Appl. Statist. (1976),
25, *No.* 3, p. 195

Comparison of Forecast and Actuality

By G. E. P. Box and G. C. Tiao

University of Wisconsin, Madison, U.S.A.

[Received February 1975. Revised November 1975]

SUMMARY

The paper shows how possible change in a system generating a time series may be studied by comparing forecasts made from a model built on data prior to the suspected change with data actually occurring. An environmental example illustrates the decomposition of the overall criterion into relevant components and shows how difficulties can occur in distinguishing alternative models for change. The relation to surveillance problems and to intervention analysis is briefly discussed.

Keywords: FORECAST; PARAMETER CHANGES; SURVEILLANCE; INTERVENTION ANALYSIS; OZONE LEVEL; POLLUTION CONTROL

1. Introduction

Suppose a system has been subjected to a change. A natural way to consider the possible effect of that change is to compare, with actuality, forecasts made from a stochastic model which was appropriate before the change. Such a model determines the probability structure of its forecast errors and a specified change in the model results in a calculable change in that structure. Hence, appropriate functions of the forecast errors may be calculated which point to specific changes in the model.

The object of this paper is to illustrate this, using some environmental data. One point that emerges is that changes in the model of different kinds may not be easily distinguishable.

2. A Time Series Model for the Ozone Data

Following notation and methodology used for example in Box and Jenkins (1970), denote a time series by the sequence $\ldots z_{t-1}, z_t, z_{t+1}, \ldots$. Also define a *white noise* series $\ldots a_{t-1}, a_t, a_{t+1}, \ldots$ as a sequence of independently and normally distributed random *shocks* with mean zero and variance σ^2. The values z_t of the time series are then supposed to be generated by the *linear filtering* operation

$$z_t = \psi(B)\, a_t, \tag{1}$$

where $\psi(B) = 1+\psi_1 B+\psi_2 B^2+\ldots$ and B is the back shift operator such that $Ba_t = a_{t-1}$. Alternatively, the model may be written in the form $\pi(B)\, z_t = a_t$, where

$$\pi(B) = \psi^{-1}(B) = 1-\pi_1 B-\pi_2 B^2-\ldots.$$

For 180 successive values from January 1956 to December 1970 of the monthly average atmospheric ozone concentration at Azusa, California, Tiao *et al.* (1975), obtained a model with

$$\psi(B) = \frac{(1-\theta_1 B)(1-\theta_2 B^{12})}{1-B^{12}}, \quad \theta_1 = -0{\cdot}15,\ \theta_2 = 0{\cdot}91 \quad \text{and} \quad \sigma^2 = 1{\cdot}00. \tag{2}$$

This model was used to produce minimum mean square error forecasts (all from the origin December 1970) for the next 24 months. They are compared in Fig. 1 with what actually happened.

The comparison is of interest because new automobile emissions standards were introduced at the end of 1970. The diagram gives the impression that after 1970 the levels were lower than expected, but caution is needed because (i) forecasts are subject to error, (ii) for forecasts

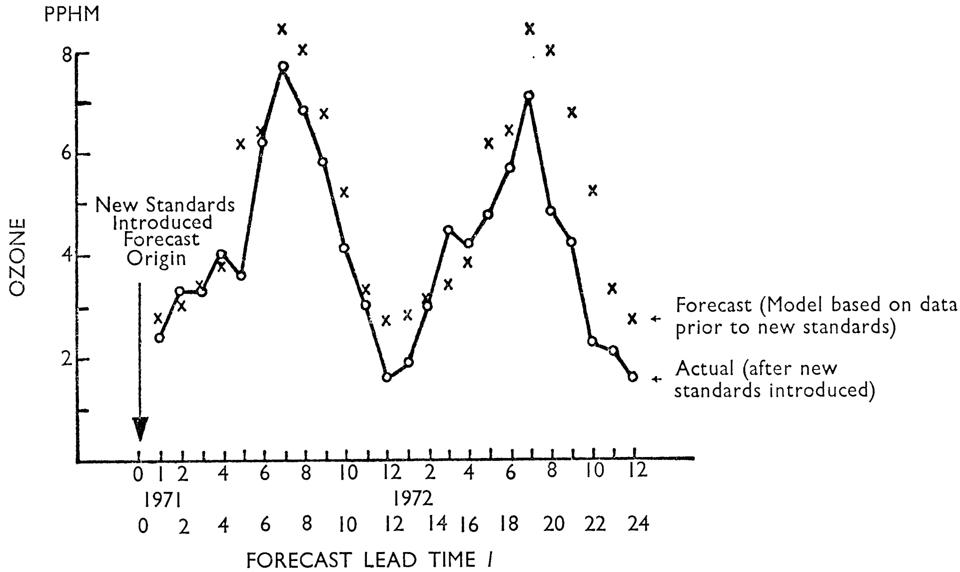

FIG. 1. Forecasts made in December 1970 of ozone concentration at Azusa, California, compared with actuality.

made from the same origin, successive forecast errors for the model in (2) are necessarily highly positively correlated giving a false impression of the consistency of discrepancies.

3. AN OVERALL CHECK

The minimum mean square error forecast of z_l made at an origin conveniently taken to be zero is denoted by $\hat{z}(l)$, where $l = 1, 2, \ldots$, is called the lead time. It is readily shown that the lead l forecast error $e_l = z_l - \hat{z}(l)$ is given by

$$e_l = \sum_{j=1}^{l} \psi_{l-1} a_j, \tag{3}$$

where $\psi_0 = 1$. Now, for the forecasts of $z_1, \ldots, z_m$, writing $\mathbf{a}' = (a_1, \ldots, a_m)$ and $\mathbf{e}' = (e_1, \ldots, e_m)$ the transformation from random shocks to forecast errors is $\mathbf{e} = \boldsymbol{\psi}\mathbf{a}$, where $\boldsymbol{\psi}$ is an $m \times m$ lower triangular matrix with diagonal elements equal to unity, first subdiagonal elements equal to ψ_1, second subdiagonal element equal to ψ_2, and so on. Conversely, $\mathbf{a} = \boldsymbol{\pi}\mathbf{e}$ where $\boldsymbol{\pi} = \boldsymbol{\psi}^{-1}$, and it is readily confirmed that $\boldsymbol{\pi}$ is an $m \times m$ lower triangular matrix with diagonal elements equal to unity and the jth subdiagonal elements equal to $-\pi_j$, $j = 1, \ldots, m-1$.

Now the $m \times m$ covariance matrix for the vector $\mathbf{e}$ is $\mathbf{V} = E(\mathbf{ee}') = \boldsymbol{\psi}\boldsymbol{\psi}'\sigma^2$. It follows that if the original model is appropriate during the period $l = 1, \ldots, m$, then $Q = \mathbf{e}'\mathbf{V}^{-1}\mathbf{e}$ is distributed as χ^2 with m degrees of freedom, where $\mathbf{V}^{-1} = \boldsymbol{\pi}'\boldsymbol{\pi}/\sigma^2$. If, on the other hand, the model changes in some way we may expect that Q will be inflated.

Now rather than compute Q from the e_l, it is easier to employ the identity

$$Q = \mathbf{e}'\mathbf{V}^{-1}\mathbf{e} = \mathbf{e}'\boldsymbol{\pi}'\boldsymbol{\pi}\mathbf{e}/\sigma^2 = \mathbf{a}'\mathbf{a}/\sigma^2 = \sigma^{-2}\sum_{l=1}^{m} a_l^2. \tag{4}$$

Thus, Q is the standardized sum of squares of the one-step ahead forecast errors, $a_1, \ldots, a_m$, and as we suggested in our joint paper with Hamming, an overall test of the continuing appropriateness of the model during the period $l = 1, \ldots, m$ is achieved by referring Q to a χ^2 table with m degrees of freedom. Further, this is equivalent to the appropriate test applied to all the lead l forecast errors e_l, $l = 1, \ldots, m$.

Since in practice σ^2 is estimated from n data values to which, say, p parameters have been fitted, a closer approximation might refer $\hat{Q}/m$, where $\hat{Q} = \hat{\sigma}^{-2} \sum_{l=1}^{m} a_l^2$, to an F table with m and $n-p$ degrees of freedom. However, when n is large, this refinement would make little difference to the result which is in any case approximate since it does not allow† for errors of estimates of the parameters of the original series.

For the ozone data and with the model fitted to the 180 observations $z_{-180}, z_{-179}, \ldots, z_{-1}$, occurring before 1971, we find that $\hat{Q} = 36{\cdot}0$ which is close to the 5 per cent value of χ^2 with 24 degrees of freedom and suggests that the deviations from the forecast are real.

4. Components of χ^2

The quantity Q provides an overall criterion having, like all overall criteria, the advantage that it is unnecessary to be specific about the nature of the feared discrepancy, but the disadvantage that it lacks sensitivity when compared with a more specific criterion which *assumes* that we have guessed correctly what to be afraid of. We now illustrate how the Q statistic may be decomposed into components associated with various relevant alternatives.

4.1. *Relevant Alternatives*

The implication of the overall χ^2 test is that after 1970 the model has changed in some way or other. In speculating on *how* it has changed, we must on the one hand confront theoretical explanations (generated by knowledge of the system) with the data, and on the other allow examination of the data, and particularly of residuals, to suggest theoretical explanations. In particular, it becomes important to know the species of basic patterns which different kinds of discrepancies would inject into the residuals.

"Off the cuff" conjectures would certainly include (a) a change in level of z_t, (*b*) a change in one or both of the stochastic parameters θ_1 and θ_2. However, those having knowledge of the chemistry and meteorology involved pointed out that it was only during the "summer" months (June–October) that the new emission standard would be expected to make much difference; also, it was known that the number of cars fitted with the new control devices would be roughly twice as high in the second year as in the first. This suggested that we might expect (c) a shift in level of z_t in the summers only, with the shift in the second summer about twice that in the first. For reference purposes we refer to this last possibility as the "Met model".

Possibilities (a) and (c) above can be allowed for by substituting $z_t - \sum_{j=1}^{k} \beta_j x_{jt}$ for z_t in the model (1), where β_j are parameters and x_{jt} are appropriate indicator variables. In all cases, x_{jt} would be zero before $t = 0$. For example, to accommodate model (a) we can set $k = 1$, $x_{1l} = 1$, $l = 1, \ldots, 24$; and to accommodate model (c) we can set $k = 2$ with $\beta_1 = 0$ and $x_{2l} = 1$, for $l = 6, 7, 8, 9, 10$, $x_{2l} = 2$ for $l = 18, 19, 20, 21, 22$ and $x_{2l} = 0$ elsewhere. The same general device can, of course, be used to model other changes believed to affect z_l directly such as time trends and new exogenous variables. Allowing these possibilities our model can be written

$$a_l = \pi(B)\left(z_l - \sum_{j=1}^{k} \beta_j x_{jl}\right). \tag{5}$$

To determine the effect of changes in the parameters of the stochastic model, suppose that, prior to time zero, $\pi(B) = \pi_0(B)$ and the stochastic parameters included in the model had

† To gain some idea of how close the approximation is, a brief investigation of this source of error was made for an autoregressive process of order p. With estimates based on n initial observations, estimation errors inflate the mean value of χ^2 calculated from (4) by a factor approximating $1 + (p/n)$.

values $\theta_{10}, \ldots, \theta_{r0}$. Further, suppose that using these values the calculated shocks were $a_{0,1}, \ldots, a_{0,m}$.

Then, after expansion, approximately

$$a_{0,l} \doteq \sum_{j=1}^{k} \beta_j X_{jl} + \sum_{i=1}^{r} (\theta_i - \theta_{i0}) W_{il} + a_l, \tag{6}$$

where

$$X_{jl} = \pi_0(B)\, x_{jl}, \quad W_{il} = \left.\frac{-\partial a_l}{\partial \theta_i}\right|_0.$$

The values of $X_{1l}, X_{2l}, W_{1l}, W_{2l}$ corresponding to β_1, β_2, θ_1 and θ_2 for the ozone example are plotted in Fig. 2 together with the values of $a_{0,l}$. The speculations mentioned above imply that the residuals $a_{0,l}$ contain a deterministic component proportional either to X_{1l}, X_{2l}, W_{1l} or W_{2l} or to some linear combinations of them.

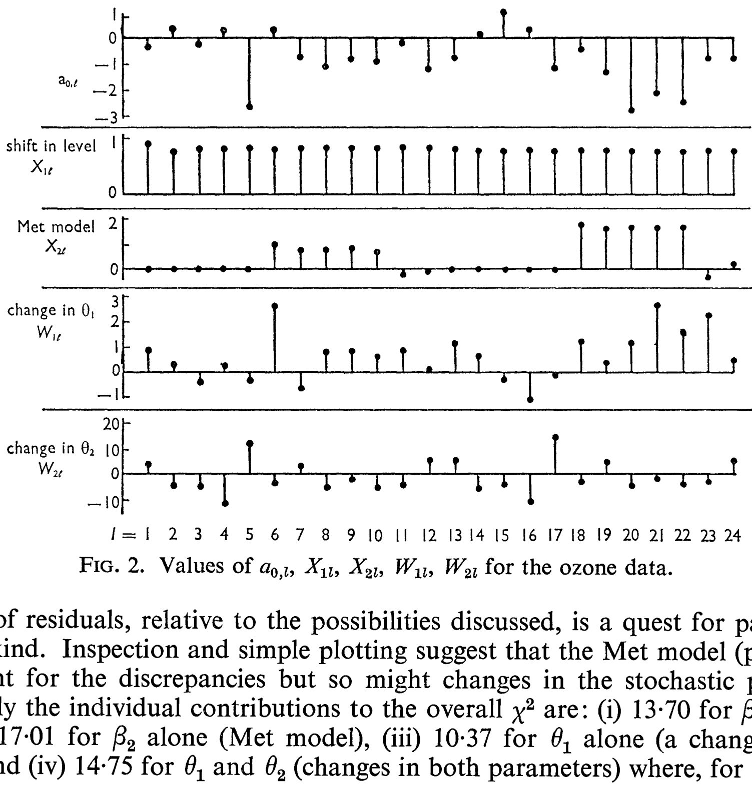

FIG. 2. Values of $a_{0,l}$, X_{1l}, X_{2l}, W_{1l}, W_{2l} for the ozone data.

Analysis of residuals, relative to the possibilities discussed, is a quest for patterns of the appropriate kind. Inspection and simple plotting suggest that the Met model (possibility (c)) might account for the discrepancies but so might changes in the stochastic parameter θ_1. More precisely the individual contributions to the overall χ^2 are: (i) 13·70 for β_1 alone (level change), (ii) 17·01 for β_2 alone (Met model), (iii) 10·37 for θ_1 alone (a change of the first parameter) and (iv) 14·75 for θ_1 and θ_2 (changes in both parameters) where, for example,

$$13{\cdot}70 = \left(\sum_{l=1}^{24} a_{0,l} X_{1l}\right)^2 \Big/ \left(\sum_{l=1}^{24} X_{1l}^2 \sigma^2\right).$$

If we fit equation (6) by least squares, we have

$$\hat{\beta}_1 = -0{\cdot}43, \quad \hat{\beta}_2 = -0{\cdot}71, \quad \theta_1 - \theta_{10} = -0{\cdot}11, \quad \theta_2 - \theta_{20} = -0{\cdot}07.$$
$$(0{\cdot}24) \qquad (0{\cdot}24) \qquad (0{\cdot}18) \qquad (0{\cdot}02)$$

From the analysis of the overall χ^2 given in Table 1 it may be concluded that:

(i) much of the forecast discrepancy can be explained by the Met model associated with β_2 alone;

(ii) after the contribution of that model is taken account of, additional contributions from level shift β_1 and shift in θ_1 are not substantial;

(iii) there is, however, evidence of a slight but significant shift in θ_2.

In connection with this last component it must be remembered that the "Met model" given here is rather rough. Ideally it should take more accurate account of the increase in cars fitted with new engines, and also of the differential effectiveness of the new engines at different seasons. These inadequacies could produce pseudo-seasonal components and hence the apparent need for a slight change in θ_2 which was found.

TABLE 1

Analysis of χ^2 showing contributions of possible discrepancies

Source	*d.f.*	χ^2
Due to β_2 (Met model)	1	17·0
Extra for β_1 (Level shift)	1	2·5
Extra for shift in θ_1	1	0·7
Extra for shift in θ_2	1	5·0
Remainder	20	10·8
Total	24	36·0

A possibility which seems most worrisome to many users of time series analysis, and tends to come to mind first, is that the stochastic parameters determining the "memory" of the system may change by amounts sufficient to cause serious inaccuracies in forecasts. Now it should be noted that if the only possibility tested in the above example had been that θ_1 and θ_2 might have changed, the large contribution of 14·75 to χ^2 might have seemed totally convincing, and also θ_1 would have been identified as the major contributor. The above analysis cannot, of course, prove that a change in θ_1 has not occurred. It does show, however, that if level changes in z_l of a kind that makes practical sense are taken account of, the evidence that a change in θ_1 is needed, vanishes.

This is, of course, because the possibilities considered are closely related. Changes in level of z_l associated with such variables as X_{1l} and X_{2l} will obviously produce serial correlation in $a_{0,l}$. But consider also the residual component produced by a change in the parameter θ_1. This is

$$W_{1l} = -\left.\frac{\partial a_l}{\partial \theta_1}\right|_0 = -(1-\theta_{10}B)^{-1}\,a_{0,l-1}.$$

Since θ_{10} is small in magnitude, the corresponding contribution to χ^2 is roughly

$$\left(\sum_{l=2}^{m} a_{0,l}\,a_{0,l-1}\right)^2 \Big/ \left(\sum_{l=2}^{m} a_{0,l-1}^2\right)\sigma^2 = \sigma^{-2}\sum_{l=2}^{m} a_{0,l-1}^2\, r_1^2,$$

where r_1 is the sample autocorrelation at lag 1 of the $a_{0,l}$. Thus, changes in level of z_t as well as changes in θ_1 are detected by the existence of serial correlation in the residuals. It follows that, although in this example the residuals $a_{0,l}$ are obviously serially correlated, this need not mean that θ_1 has changed.

Investigations as to whether, where and how often changes in stochastic parameters might occur *in real series* are not easy to carry out and little seems to have been done. One result that is clear, however, from the above discussion is that in any such investigation all plausible possibilities ought to be considered and we should not jump too hastily to the conclusion that shifts are necessarily due to changes in stochastic parameters.

5. Relation to Surveillance of Forecasting Systems

The ozone data used for illustration are here presented in the manner which we first received them—complete in December 1972. The problem we have discussed is, however, related to, but is different from, the problem of *sequential* surveillance of routine forecasting schemes. In that problem, data in the form of one-step ahead forecast errors from an operating scheme are available sequentially and a continuous monitoring is carried out to detect possible changes in the model. An early suggestion was that the a_t be plotted on a cumulative sum chart. This usually provides a good catch-all criterion and for the present data such a chart points strongly to the possibility of change although not, of course, to its specific nature. The essential idea involved in the CUSUM chart is the sequential plotting of the likelihood ratio statistic. If a change in *level* of the a_t themselves is being sought this leads at once to the criterion of the cumulative sum. However, as was pointed out, for example by Box and Jenkins (1966), for other alternatives such as a change in levels of one or more of the stochastic parameters, different cumulative plotting methods are appropriate. Some of these are discussed by Ledolter (1975) in a recent thesis.

6. Relation with Intervention Analysis

In our earlier work (1965, 1975) methods were described for estimating the effect of an "intervention" at a known point in a time series.

This situation differed from that discussed here in that the parameters of the time series model were estimated from substantial quantities of data available after, as well as before, the intervention.

However, these earlier methods can perfectly well be applied to examples like the present one and the results will be essentially similar if the period after the intervention is short. We believe the present procedure is worth separate consideration because of its simplicity and intuitive appeal. It is very natural to learn about a system by comparing a set of forecasts made at some point of possible change with actuality.

Acknowledgement

This research was supported in part by a grant from the American Petroleum Institute and in part by U.S. Army Research Office under Grant DA-ARO-D-31-124-72-G162.

References

Box, G. E. P. and Jenkins, G. M. (1966). Models for prediction and control. VI. Diagnostic checking. Tech. Report No. 99, Department of Statistics, University of Wisconsin, Madison.

—— (1970). *Time Series Analysis, Forecasting and Control.* San Francisco: Holden-Day.

Box, G. E. P. and Tiao, G. C. (1965). A change in level of a non-stationary time series. *Biometrika*, **52**, 181–192.

—— (1975). Intervention analysis with applications to economic and environmental problems. *J. Amer. Statist. Ass.*, **70**, 70–79.

Ledolter, J. (1975). Topics in time series analysis. Ph.D. Thesis, Department of Statistics, University of Wisconsin, Madison.

Tiao, G. C., Box, G. E. P. and Hamming, W. J. (1975). Analysis of Los Angeles photochemical smog data: a statistical overview. *J. Air Pollution Control Ass.*, **25**, 260–268.