular, no outliers are detected in the irregular component).

A possible solution to the problem of making trading day estimation and the balancing constraint compatible is to achieve balancing by means of "prior adjustments". We can look at the diagnostic plots on the counter parts individually, to find out which counter parts contain trading day effects. Figure 15 shows the spectra of the irregular component of M4 and its five counter parts. It emerges from the plots that the presence of trading days in the M4 series is the result of the presence of trading days in the SLPS (Sterling Lending to M4 Private Sector) series. The remaining series appear to be free from trading day effects. This suggests that trading days could be estimated and removed from the SLPS series by means of "prior adjustments".[29]

[29] The M4 set of series to be seasonally adjusted include both sides of the monetary sector balance sheet: assets (counterparts) and liabilities. Such prior adjustments therefore need to be balanced on both sides and removed prior to running a seasonal adjustment program.

Variable	$\hat{\alpha}_k$	t-ratio
Number of Days in a Month	−1.12	-0.35
Number of SATURDAYS (-#FRI)	−118	-0.45
Number of SUNDAYS (-#FRI)	−369	-1.38
Number of MONDAYS (-#FRI)	+74	+0.28
Number of TUESDAYS (-#FRI)	+29	+0.11
Number of WEDNESDAYS (-#FRI)	−504	-1.86
Number of THURSDAYS (-#FRI)	+489	+1.83
Robust $R^2 = 0.16$.		

Table 5: Trading days coefficients estimated by GLAS (under S-PLUS).

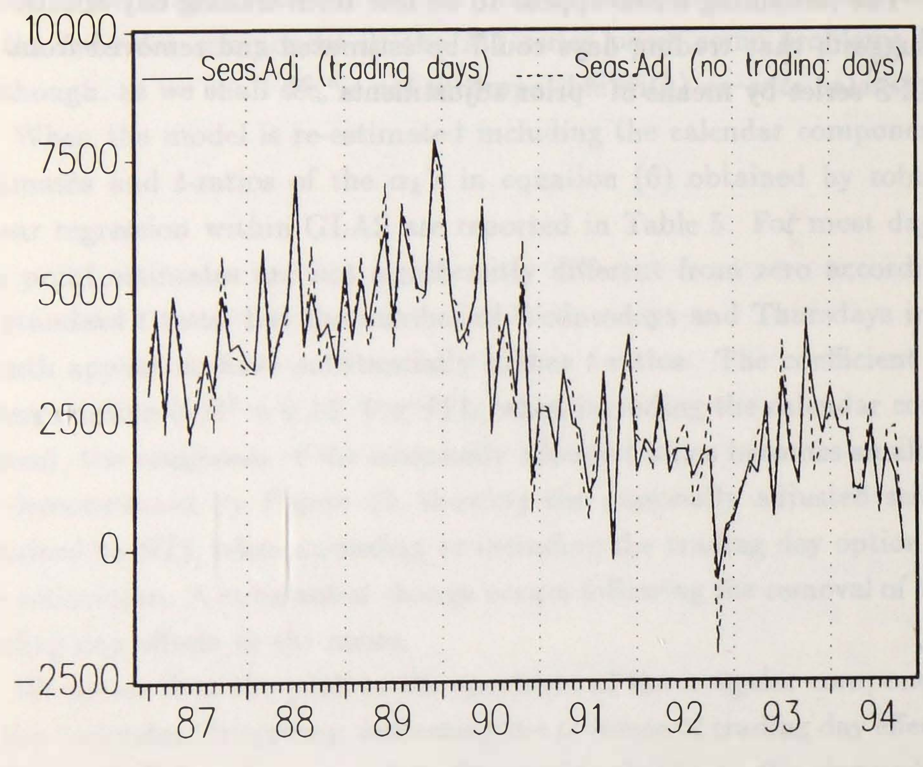

Figure 13: Seasonal adjusted series obtained by STL when including (dashed line) or excluding (solid line) trading days estimation.

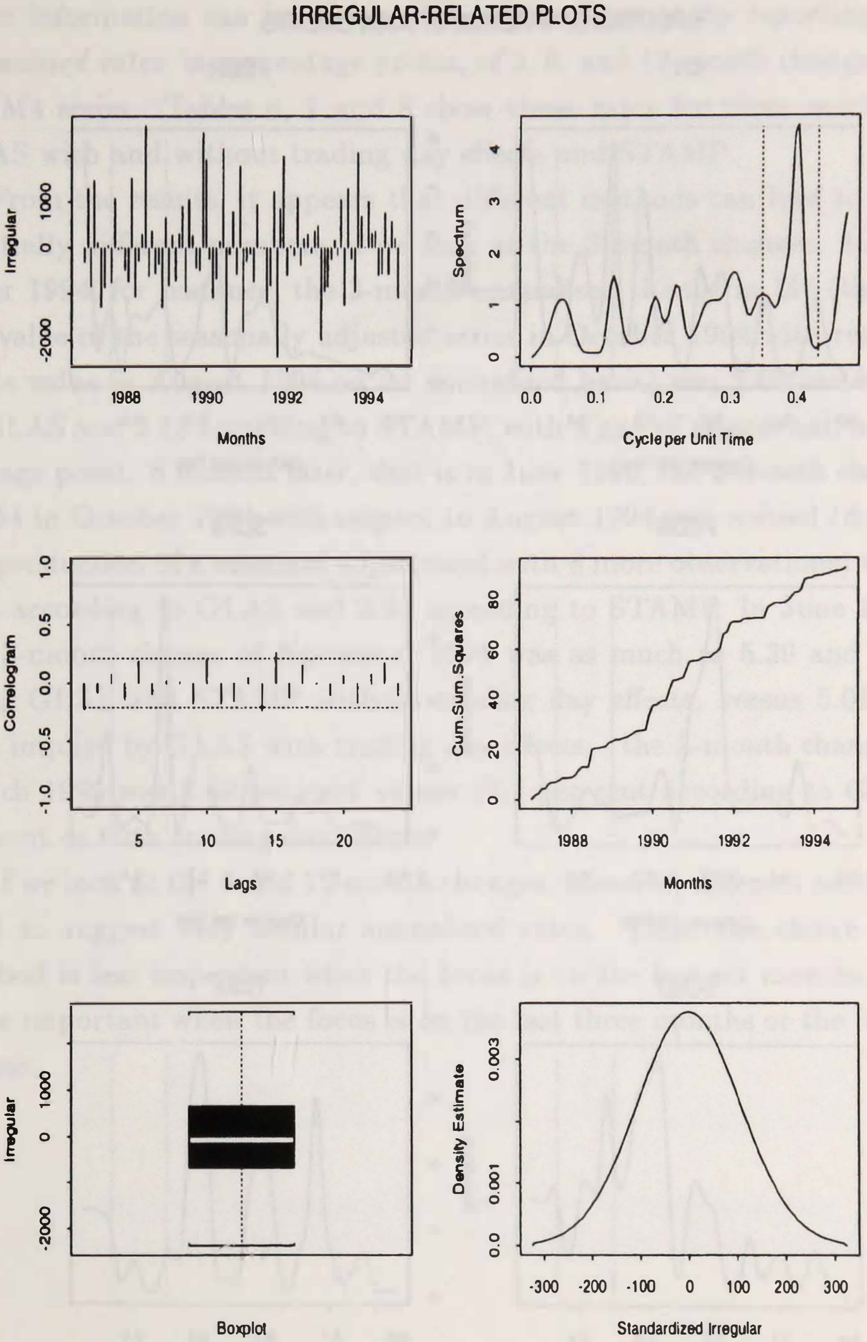

Figure 14: Irregular–related diagnostic plots. Trading days are estimated, thus no peaks at the critical frequencies appear.

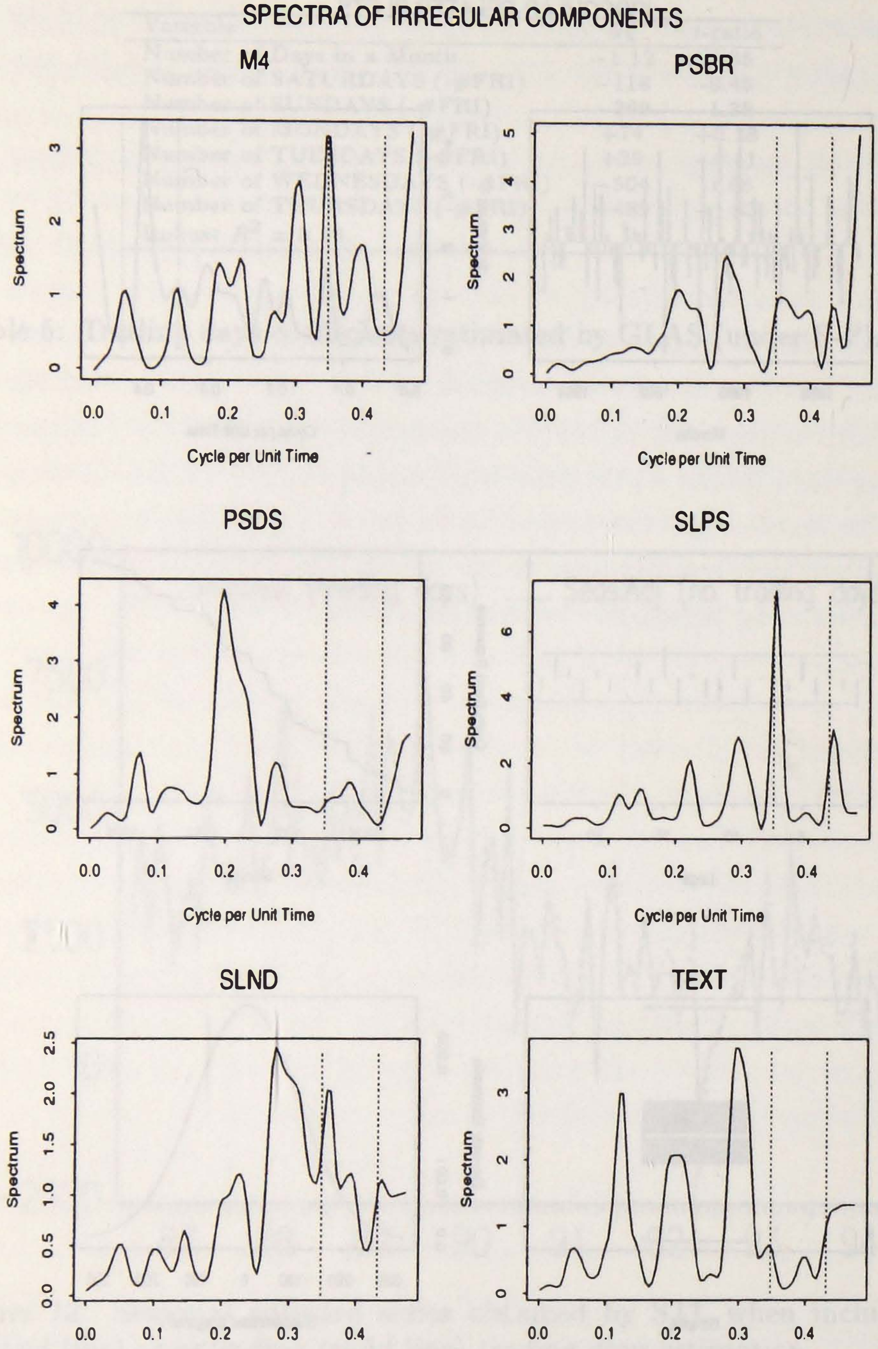

Figure 15: A check for trading day effects in the M4 series and its five counter parts.

7 Further Results

More information can be obtained from the live-test by reporting the *annualised rates*, in percentage points, of 3, 6, and 12-month changes for the M4 series. Tables 6, 7 and 8 show these rates for three methods: GLAS with and without trading day effects and STAMP.

From the results, it appears that different methods can lead to substantially different numbers, if we look at the 3-month changes. In October 1994, for instance, the 3-month annualised change in M4 (that is the value of the seasonally adjusted series in October 1994 with respect to its value in August 1994 on an annualised basis) was 2.60 according to GLAS and 2.13 according to STAMP, with a gap of almost half a percentage point. 8 months later, that is in June 1995, the 3-month change of M4 in October 1994 with respect to August 1994 was revised (due to the production of a seasonal adjustment with 8 more observations) to be 3.35 according to GLAS and 2.91 according to STAMP. In June 1995, the 3-month change of November 1994 was as much as 5.39 and 5.38 using GLAS and STAMP without trading day effects, versus 5.01 per cent implied by GLAS with trading day effects; the 3-month change of March 1995 was 8.69 per cent versus 10.1 per cent according to GLAS without or with trading day effects.

If we look at the 6 and 12-month changes, however, different methods tend to suggest very similar annualised rates. Thus, the choice of a method is less important when the focus is on the last six months, but more important when the focus is on the last three months or the latest month.

	94:10	94:11	94:12	95:1	95:2	95:3	95:4	95:5	95:6
OCTOBER 1994:									
3-month change									
G	2.60	3.04	3.03	2.37	2.67	3.42	3.16	3.28	3.35
G*	2.54	2.98	2.90	2.27	2.58	2.95	2.79	2.94	3.04
S	2.13	2.56	2.61	2.55	2.69	2.92	2.85	2.91	2.91
6-month change									
G	2.53	2.87	2.90	2.40	2.62	3.14	2.97	2.77	2.77
G*	2.46	2.83	2.78	2.32	2.53	2.78	2.68	2.45	2.33
S	1.83	2.14	2.61	2.14	2.25	2.39	2.34	2.22	2.21
12-month change									
G	3.97	4.00	4.02	3.63	3.67	4.01	3.89	3.88	3.95
G*	3.97	4.01	4.00	3.62	3.67	3.95	3.85	3.87	3.94
S	3.82	3.86	3.86	3.86	3.86	3.87	3.87	3.87	3.87
NOVEMBER 1994:									
3-month change									
G		4.99	4.95	4.15	4.53	5.47	5.13	5.28	5.39
G*		4.89	4.74	3.98	4.37	4.81	4.61	4.89	5.01
S		4.97	5.05	4.98	5.20	5.45	5.32	5.38	5.38
6-month change									
G		4.17	4.17	3.56	3.83	4.52	4.28	4.38	4.36
G*		4.12	4.04	3.45	3.73	4.06	3.92	4.08	3.93
S		3.59	3.65	3.59	3.74	3.93	3.85	3.89	3.89
12-month change									
G		4.43	4.44	4.02	4.08	4.45	4.32	4.31	4.39
G*		4.43	4.41	4.00	4.07	4.37	4.26	4.29	4.36
S		4.33	4.33	4.33	4.34	4.35	4.35	4.35	4.35
DECEMBER 1994:									
3-month change									
G			6.24	4.29	4.76	5.89	5.47	5.67	5.81
G*			5.18	4.28	4.76	5.29	5.04	5.33	5.58
S			5.74	5.64	5.88	6.24	6.11	6.22	6.22
6-month change									
G			4.27	3.54	3.88	4.73	4.43	4.56	4.66
G*			4.33	3.63	3.98	4.38	4.20	4.40	4.57
S			4.14	4.07	4.26	4.50	4.40	4.46	4.47
12-month change									
G			4.46	4.00	4.08	4.49	4.34	4.35	4.43
G*			4.45	4.00	4.10	4.41	4.29	4.33	4.42
S			4.32	4.32	4.33	4.35	4.35	4.35	4.35

Table 6: Annualised rates of M4 in percentage points. Legend: G = GLAS without trading days; GLAS* = GLAS with trading days; S = STAMP.

	94:10	94:11	94:12	95:1	95:2	95:3	95:4	95:5	95:6
JANUARY 1995:									
3-month change									
G				5.61	6.18	7.52	7.00	7.25	7.45
G*				5.54	6.13	6.74	6.44	6.79	7.12
S				6.32	6.64	7.03	6.83	6.94	6.94
6-month change									
G				3.96	4.39	5.42	5.04	5.21	5.35
G*				3.87	4.32	4.80	4.58	4.82	5.03
S				4.40	4.63	4.93	4.80	4.88	4.88
12-month change									
G				3.89	3.99	4.46	4.29	4.30	4.39
G*				3.88	3.99	4.33	4.20	4.25	4.35
S				4.22	4.25	4.28	4.26	4.27	4.27
FEBRUARY 1995:									
3-month change									
G					5.19	6.73	6.13	6.42	6.66
G*					5.41	6.10	5.75	6.14	6.53
S					6.43	6.95	6.74	6.91	6.91
6-month change									
G					4.83	6.05	5.59	5.81	5.98
G*					4.86	5.42	5.15	5.45	5.73
S					5.77	6.15	5.99	6.10	6.10
12-month change									
G					4.12	4.65	4.45	4.48	4.58
G*					4.13	4.50	4.35	4.42	4.54
S					4.46	4.50	4.49	4.50	4.50
MARCH 1995:									
3-month change									
G						8.75	8.05	8.39	8.69
G*						9.58	9.16	9.66	10.1
S						9.93	9.65	9.83	9.83
6-month change									
G						7.25	6.69	6.96	7.18
G*						7.35	7.02	7.41	7.17
S						7.99	7.79	7.93	7.93
12-month change									
G						5.27	5.05	5.09	5.21
G*						5.29	5.13	5.22	5.35
S						5.22	5.20	5.21	5.21

Table 7: Annualised rates of M4 in percentage points. Legend: G = GLAS without trading days; GLAS* = GLAS with trading days; S = STAMP.

	94:10	94:11	94:12	95:1	95:2	95:3	95:4	95:5	95:6
							APRIL 1995:		
							3-month change		
G							7.79	8.19	8.55
G*							9.08	9.68	10.3
S							9.66	9.90	9.91
							6-month change		
G							7.33	7.65	7.92
G*							7.68	8.14	8.59
S							8.16	8.33	8.33
							12-month change		
G							5.08	5.13	5.26
G*							5.10	5.21	5.36
S							5.16	5.18	5.18
							MAY 1995:		
							3-month change		
G								9.24	9.67
G*								10.5	11.1
S								10.3	10.3
							6-month change		
G								7.75	8.08
G*								8.20	8.71
S								8.50	8.50
							12-month change		
G								5.97	6.12
G*								6.04	6.21
S								6.09	6.09
							JUNE 1995:		
							3-month change		
G									8.96
G*									7.88
S									8.23
							6-month change		
G									8.73
G*									8.90
S									8.93
							12-month change		
G									6.58
G*									6.62
S									6.58

Table 8: Annualised rates of M4 in percentage points. Legend: G = GLAS without trading days; GLAS* = GLAS with trading days; S = STAMP.

7.1 An application to a longer M4 series

We consider a longer M4 (flow) series from 1983:1 to 1995:6 and look at the performance of GLAS and STAMP in the year 1990 (we know that in this year there has been a turning point in M4, with the beginning of a downward trend). We estimate the seasonal component by GLAS and STAMP using all the data and record the values in 1990. Then, we estimate the seasonal factor in January 1990 using the data from 1983:1 to 1990:1, in February using the data from 1983:1 to 1990:2, etc. We obtain the results reported in Table 9.

It appears from the table that similar estimates of the seasonal component in 1990 are obtained when using both methods over the period 1983:1–1995:6, except for the values in November and December. We can interpret the numbers reported under the column "all obs" as a proxy for the "true", unknown value of the seasonal, whereas the numbers under the column "1 obs" can be interpreted as a first guess of the seasonal in January when the January unadjusted data is made available, in February once the February data is made available, etc. The discrepancy between the first guess and the proxy for the true value of the seasonal component in different months is expressed in percentage points in the second and the last columns. It appears that STAMP has larger revisions than GLAS on average. This is consistent with the findings previously reported in Figure 10. The reason for large revisions, far back in time, in STAMP is two-fold: (i) STAMP attaches non-zero weights to observations, even those in the distant past; (ii) the ratios of the hyperparameters (sometimes called q-ratios) estimated in the periods 1983:1–1989:12 and 1983:1–1995:6 are very different. In the period 1983:1–1989:12, the ratios are $\sigma_T/\sigma_\epsilon = 0.280$ and $\sigma_S/\sigma_\epsilon = 0.219$, whereas in the period 1983:1–1995:6 we have $\sigma_T/\sigma_\epsilon = 0.300$ and $\sigma_S/\sigma_\epsilon = 0.124$. Revisions are a consequence, in STAMP, of wanting to compute optimal current estimates.

	GLAS			STAMP		
	gap %	1 obs	all obs	all obs	1 obs	gap %
Jan 90	4	-2476	-2577	-2584	-2285	12
Feb 90	37	-2808	-2052	-1971	-2717	38
Mar 90	10	+4334	+4838	+4912	+4367	11
Apr 90	71	-2839	-1656	-1957	-3214	64
May 90	97	+15	+444	+496	-154	131
Jun 90	5	+3480	+3682	+3924	+4319	10
Jul 90	14	-1877	-2192	-2712	-4131	52
Aug 90	11	-863	-967	-766	-1110	45
Sep 90	29	+1942	+2735	+2475	+1792	28
Oct 90	2	-2813	-2859	-2969	-3981	34
Nov 90	15	-1001	-749	+78	-41	153
Dec 90	56	+644	+1471	+806	-988	223

Table 9: Values of the seasonal factor estimated by GLAS and STAMP with limited or full information (that is using just one more or all available observations).

7.2 A Small Simulation Study

It is sometimes difficult to judge about the relative performance of the different methods if the underlying data generating process is unknown. For this reason, we conduct in this section some simulation experiments where we generate a series[30]

$$y_t = Trend_t + Seas_t + Irr_t$$

with $Trend_t = (T_t - \text{mean}(T_t))/\text{stand.dev.}(T_t)$, where T_t is the trend of the series obtained by applying a smoothing spline to the M4 series over the period 1983:1 to 1990:12 and $Irr_t \sim NID(0, 0.5^2)$. In the different experiments we have then the following specifications for the seasonal:

[30] These experiments are very much exploratory in this relatively short section, and do not pretend therefore to provide comprehensive results on the ability of the various methods to extract the seasonal component from a series.

Exp I(a): $S_t = sin(t)$ for $t = 1, \ldots, 96$;

Exp I(b): $S_t = sin(t) + d_t$, with $d_t = d_0 + \sum_{t=2}^{T} \Delta_t$, $d_0 = 0.1$ and $\Delta_t = 0.01$;

Exp II(a): $S_t = -1$ for $t = 1, 4, 7, \ldots$; $S_t = -1$ for $t = 2, 5, 8, \ldots$; $S_t = +2$ for $t = 3, 6, 9, \ldots$;

Exp II(b): S_t as in Exp II(a), plus a deterministic factor $d_t = d_0 + \sum_{t=2}^{T} \Delta_t$ with $d_0 = 0.2$ and $\Delta_t = 0.005$;

Exp III(a): S_t as in Exp I(a) for $t = 1, \ldots, 48$ and as in Exp II(a) for $t = 49, \ldots, 96$;

Exp IV(a): we consider the PSBR series from 1987:1 to 1994:12 (we know that for these series trading day effects are absent) and seasonally adjust it using GLAS; then, the simulated series is generated as $y_t = T_t + S_t + Irr_t$, where T_t and S_t are respectively the trend and seasonal components obtained by applying GLAS to the PSBR series and $Irr \sim N(0,1) \cdot mean(T_t)$.[31] On the series generated in such a way, we run the seasonal adjustment using GLAS and STAMP;

Exp IV(b): As in Exp IV(a), but using STAMP to generate T_t and S_t, and with the same irregular component Irr_t.

[31] In other words, the generated series differs from the actual PSBR series by substituting a white noise error component for the previous irregular component.

We stress that the seasonal component has a cycle of 6 months in Experiment I and a cycle of 3 months in Experiment II. The actual series with trend, seasonal and irregular components in experiments I(a) and I(b) are shown in Figure 16. In Figure 17, where the performance of GLAS and STAMP is evaluated by comparing the estimated with the true seasonal components, it appears that GLAS gets closer to the true seasonal component than STAMP. In particular, this can be inferred based on the box plot charts summarising the distribution of the error series. In Figure 18, concerning Experiment II, the opposite appears to be true, whereas a similar performance of the two methods is found in Experiment III (see Figure 19).

The results for Experiment IV are reported in Figure 20. It appears that STAMP can better estimate the seasonal component generated by GLAS, while GLAS can better estimate the seasonal component generated by STAMP.[32]

[32] These results are certainly counter-intuitive, as we would expect GLAS (STAMP) to perform better when the simulated series is the sum of the unobserved trend and seasonal components generated by GLAS (STAMP). Instead, we just have the opposite result! However, Findley (1983), who was the first to suggest this type of experiment, found similar results when comparing two methods for seasonal adjustment such as X-11 and a Bayesian method developed by Akaike.

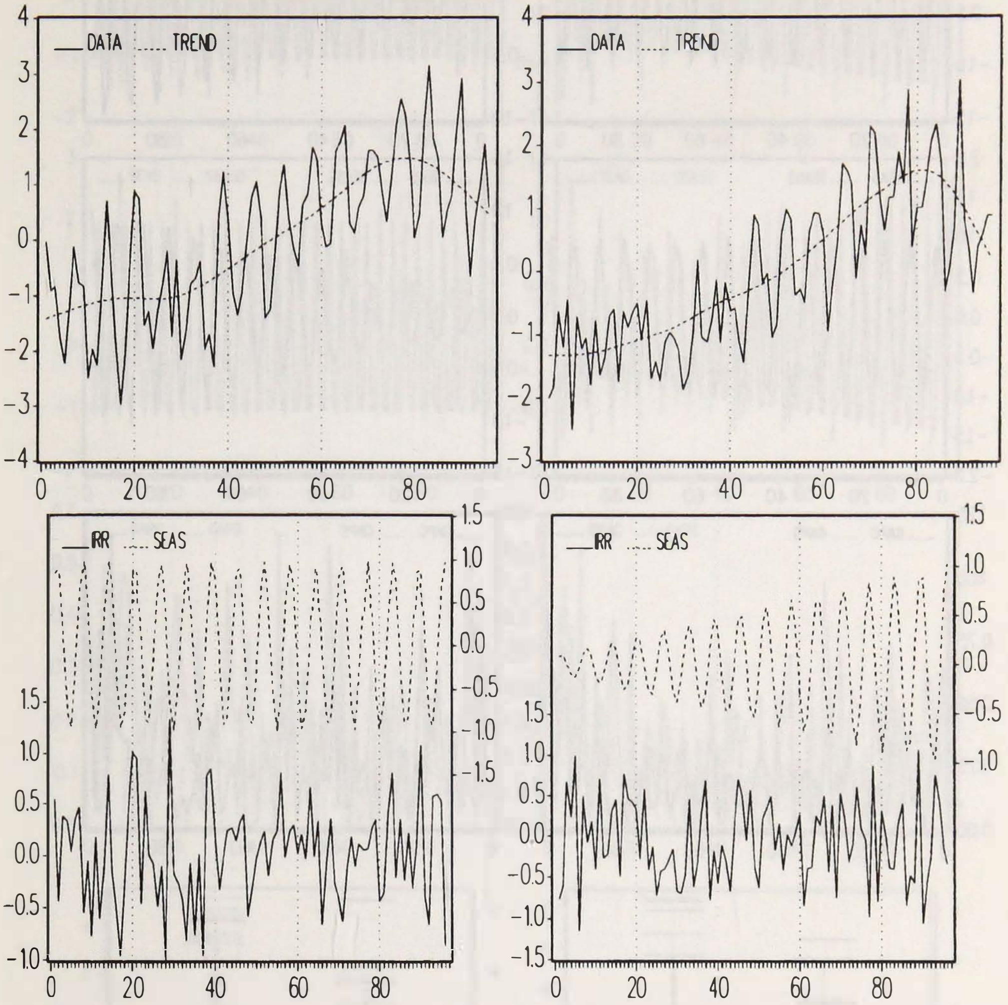

Figure 16: Simulated series with constant seasonal factor (left charts) and increasing seasonal factor (right charts).

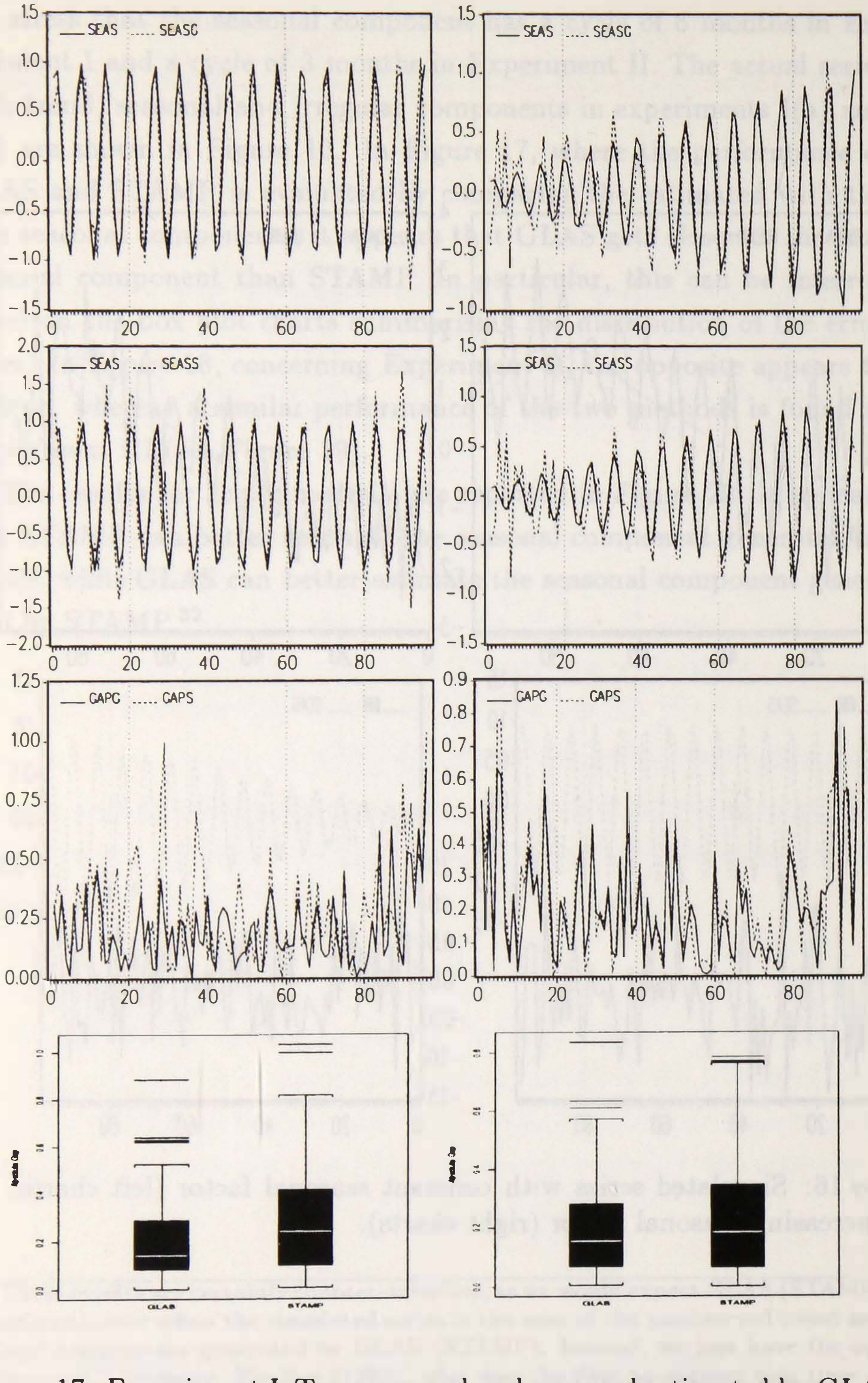

Figure 17: Experiment I. True seasonal and seasonal estimated by GLAS (G) and STAMP (S). Gaps between true and estimated series are in absolute value. Boxplots are of the gap series. Note: outliers are represented by horizontal lines.

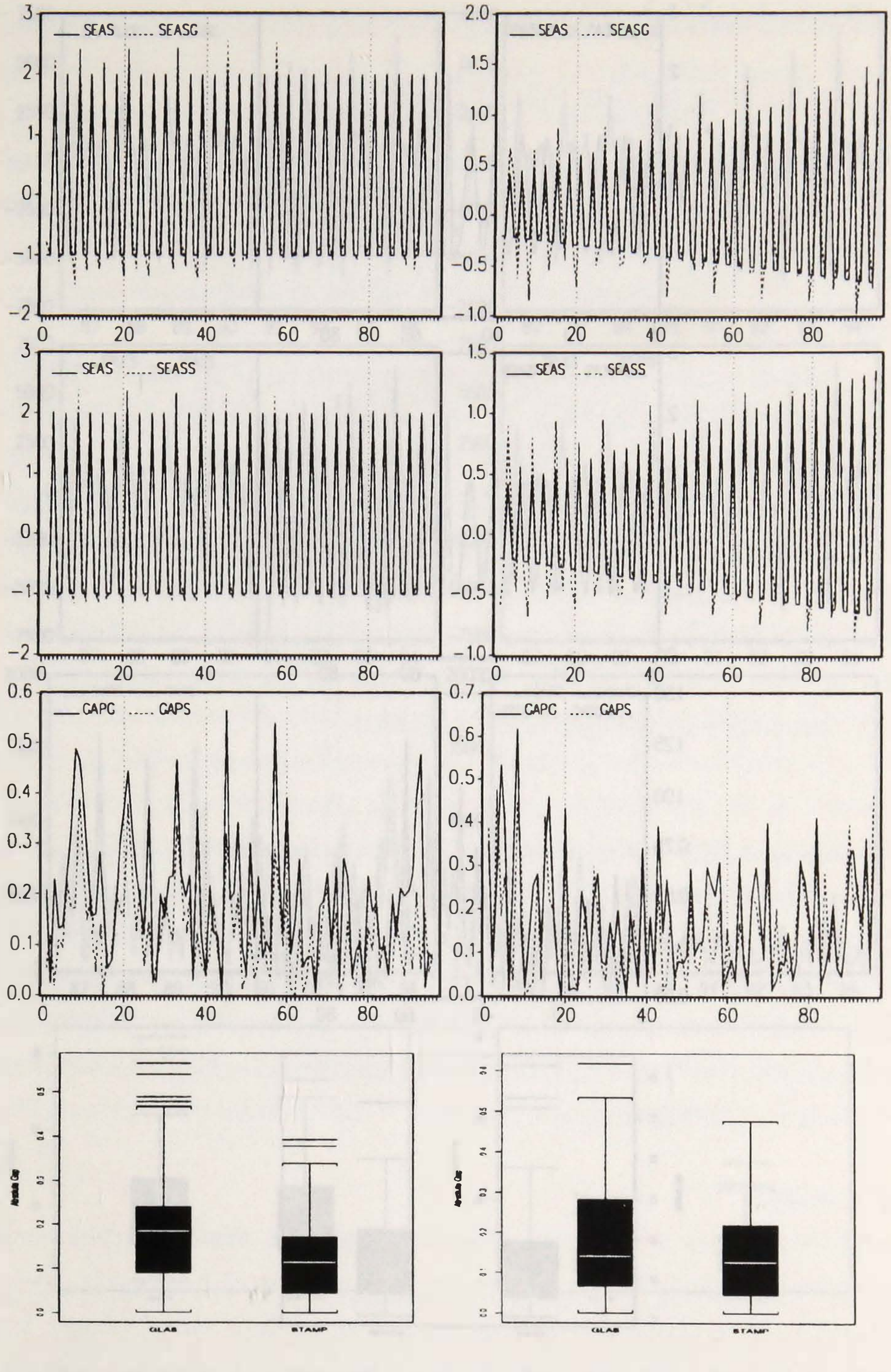

Figure 18: Experiment II. True seasonal and seasonal estimated by GLAS (G) and STAMP (S). Gaps between true and estimated in absolute value. Boxplots of the gap series.

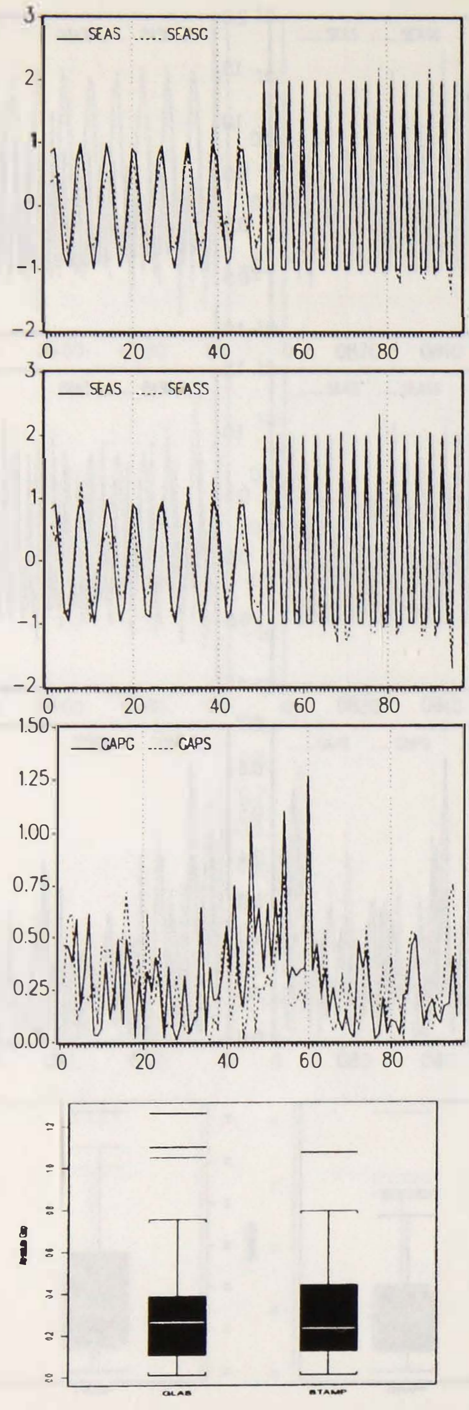

Figure 19: Experiment III. True seasonal and seasonal estimated by GLAS (G) and STAMP (S). Gaps between true and estimated in absolute value. Boxplots of the gap series.

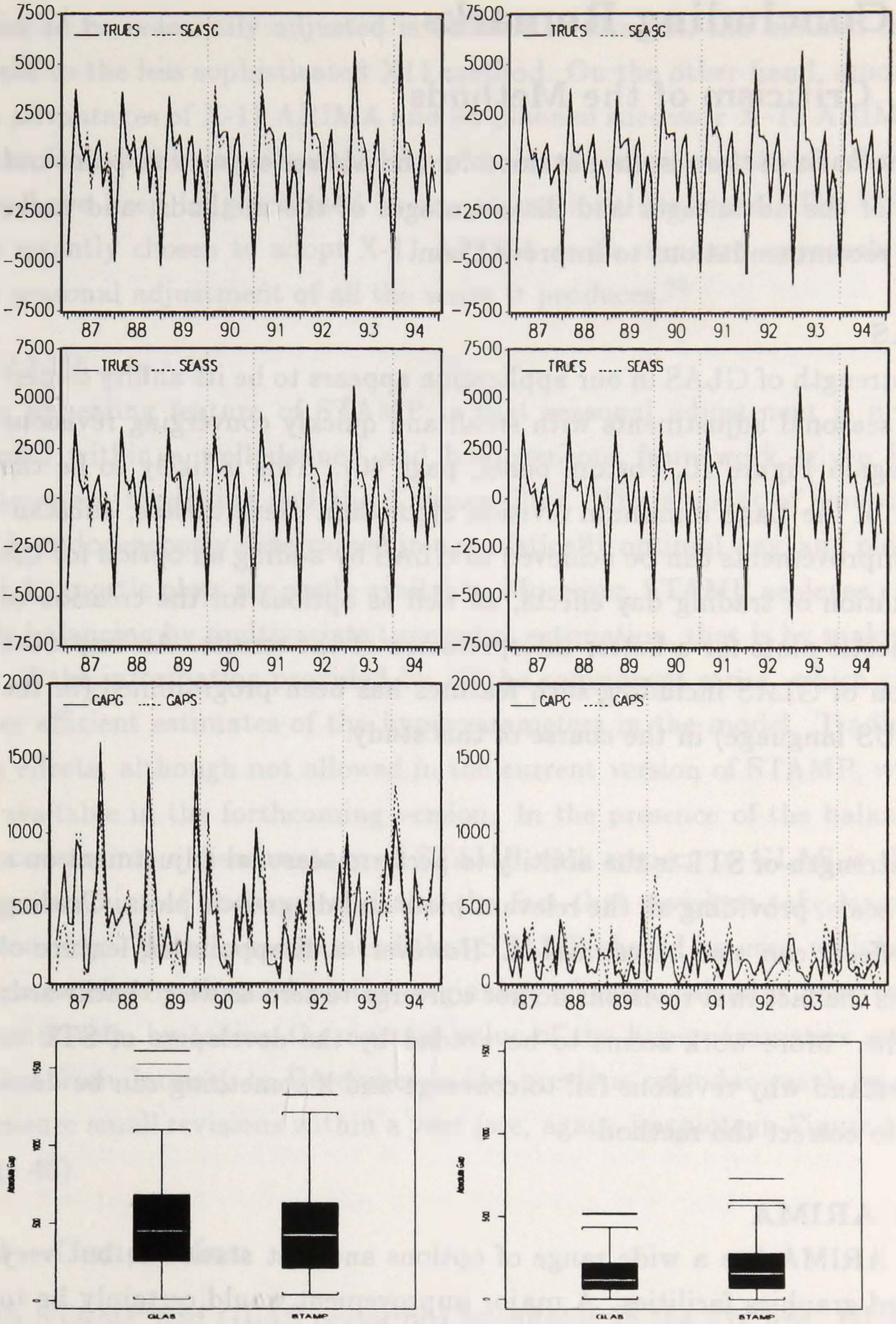

Figure 20: Experiment IV. True seasonal and seasonal estimated by GLAS (G) and STAMP (S). Gaps between true and estimated in absolute value. Boxplots of the gap series. Note: Seasonal generated by GLAS (left column) and by STAMP (right column).

8 Concluding Remarks

8.1 Criticism of the Methods

On the basis of the results obtained for the M4 series, we can point out some of the advantages and disadvantages of the methods, and make some recommendations to improve them.

GLAS

The strength of GLAS in our application appears to be its ability to perform seasonal adjustments with small and quickly converging revisions (see again Figure 10, bottom panel, page 40). This is likely to be the effect of the Lane minimum revision algorithm. Nevertheless, substantial improvements can be achieved in GLAS by adding an option for the estimation of trading day effects, as well as options for the creation of diagnostic plots for a better interpretation of the results. An improved version of GLAS including such features has been programmed (in the S-PLUS language) in the course of this study.

STL

The strength of STL is the abiltity to perform seasonal adjustment on a large scale, providing all the relevant plots and diagnostic plots. Trading day effects can also be estimated. However, a disappointing feature of STL is the fact that revisions do not converge to zero as we go backwards in time. More work seems to be needed by the developers of STL to understand why revisions fail to converge and if something can be done here to correct the method.

X-11 ARIMA

X-11 ARIMA has a wide range of options and test statistics, but very limited graphics facilities. A major improvement would certainly be to incorporate the statistical methods within a modern environment for graphics. Another difficulty with X-11 ARIMA is its somewhat complicated design. Moreover, in the presence of *ex-ante* balancing, or if the series to be seasonally adjusted is difficult to forecast, the method collapses to the less sophisticated X11 method. On the other hand, among the advantages of X-11 ARIMA and its planned successor X–12 ARIMA is the fact that being developed by prominent central statistical offices, they have been long accepted as an international standard. The CSO has recently chosen to adopt X-11 ARIMA as its standard approach to the seasonal adjustment of all the series it produces.[33]

STAMP

The appealing feature of STAMP is that seasonal adjustment is performed within a well-defined and homogeneous framework, given by state-space modelling and the Kalman filter. The amount of smoothing is endogeneously determined in a statistically optimal way, and plots and diagnostic plots are easily available. Moreover, STAMP achieves *ex-ante* balancing by multivariate time series estimation, that is by making use of the information provided by *all* the component series, which ensures efficient estimates of the hyperparameters in the model. Trading day effects, although not allowed in the current version of STAMP, will be available in the forthcoming version. In the presence of the balancing constraint, a disadvantage of STAMP with respect to GLAS is the larger amount of revisions and also the fact that revisions take longer to converge. We also recommend that STAMP should be used with hyperparameters estimated at the beginning of every calendar year (or, alternatively, by taking the *average* value of the hyperparameters estimated from January to December in the previous calendar year), so as to ensure small revisions within a year (see, again, boxplots in Figure 11, page 46).

8.2 Conclusion

Both STAMP and GLAS performed acceptably in the live test. GLAS has smaller and quickly converging revisions, is simple to use and explain, and has had a good record in operational use over the last four

[33]See the "Report of the Task Force on Seasonal Adjustment", Government Statistical Service Methods Committee, February 1996.

years. On the other hand, at the theoretical level STAMP should produce better seasonally adjusted estimates for the latest observation. If the estimation of the current trend and current seasonal is taken to require forecasting of future movements in the series, then STAMP uses much more sophisticated forecasting tools than GLAS.

8.3 Some Directions for Future Research

The results presented in this paper constitute a basis for a further discussion of the comparison of seasonal adjustment methods, but they also implicitly suggest improvements and directions for further research that we could summarise as follows:

1. In principle, whichever method is used for seasonal adjustment, it should have the capacity to estimate trading day effects. As shown in our application on the M4 and the Sterling Lending to Private Sector series, trading day effects may account for a substantial variation in the data. Of course, the possibility of removing trading day effects raises an issue of compatibility with the balancing constraint. More investigation is clearly needed to understand how one can retain balancing with trading day adjustments.

2. Should we rethink the balancing constraint? How do other countries and central banks handle the problem in other applications (such as flow of funds data)? A reason for the perhaps disappointing performance of X-11 ARIMA may be the balancing constraint. The CSO favour post-balancing and we are aware that some other central banks do not apply it fully in the seasonal adjustment of monetary data. The above questions will need to be addressed in future work.

3. Despite the results presented in the paper and the effort made to shed light on different methods, a clear recommendation in favour of a particular method cannot be made here. This paper looks only at technical considerations, other factors such as cost of disruption, maintenance costs and availability of support, and the

potential advantages of using an industry standard need also to be considered.

4. As regards future technical research, we intend to carry out new live-tests using the improved version of GLAS, the improved version of STAMP, and the improved version of X-11 ARIMA (the latter is actually called X-12 ARIMA, and it has just been officially released by the Bureau of Census in the US).[34]

5. Whichever method is chosen, further consideration should be given to the estimation and the possible publication of trend series, since it is likely to be of particular interest to policy makers.

6. The seasonal adjustment of the weekly (Wednesday–observed) M0 series, which has not been looked at in the present study, would also need to be addressed in future work.

[34] Some main advantages of X-12 ARIMA with respect to X-11 ARIMA are: a good instruction manual is available for X-12 ARIMA; and ability to handle jumps in the series.

Appendix

Diagnostic Plots

The seasonal–related and irregular–related plots presented in Figures 6–7 are a routine output in STL and in the new GLAS program (both running under the S-PLUS package) and therefore deserve some explanation. The seasonal–related plots provide information on the seasonal pattern in the series and serve to facilitate the interpretation of the results of the seasonal adjustment. The irregular–related plots are designed, instead, for diagnostic checking purposes, or model adequancy.

The seasonal–related plots are composed of: plot of the seasonal component, spectrum of the seasonal, plot of the seasonal sub–series, month by month and over time. The spectrum of the seasonal reveals how much of the total variability in the seasonal factor is explained at different frequencies, labeled "cycles per unit of time" on the x-axis. These frequencies can be converted into months by taking the reciprocal; thus, we have for example that 0.0833333 cycles per unit of time is equivalent to a length of cycle of $1/0.0833333 = 12$ months. Other important frequencies, corresponding to cycles with length of 6, 4 and 3 months are stressed by the vertical dotted lines shown together with the plot of the spectrum. The plot of the seasonal sub–series provides useful information about the mean value of the seasonal factor in different months (represented by the horizontal lines) as well as the variability about this mean value in different years (protrayed by the vertical lines). It can be seen in Figure 6, for example, that February, March, June and July are months with large variability in the seasonal factor, as opposed to April, May and December which are characterised by small variability. The plots of each seasonal subseries over time provide another way of summarising the information given by the subseries plot, by focusing more on the evolution of the seasonal factors over time.

The irregular–related plots are composed of: plot of the irregular component with upper and lower bands ($\pm 2\sigma(\epsilon)$) for the identification of outliers, spectrum of the irregular component (with outliers clipped) for a check of trading day effects; correlogram and cumulative sum of squares of the irregular component for a check of the white noise process; boxplot and normal–quantile (or density estimation) plot for a check of the normality of the irregular component. In view of the detection of possible trading day effects, the most important plot is given by the spectrum of the irregular component. In particular, for a series which has a weekly cycle but is recordered monthly, the weekly harmonic occurs with frequency

$$\frac{365.25}{7 \times 12} = 4.348$$

which correspond to a frequency (called "critical frequency") of 0.348. Another critical frequency for trading day effects is 0.432 (see Cleveland and Devlin, 1982). Vertical dotted lines corresponding to these two frequencies are drawn with the spectrum of the irregular component in our graphical representation; peaks in the spectrum at the critical frequencies 0.348 and 0.432 indicate the presence of trading day effects.

More on the Balancing Constraint

It is argued in the paper that the potential conflict between inclusion of trading day effects and balancing has to be addressed more deeply. In principle, there are a number of possible ways of addressing this in the seasonal adjustment of M4 and its counter parts:

0. *Drop the balancing constraint.*

1. *Balancing by prior adjustments.* Trading day effects are treated like prior adjustments, that is they are estimated and removed from the Sterling Lending to Private Sector series, which is then redefined accordingly. The seasonal adjustment with balancing is then applied to M4 as the sum of the five counter parts, with SLPS redifined as indicated.

2. *Balancing by neglecting a series.* The series of total external transactions, a less monitored series among the M4 counter parts, could be seasonally adjusted in a residual manner (that is, in practice,

the seasonal adjustment program is not applied to this series). This might be a workable solution if we think that, both from a statistical (seasonality test) and an economic viewpoint, it is not clear at all whether this series is seasonal.

3. *Indirect balancing.* This method is recommended by the CSO in those situations where component series, say A, B, C add up to a single series D. The method consists of seasonally adjusting each of the component series individually using the best possible options (i.e. accounting for trading days where necessary, using different smoothing parameters for different seasonal subseries, etc). From the seasonal component obtained in this way, $S(A)$, $S(B)$, $S(C)$, the seasonal component for the aggregate is defined as the sum of the seasonal for the components, that is $S(D) = S(A) + S(B) + S(C)$. The seasonally adjusted series for the aggregate is obtained as $D - S(D)$. Since the aggregate series is adjusted in an indirect way, a check for the absence of residual seasonality in the D series is in order. If residual seasonality is detected, the seasonal adjustment of the component series is revised again and again, using different options, until no residual seasonality is left.

4. *Balancing through a residual component.* This type of balancing can be used in a situation like: $A + B + C = D + E$. Here, all the components from A to E are individually seasonally adjusted and a residual series is created as: $R(S) = [A(S) + B(S) + C(S)] - [D(S) + E(S)]$. The residual series is then redistributed to all individual series in a way which is proportional to the variance of the series in comparison to the trend. We have therefore $A^*(S) + B^*(S) + C^*(S) = D^*(S) + E^*(S)$, where $A^*(S) = a \cdot R(S) + A(S)$, $B^*(S) = b \cdot R(S) + B(S)$, etc., with $a + b + c + d + e = 1$. A check for residual seasonality is run on all series $A - A^*(S)$, $B - B^*(S)$, ..., $E - E^*(S)$.

Among the different possibility, options 0 and 3 are apparently in use at the Bundesbank, whereas options 3 and 4 are recommended by the CSO.

Effect of the balancing constraint in STAMP

Figure 21 presents the absolute revisions obtained for the M4 series by STAMP (relative to GLAS) when estimating the model under the balancing constraint (that is using the "homogeneity" option in the STAMP multivariate model estimation menu) or unconstrained for balancing (univariate estimation with fixed hyperparameters). The figure reveals a somewhat substantial price to be paid for the accounting balancing in STAMP, in terms of larger absolute revisions. A practical solution in applications could be to cut down on revisions by not doing them after a certain point.

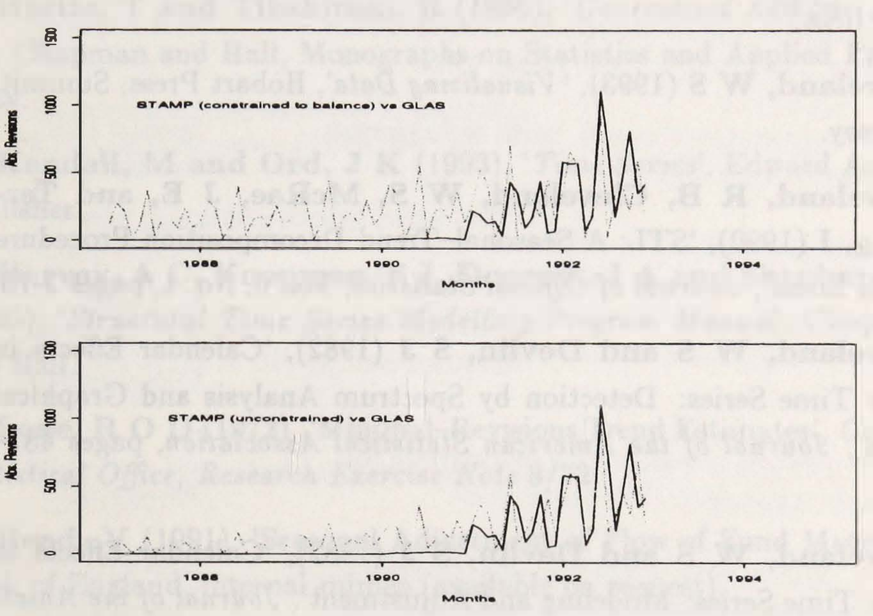

Figure 21: Revisions (in absolute value) obtained for the M4 series when subtracting the seasonally adjusted series obtained when running the seasonal adjustment in the samples 1987:1–1992:12 and 1987:1–1994:12. Note: GLAS (solid line) is taken as the reference method.

References

Chambers, J M and Hastie, T (1993), '*Statistical Models in S*', Chapman and Hall, Computer Science Series.

Cleveland, W S (1979), 'Robust Locally Weighted Regression and Smoothing Scatterplots', *Journal of the American Statistical Association*, pages 829–836.

Cleveland, W S (1983), 'Seasonal and Calendar Adjustment', in D.R. Brillinger and P.R. Krishnaiah editors, *Handbook of Statistics*, pages 39–72, Volume 3, Elsevier Science Publishers, North–Holland.

Cleveland, W S (1983), "Comments on 'Modelling considerations in the seasonal adjustment of economic time series' by S.C. Hillmer, W.R. Bell, and G.C. Tiao", in *Modeling Seasonal Economic Data*, Bureau of Census, USA.

Cleveland, W S (1993), '*Visualizing Data*', Hobart Press, Summit, New Jersey.

Cleveland, R B, Cleveland, W S, McRae, J E, and Terpenning, I (1990), 'STL: A Seasonal–Trend Decomposition Procedure Based on Loess', *Journal of Official Statistics*, Vol. 6, No. 1, pages 3-73.

Cleveland, W S and Devlin, S J (1982), 'Calendar Effects in Monthly Time Series: Detection by Spectrum Analysis and Graphical Methods', *Journal of the American Statistical Association*, pages 487–496.

Cleveland, W S and Devlin, S J (1982), 'Calendar Effects in Monthly Time Series: Modeling and Adjustment', *Journal of the American Statistical Association*, pages 520–528.

Cleveland, W S and Terpenning, I (1982), 'Graphical Methods for Seasonal Adjustment', *Journal of the American Statistical Association*, pages 52–62.

Cleveland, W S, Devlin, S J and Terpenning, I (1981), 'The Details of the SABL Transformation, Decomposition, and Calendar Methods', available from the Computer Information Library, Bell Laboratories, 600 Mountain Ave., Murray Hill, N.J. 07974.

Dagum, E (1988), 'The X-11 ARIMA/88 Seasonal Adjustment Method: Foundations and User's Manual', *Statistics Canada.*

Findley, D F (1983), "Comments on 'Comparative Study of the X-11 and Baysea Procedures of Seasonal Adjustment' by H. Akaike and M. Ishiguro", in "Applied Time Series Analysis of Economics Data", pages 36-45, edited by Arnold Zellner and published by the Bureau of the Census (Washington, D.C.).

Harvey, A C (1989) '*Forecasting, structural time series models and the Kalman filter*', Cambridge University Press.

Hastie, T and Tibshirani, R (1990), '*Generalized Additive Models*', Chapman and Hall, Monographs on Statistics and Applied Probability.

Kendall, M and Ord, J K (1993), '*Time Series*', Edward Arnold publisher.

Harvey, A C, Koopman, S J, Doornik, J A and Shephard, N (1995), '*Structural Time Series Modelling Program Manual*', Chapman and Hall.

Lane, R O D (1972), 'Minimal-Revisions Trend Estimates', *Central Statistical Office, Research Exercise Note* 8/72.

Read, V (1991), 'Seasonal Adjustment of Flow of Fund Matrices', Bank of England, internal mimeo (available on request).

Working Party (1992), 'Report of the Seasonal Adjustment Working Party', *Bank of England, Occasional paper no.* 2, October 1992.

Young, T (1992), 'Seasonal Adjusting of Flow of Fund Matrices', Bank of England, internal mimeo (available on request).

Bank of England Working Paper Series

Publication date in italics

1	Real interest parity, dynamic convergence and the European Monetary System *(June 1992)*	Andrew G Haldane Mahmood Pradhan
2	Testing real interest parity in the European Monetary System *(July 1992)*	Andrew G Haldane Mahmood Pradhan
3	Output, productivity and externalities—the case of banking *(August 1992)*	R J Colwell E P Davis
4	Testing for short-termism in the UK stock market *(October 1992)*	David Miles
5	Financial deregulation and household saving *(October 1992)*	Tamim Bayoumi
6	An investigation of the effect of funding on the slope of the yield curve *(January 1993)*	D M Egginton S G Hall
7	A simple model of money, credit and aggregate demand *(April 1993)*	Spencer Dale Andrew G Haldane
8	Bank credit risk *(April 1993)*	E P Davis
9	Divisia indices for money: an appraisal of theory and practice *(April 1993)*	Paul Fisher Suzanne Hudson Mahmood Pradhan
10	The effect of official interest rate changes on market rates since 1987 *(April 1993)*	Spencer Dale
11	Tax specific term structures of interest rates in the UK government bond market *(April 1993)*	Andrew J Derry Mahmood Pradhan
12	Regional trading blocs, mobile capital and exchange rate co-ordination *(April 1993)*	Tamim Bayoumi Gabriel Sterne
13	Temporary cycles or volatile trends? Economic fluctuations in 21 OECD countries *(May 1993)*	Gabriel Sterne Tamim Bayoumi
14	House prices, arrears and possessions: A three equation model for the UK *(June 1993)*	F J Breedon M A S Joyce

15	Tradable and non-tradable prices in the UK and EC: measurement and explanation *(June 1993)*	C L Melliss
16	The statistical distribution of short-term libor rates under two monetary regimes *(September 1993)*	Bahram Pesaran Gary Robinson
17	Interest rate control in a model of monetary policy *(September 1993)*	Spencer Dale Andrew G Haldane
18	Interest rates and the channels of monetary transmission: some sectoral estimates *(September 1993)*	Spencer Dale Andrew G Haldane
19	The effect of futures trading on cash market volatility: evidence from the London stock exchange *(December 1993)*	Gary Robinson
20	M0: causes and consequences *(December 1993)*	F J Breedon P G Fisher
21	An empirical analysis of M4 in the United Kingdom *(December 1993)*	P G Fisher J L Vega
22	A model of building society interest rate setting *(June 1994)*	Joanna Paisley
23	Deriving estimates of inflation expectations from the prices of UK government bonds *(July 1994)*	Mark Deacon Andrew Derry
24	Estimating the term structure of interest rates *(July 1994)*	Mark Deacon Andrew Derry
25	Potential credit exposure on interest rate swaps *(August 1994)*	Ian Bond Gareth Murphy Gary Robinson
26	New currencies in the Former Soviet Union: a recipe for hyperinflation or the path to price stability? *(September 1994)*	C L Melliss M Cornelius
27	Inflation, inflation risks and asset returns *(November 1994)*	Jo Corkish David Miles

28	The construction of RPIY *(February 1995)*	R Beaton P G Fisher
29	Pricing deposit insurance in the United Kingdom *(March 1995)*	David Maude William Perraudin
30	Modelling UK inflation uncertainty: the impact of news and the relationship with inflation *(April 1995)*	M A S Joyce
31	Measuring core inflation *(April 1995)*	Danny T Quah Shaun P Vahey
32	An assessment of the relative importance of real interest rates, inflation and term premia in determining the prices of real and nominal UK bonds *(April 1995)*	David G Barr Bahram Pesaran
33	Granger causality in the presence of structural changes *(May 1995)*	Marco Bianchi
34	How cyclical is the PSBR? *(May 1995)*	Joanna Paisley Chris Salmon
35	Money as an Indicator *(May 1995)*	Mark S Astley Andrew G Haldane
36	Testing for convergence: evidence from nonparametric multimodality tests *(June 1995)*	Marco Bianchi
37	Wage interactions: comparisons or fall-back options *(August 1995)*	Jennifer C Smith
38	The microstructure of the UK gilt market *(September 1995)*	James Proudman
39	Valuation of underwriting agreements for UK rights issues: evidence from the traded option market *(September 1995)*	Francis Breedon Ian Twinn
40	Rules, discretion and the United Kingdom's new monetary framework *(November 1995)*	Andrew G Haldane
41	Optimal commitment in an open economy credibility vs flexibility *(December 1995)*	Sylvester Eijffinger Eric Schaling

42	Bidding and information: Evidence from Gilt-Edged auctions (*January 1996*)	Francis Breedon Joe Ganley
43	International bank lending to LDCs—an information-based approach (*February 1996*)	Prasanna Gai
44	A comparison of methods for seasonal adjustment of the monetary aggregates *(March 1996)*	Marco Bianchi