Applied Economics Letters, 1999, **6**, 103–109

A comparison of methods for trend estimation

MARCO BIANCHI, MARTIN BOYLE* and DÉIRDRE HOLLINGSWORTH‡

*Bank of England, Monetary Analysis, Division, *Monetary Financial Statistical Division, London, UK and ‡Oxford University, Department of Mathematics, Oxford, UK*

Received 9 July 1997

This paper analyses a number of methods for trend estimation focusing on their ability to pick up turning points quickly at the end of a series. An application to the Bank of England flows M4 series is provided which shows that some of the proposed methods may be more reliable than others for this task.

I. INTRODUCTION

The aim of this report is to present proposals for a method of estimating trends for the Bank's monetary statistics. Trend estimation is a potentially useful technique to aid interpretation of the data and would complement the existing seasonally adjusted statistics. The initial aim of the project has been to discriminate between competing estimates to select a method that would inform internal policy analysis. If the results obtained are judged to add value then there must be a presumption in favour of wider dissemination.

Policy makers and final users of trend figures frequently tend to evaluate trend estimation methods based on a number of characteristics such as the ability to pick up turning points quickly, but also minimizing the risk of false turning points, smoothness, quick convergence to final trend (i.e. trend revisions rapidly declining to zero as more observations become available) and trend unaffected by outliers. There is likely a trade-off between some of these characteristics, but it appears that the majority of policy makers, central statistical offices and central banks are especially concerned with the ability of a short term trend to pick up turning points quickly at the end of the series, minimizing however the risk of picking up false turning points. As this is also the prevailing view within the Bank, trend estimation techniques for quick detection of turning points is the main focus of our paper.

There are many techniques for depicting a trend. For some purposes, relatively simple techniques such as moving averages or scatter diagrams can provide acceptable results. But, where data are volatile, or where the early identification of turning points is critical, it is usually necessary to make use of more sophisticated mathematical techniques. Broadly speaking, trend estimation methods fall into two categories: parametric (or model-based) and nonparametric. We look in this study at the performance of the following nonparametric methods: (i) GLAS weighted moving average filters; (ii) Henderson weighted moving average filters; (iii) LOcally WEighted Scatterplot Smoothers (LOWESS); (iv) Smoothing splines. As for the parametric methods, we only look at the Kalman filter approach. For all the methods, trend estimates are derived from seasonally adjusted series, what is known in the literature as *ex-post* trend smoothing (see Cleveland *et al.*, 1994).

The remainder of this paper is organized as follows. In Section II we give a brief overview on a number of methods for trend estimation. In Section III we evaluate the performance of the methods by focusing on their ability to quickly detect a turning point in the Bank of England M4 flows series. Section IV summarizes and concludes.

II. TREND ESTIMATION METHODS

Let y_t denote a seasonally adjusted series from which a trend (or trend-cycle) unobserved component has to be extracted. We review the different methods of trend estimation in the following subsections.

GLAS (GL)

GLAS stands for 'General Linear Abstraction of Seasonality'. It represents the package currently used at the Bank of England for seasonal adjustment and trend estimation of the monetary series (see Young, 1992). 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). The number of points used in the moving average (denoted by n_t and sometimes called trend window width) governs the 'degree of smoothness' of the trend; thus, increasing n_t (by definition an odd integer number) makes the trend smoother.

To understand how an estimate of the trend at a given point in time, say t_0, is obtained in GLAS, we give a simple illustrative example. Given the trend window width n_t, the set of at nearest neighbour points in time to t_0 (including t_0) is identified. Call this set $N(t_0)$. Define the triangular weighting function

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

where $u = 2(t_0 - t)/n_t$, 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

$$T(t_0) = \sum_{t \in N(t_0)} w_t y_t \tag{2}$$

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 direction.

A feature 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 asymmetric 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 minimized. This 'minimum revision algorithm' employed by GLAS may be in contrast with the objective of quick detection of turning points at the end of the series if there is a trade-off between ability of a trend to minimize revisions and its flexibility to adapt its shape to new data at the end of the series.

Henderson filter (HF)

The Henderson filter used in the X11-ARIMA and X-12-ARIMA packages is also a weighted moving average smoother, but using a weighting pattern different from the triangular one employed by GLAS. Given the trend as

$$T_t = \sum_{k=-m}^{m} w_k y_{t+k} \tag{3}$$

the formula for the symmetric Henderson weights applicable to the kth term is (see Kenny and Durbin, 1982):

$$\begin{aligned} w_k \propto & \{(m+1)^2 - k^2\}\{(m+2)^2 - k^2\} \times \\ & \{(m+3)^2 - k^2\}\{3(m+2)^2 - 16 - 11k^2\} \end{aligned} \tag{4}$$

where the constant of proportionality is chosen to ensure that $\sum_k w_k = 1$. Here, $n_t \equiv 2m + 1$ is the number of terms or trend window width, so that, for example, $m = 6$ for the $n_t = 13$ term moving average.

Contrary to GLAS, where progressively more asymmetric versions of the triangular weighting pattern are used, the Henderson filter employed in this paper is always symmetric due to forecasting and backcasting in X-12-ARIMA.

Lowess (LW)

Lowess identifies a certain number of nearest-neighbours to a given point, x_0, and assigns a weight to each neighbour based on the distance of that neighbour to the point. A value of the trend at x_0 is then calculated based on these weights. The number of nearest neighbours which are used is the smoothing parameter. Again, the bigger the number, the smoother the trend. In fact, the size of neighbourhood governs a fundamental trade-off between bias and variance of the estimator. If a large neighbourhood is used, the trend is very smooth (that is the variance is low), but at the possible cost that it is not flexible enough representation to adapt to the underlying pattern of the data (that is the bias can be high).

The Lowess smoother fitted at a given point is derived by locally averaging the data in a neighbourhood of that point. A polynomial is fitted to the data using (iterative) weighted least squares, with the weights computed according to a 'tri-cube' weight function. The estimator is constructed through the following steps (see Hastie and Tibshirani, 1990, sec. 2.11; Cleveland, 1994, pp. 94–101):

(i) Given the value y, the k nearest neighbours of y are identified, denoted by $N(y)$.

(ii) $\Delta(y) = \max_{N(y)} |y - y_t|$ is computed, the distance of the farthest near-neighbour from y.

(iii) Weights w_t are assigned to each point in $N(y)$, using the so-called 'tri-cube' weight function

$$W\left(\frac{|y - y_t|}{\Delta(y)}\right)$$

where, for any u,

$$W(u) = \begin{cases} (1 - u^3)^3 & \text{for } 0 \le u < 1 \\ 0 & \text{otherwise} \end{cases}$$

(iv) The fitted Lowess curve at y, $\hat{g}(y)$, is the value of a polynomial of dth degree fitted to the data using (iterative) weighted least squares, with the weights computed as in (iii). So, if $d = 1$ the values of a and b, $\hat{a}$ and $\hat{b}$ respectively, are found that minimize

$$\sum_{j=1}^{T} w_t(y)(y_t - a - by_t)^2$$

Then, the fit at r is $\hat{g}(y) = \hat{a} + \hat{b}y$.

The parameter k represents here the smoothing parameter, being the number of nearest neighbour points. It is common practice for Lowess to express this smoothing parameter as the fraction of nearest neighbour points over the total number of points. This leads to the definition of the smoothing parameter $\alpha \equiv k/T$. It is worth noticing that in practical applications α is selected *based on a combination of judgement and of trial and error* (Cleveland, 1994, p. 96) bearing in mind some certain guidelines, such as for example *to produce a fit that is as smooth as possible without unduly distorting the underlying pattern in the data* (Cleveland, p. 98). Thus, even though the choice of parameter is somehow subjective, it must be well suited to the series.

Smoothing splines (SS)

The smoothing spline smoother is derived as the explicit solution to the functional minimization problem

$$\min_{m(z)} \left[\sum_{t=1}^{T} \left[y_t = g(y_t) \right]^2 + \lambda \int_{-}^{+} \left[g''(z) \right]^2 dz \right]$$

where λ represents the smoothing parameter, which is the trade-off between the smoothness of the curve (the second derivative term in the integral) and the fidelity to the data (the residual sum of squares). As discussed by Buja *et al.* (1989), for any given λ the solution of the minimization problem is $g(\mathbf{y}) = TREND = \mathbf{Sy}$ where $\mathbf{y} = (y_1, \ldots, y_t, \ldots, T)'$ and $\mathbf{S}$ is a smoothing matrix given by $\mathbf{S} = (\mathbf{i} + \lambda\mathbf{K})^{-1}$, where $\mathbf{S}$ is a matrix of weights with $\mathbf{K} = \mathbf{D}'\mathbf{C}^{-1}\mathbf{D}$. Denoting by $h_t = y_{t+1} - y_t$, $t = 1, 2, \ldots, T$, $\mathbf{D}$ is a $(T-2) \times T$ tridiagonal matrix with $D_{tt} = 1/h_t$, $D_{t,t+1} = -(1/h_t + 1/h_{t+1})$, $D_{t,t+2} = -(1/h_{t+1})$, while $\mathbf{C}$ is a symmetric tridiagonal matrix of order $T-2$ with $C_{t,t-1} = C_{t,t-1} = h_t/6$, and $C_{tt} = (h_t + h_{t+1}/3$.[1] The trace of the smoothing matrix $\mathbf{S}$ defines the 'degrees of freedom' of the smoothing spline smoother; the bigger the value of the smoothing parameter λ, the higher the number of degrees of freedom.

Kalman filter (KF)

The Kalman filter approach employs the idea of structural time series modelling where the unobserved component of trend is assumed to follow a well-defined stochastic process (Harvey, 1989). A fairly general form for the trend component is given by

$$y_t = T_t + \varepsilon_t \tag{5}$$

with

$$T_t = T_{t-1} + \tau_{t-1} + v_t \tag{6}$$

and

$$\tau_t = \tau_{t-1} + u_t \tag{7}$$

where v and u are zero mean, normally distributed 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.

The variances of the error terms, σ_ε^2, σ_T^2, σ_τ^2, are called the *hyperparameters* of the model. The hyperparameters play here the same role of the smoothing parameters in the context of nonparametric methods. The bigger the value of the variance σ_T^2 relative to σ_ε^2, the more are past observations discounted in constructing the trend pattern for the forecast function. In this sense, we can define $n_t \propto q_T = \sigma_T/\sigma_\varepsilon$, where $q_{\hat{\tau}}$ is called the q-ratio for the trend. If $\sigma_\tau^2 = 0$, a null hypothesis which can be tested from the data, no stochastic slope is specified for the trend.

The appealing feature of the Kalman filter approach is that the smoothing parameter is estimated by maximum likelihood. The smoothing parameter is thus indigenously determined, rather than fixed *a priori*. Given the hyperparameters, the estimates of the trend component, which are obtained by the Kalman filter, are optimal in the sense that they minimize 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.

III. RESULTS

As an application, we consider the Bank of England seasonally adjusted M4 flows series from January 1987 to September 1996 (117 observations). The series was seasonally adjusted using GLAS. Trends are calculated on this series by smoothing splines (SS), Lowess (LW), the Kalman filter (KF), the Henderson filter (HF) and GLAS (GL), which are shown in Fig. 1.[2] For smoothing splines, the smoothing parameter λ was selected based on residual diagnostics plots (see the Appendix).[3] The fitted trends show two turning points about 1989:08 and 1992:11, where two vertical lines are drawn in Fig. 1. It is also worth pointing out the different behaviour of the Lowess smoother, relative to the other methods, in 1995; by its nature, the shape of the Lowess trend

[1] It is worth noting if there are ties in the predictor values the h_j take zero values for some j, what implies the D_{jj} are not computable. As suggested by Hastie and Tibshirani (1990 p. 74), the solution is to create a new data set with $m < T$ observations, each one representing a set, and redefining the values of the response variable as weighted sums. Then, a *weighted* smoothing procedure would be applied to the new data set.

[2] SS and LW were estimated using S-PLUS, see Chambers and Hastie (1993) – for both methods a plot with a 95% pointwise confidence band around the trend is produced; KF using STAMP, see Harvey *et al.* (1995); GL using GAUSS and HF by X-12-ARIMA, see Findley *et al.* 1996.

[3] These indicated a value of 0.0005, which has an equivalent number of degrees of freedom of approximately 8.8 parameters. For a mathematical definition of degrees of freedom in non-parametric regression, see Hastie and Tibshirani (1990). Intuitively, degrees of freedom are interpretable as the number of parameters in a parametric regression model. An equivalent number of degrees of freedom in Lowess was obtained using $\lambda = 0.20$, whereas for the Henderson filter we have used 33 points. In GLAS, the default value was employed (23 points), whereas the Kalman filter was implemented based on maximum likelihood estimates of the hyperparameters.

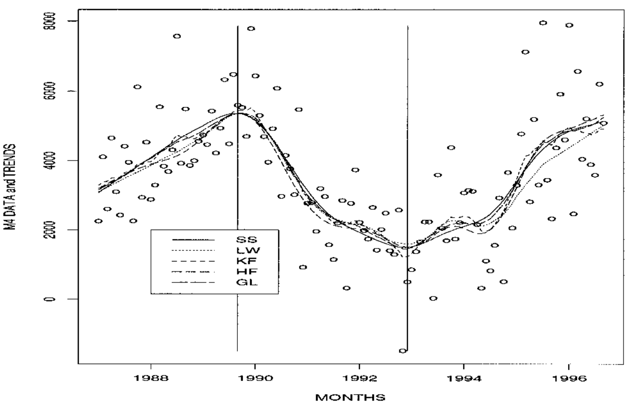

Fig. 1. *Seasonally adjusted M4 series (points) with fitted trends using smoothing splines (SS), Lowess (LW), the Kalman filter (KF), GLAS (GL) and the Henderson filter (HF). Turning points appear in 1989:08 and 1992: 10*

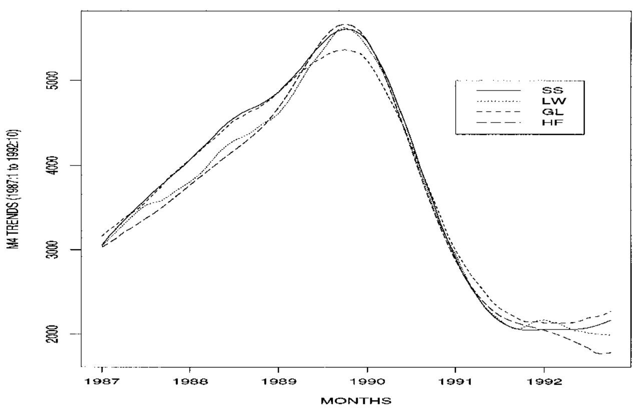

Fig. 2. *Trend fitted by different nonparametric methods on the seasonally adjusted M4 series from 1987:1 to 1992:10*

was not distorted by three high values of the seasonally adjusted M4 series at around the 8000 level.

Given the above information, the series is seasonally adjusted in a number of sequential runs as each month's data were added, starting from October 1992 (i.e. using data from 1987:1 to 1992:10). This seasonal adjustment led to changes throughout the run of data, which then affected the trend. This is what the trend estimation procedure would have to deal with in application, and did not affect the comparison of methods since they were all reacting to the same data. The value of the smoothing parameter for the smoothing spline method is chosen (as before) based on the usual set of diagnostic plots. On the first run of M4 data to October 1992, this led us to choose $\lambda = 0.0005$, the same smoothing parameter which was applied to the whole series. Again, smoothing parameters implying an equivalent number of degrees of freedom (i.e. degree of smoothness of the trend) were selected for Lowess and the Henderson filter, whereas default values were used in GLAS. The results obtained for the nonparametric methods are shown in Fig. 2: here, the trends appear to have a very similar degree of smoothness with GLAS being perhaps a little bit smoother than the other methods.[4] For all these methods the selected smoothing parameter was then kept fixed in the sequential runs. In STAMP, on the other hand, maximum likelihood estimates of the smoothing parameter were calculated in each run.

The results are presented in Fig. 3 which shows how the different trends evolve as each month of data is added around the turning point estimated to have occurred around the end of 1992/beginning of 1993. The Lowess trend, for example, changes from a decreasing to a level trend over the months of October 1992 to February 1993 (see curves a, b, c, d, e). An upward trend is first shown when March 1993 data are added (f curve). This is just a few months after the turning point is considered to have occurred, in November–December 1992. From March 1993 to September 1993, all trend lines consistently display a global minimum, thus reinforcing the evidence for a turning point, with the only exception in June 1993 (i curve) where the trend is decreasing.

The overall shape of smoothing spline trend estimates are similar to Lowess – remember we have used the same amount of smoothing – but with a number of important differences too. In the period where the trend was increasing month on month (see curves h, j, k, l), from June to September 1993, the trend lines are alike. However, where the position of the data points was changing rapidly, the two methods varied in the trend they produced.[5] They gave conflicting information in some months, for example in October 1992 (curve a) and March and April 1993 (curves f and g). The turning point in the beginning of 1992, in particular, was a *false* turning point, as successive observations failed to confirm it as a genuine turning point. The Lowess trend (curve a in the top-right panel of Fig. 3) shows a negative derivative, whereas according to the smoothing spline trend the derivative evaluated at the most recent observations is positive. In this sense, the Lowess method appears to have a letter performance lay (correctly) not detecting the false turning point. In March and April 1993 (curves f and g), the SS trend lines are still decreasing (although at a very small rate), whereas LW trend lines already indicate the presence of a turning point. This appears to indicate an advantage of LW over SS to pick up turning points more quickly.

Compared to SS and LW the KF method is not as smooth. This is expected due to the stochastic nature (random walk

[4] The program that we have used in this paper for GLAS can only fit trends using the default smoothing parameter. But the program should perhaps be generalized to estimate a trend with general smoothing parameter.

[5] These are the areas where the weight given to one data point – not supported by several others – distinguish the two methods most clearly.

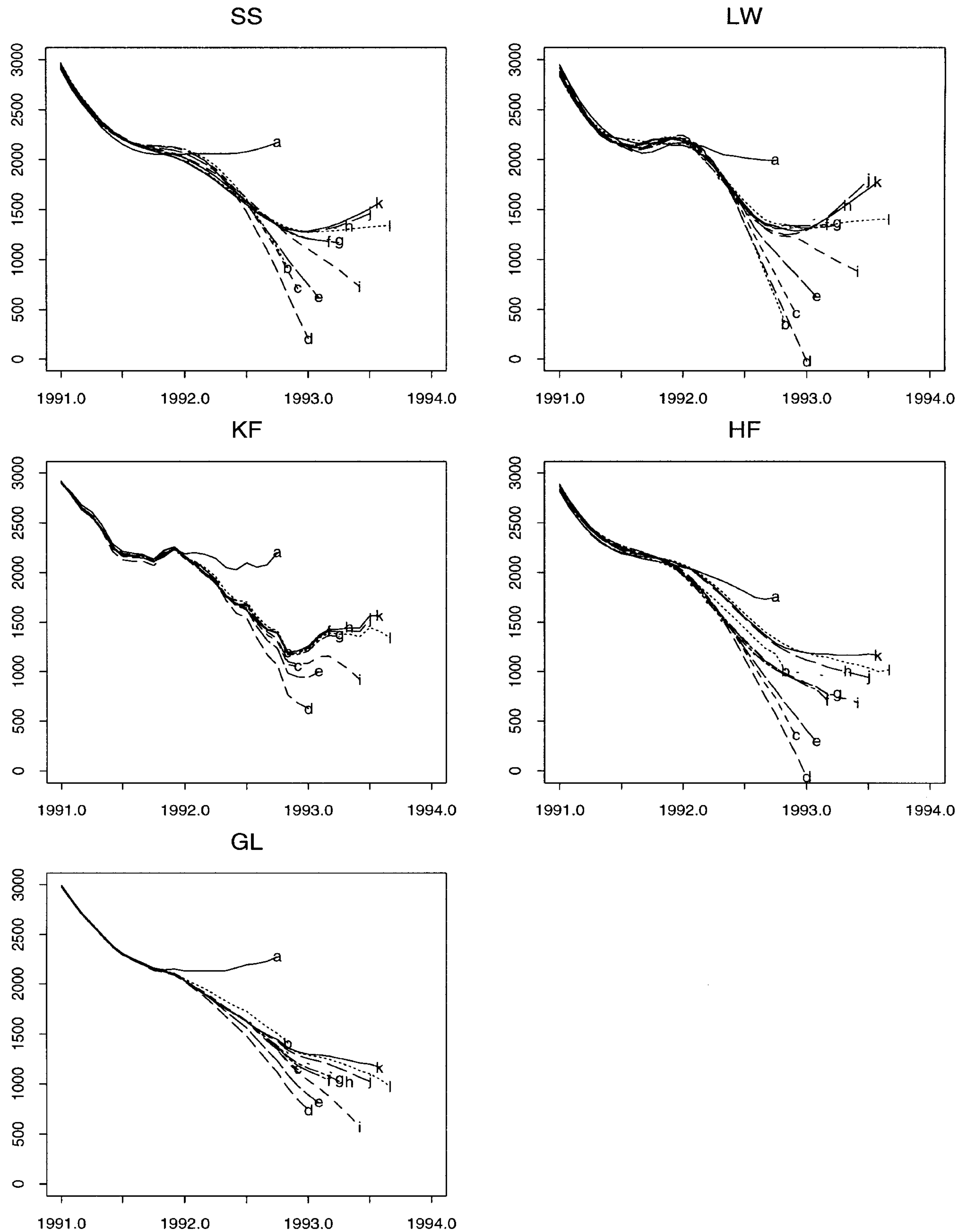

Fig. 3. *Trend estimation results in sequential runs starting from 1992:10 (denoted by letter ' a') until 1993:09 (denoted by letter ' l')*

Table 1. *Sign patterns in the detection of the turning point*

	Oct 92 (a)	Nov (b)	Dec (c)	Jan (d)	Feb (e)	Mar (f)	Apr (g)	May (h)	Jun (i)	Jul (j)	Aug (k)	Sep (l)
GL	+	–	–	–	–	–	–	–	–	–	–	–
HF	–	–	–	–	–	–	–	–	–	–	=	+
SS	+	–	–	–	–	=	=	–	+	+	+	+
LW	–	–	–	–	–	+	+	–	+	+	+	+
KF	+	–	–	–	+	+	+	–	+	+	+	+

Legend: – indicates a decreasing trend; = indicates a flat trend; + indicates an increasing trend. Therefore, a – followed by a + or a – followed by a = and a + indicate a turning point (local minimum in the trend).

process) of the trend component in the KF method. Nevertheless, the method is very flexible and appears quick at picking up turning points – in this case the turning point was found in February 1993, only around 3 months after it actually occurred. The results for the HF and GL methods do not seem to be as good, while they both deliver very smooth trends with small revisions they are totally ineffective at finding the turning point. This is particularly true for the GL method, as by September 1993 there was still no visible evidence that the method had found the turning point. There appears to be a trade off here in that the GL and HF methods have minimal revisions at the price of being unable to detect turning points quickly.

Table 1 summarizes the results in terms of the ability of the various trend methods to detect the turning point estimated (in September 1996) to have occurred in between November 1992 and January/February 1993. The table confirms the above discussion that all methods would have picked up the turning point relatively quickly, with the only exception of the GL and the HF methods, the worst performers of all our methods. The quickest method would have been the KF method, which appears to detect the turning point at the fifth run, in February 1993; at that point in time no evidence of a turning point is signalled by the SS and the LW methods. The SS trend shows some weak evidence for a turning point starting from runs f and g, which becomes stronger only at run j, in July 1993.

IV. CONCLUSIONS

We have examined in this paper a number of methods for trend estimation by applying the methods to the task of detecting a turning point in the Bank of England M4 flows series, located around the end of 1992. The results of our analysis have indicated that methods like the Lowess smoother and the Kalman filter smoother would have captured rather well (that is with only a few months delay) the turning point. The former method would have also produced a smoother trend curve. The smoothing spline method would have led to similar results but its performance was somehow inferior to the previous two. Weighted moving average methods such as GLAS and the Henderson filter have delivered the worst performance overall by failing to detect the turning point even after a member of periods. Of course, these conclusions are specific to our M4 series with a, turning point in November/December 1992; further work would require to repeat the analysis of this paper to various series before more general conclusions about the methods can be drawn.

ACKNOWLEDGEMENT

We thank Jackie Knowles for useful comments and suggestions.

REFERENCES

Chambers, J.M. and Hastie, T. (1993) *Statistical Models in S*, Chapman and Hall, Computer Science Series, London.

Cleveland, W.S. (1979) Robust locally weighted regression and smoothing scatterplots, *Journal of the American Statistical Association*, 829–36.

Findley, D.F., Monsell, B.C., Bell, W.R., Otto, M. C. and Bor-Chung, C. (1996) New capabilities and methods of the X-12-ARIMA seasonal adjustment program, Bureau of the Census, Washington DC, USA.

Harvey, A.C. (1989) *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press, Cambridge.

Hastie, T. and Tibshirani, R. (1990) *Generalized Additive Models*, Chapman and Hall, Monographs on Statistics and Applied Probability, London.

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

Kenny, P.B. and Durbin, J. (1982) Local trend estimation and seasonal adjustment of economic and social time series, *Journal of the Royal Statistical Society*, Series A, **145**, 1–41.

Lane, R.O.D. (1972) Minimal-revisions trend estimates, *Central Statistical Office*, *Research Exercise Note* 8/72.

Young, T. (1992) Seasonal adjusting of flow of fund matrices, Bank of England, internal mimeo (available on request).

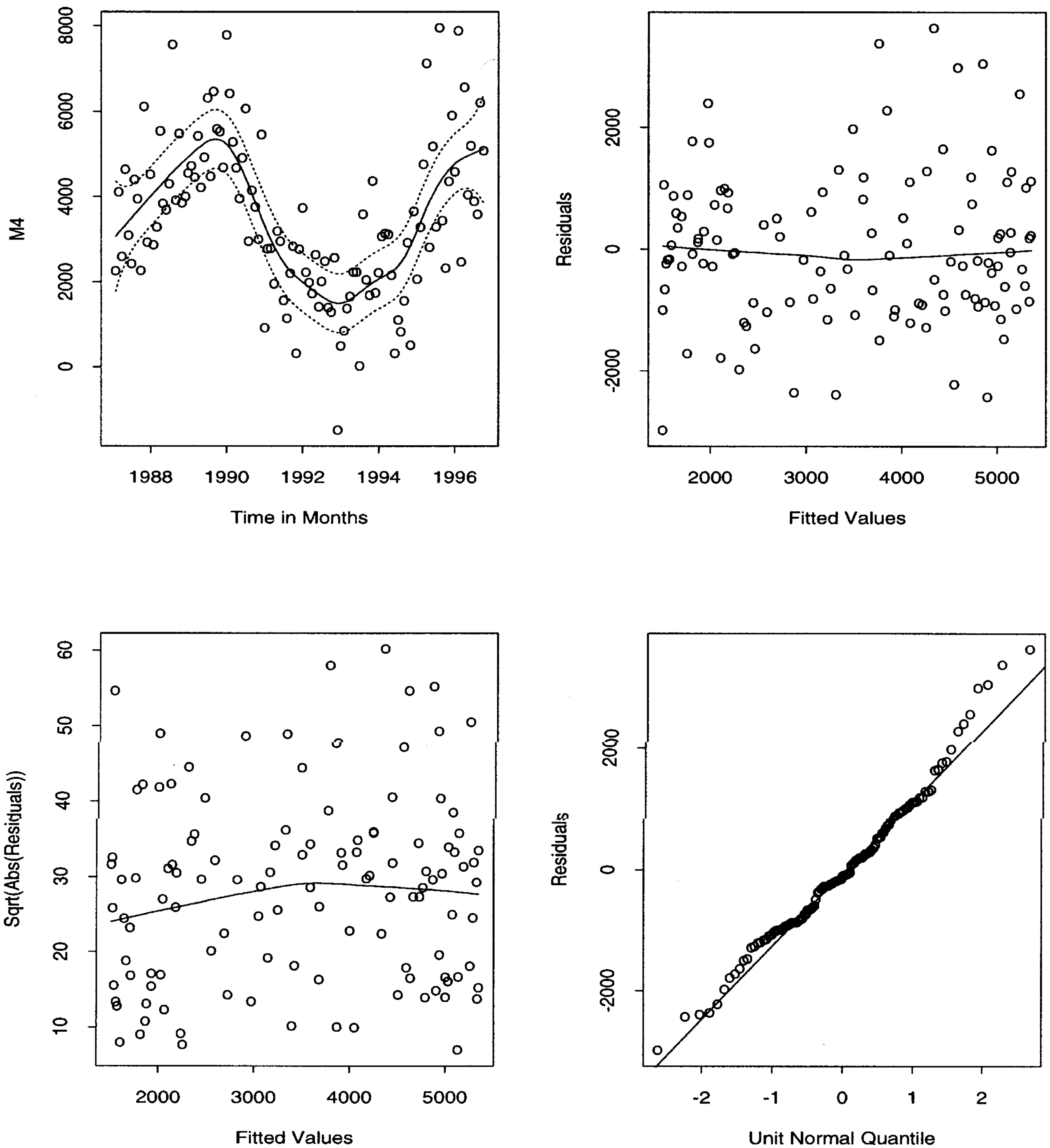

Fig. A1. *Trend estimation on the seasonally adjusted M4 series from 1987.01 to 1996.09 using smoothing splines with* $\lambda = 0.0005$ *(top left panel) and diagnostic plots to check for zero-mean (top right panel), constant variance (bottom left panel) and normally distributed (bottom right panel) residuals. Remark: some heteroscedasticity, as well as non-normality, is apparent in the residuals*