Letters

ARIMA models in public health surveillance

Sir,

Jiang et al.[1] have recently published an interesting paper in which time-series analysis, autoregressive integrated moving average (ARIMA) models in particular, was used for public health surveillance, specifically for the surveillance of Guillain–Barré syndrome in Sweden.

Nowadays, it is very unusual to find an application of time-series analysis in scientific journals and even more unusual for ARIMA models to be used in public health research. Thus, it is disheartening that by using an inaccurate model-building process the authors failed to transmit precisely the potential of such types of model.

As is known[2] the strategy for constructing ARIMA models is based on a three-step iterative cycle of i) model identification, ii) model estimation and iii) diagnostic checks on model adequacy. We regret that the authors did not care to check the models, that is determine whether or not the chosen models adequately represented the given set of data. Unfortunately, there are inadequacies that might suggest alternative models and, therefore, might invalidate the consequent inferences drawn by the authors.

The authors did not use all nor the most important diagnostic checks. In fact, the only tool they used was the statistical significance of the model parameters, as they pointed out as follows, that "in certain cases, the models were modified (e.g. deleting insignificant parameters) by reidentifying and/or respecifying them" (p. 168). There are at least two other main diagnostic tools: the checking of the stationarity and invertibility of the model and the checking of the white noise behaviour of the residuals. Although the authors did not provide ways for checking the latter, it is easy to check how most of the fitted models were in fact not invertible. As one can see from table 1 (p. 200), it is not possible to reject the null hypothesis that the regular moving average parameters of all the models were equal to one, that is all the models were over-differentiated. In fact, it is possible that the series do not present actual trends but perhaps other systematic patterns that induce non-stationarities (see figures 1 and 2, p. 199 and figure 3, p. 200).

The consequence of such inadequacies is that the fit of the models is in general very poor, in spite of the authors' opinion that "in general, the predicted values for 1993 fitted well with the observed figures" (p. 197). One can derive from table 2 (p. 201) that the mean absolute percentage error ranges from 25.75% (Sweden, incidence per 100,000 ≥40 years) to 84.20% (Sweden, incidence per 100,000 <40 years). The authors should have tried other models before drawing any conclusion.

Summing up, we think the authors lost a great opportunity, that of giving to ARIMA models the position they should have in public health research.

Marc Saez, Maria Antònia Barceló, Department of Economics, University of Girona, Campus de Montilivi, 17071 Girona, Spain, tel. +34 972 418736, fax +34 972 418032, e-mail: msaez@gnomics.udg.es

References

1 Jiang GX, Cheng Q, De Pedro-Cuesta J. Basis for public health surveillance of Guillain–Barré syndrome in Sweden. Eur J Public Hlth 1998,8(3):197–202.

2 Granger CWJ, Newbold P. Forecasting economic time series, 2nd edn. Orlando, FL: Academic Press, 1996.

ARIMA models in public health surveillance: reply to readers' comments

Sir,

We appreciate the interest in our article which was published recently in this journal.[1]

As stated in our paper (p. 198), autoregressive integrated moving average (ARIMA) models in our analyses were constructed by using the 'Statistical Software for Public Health Surveillance' (SSS1)[2] and following the steps of stabilisation of the series mean and variance, model identification, parameter estimation, diagnostic checking and forecast. Reidentifying a model indicates that the new model has to be re-estimated and rechecked.

The concept of stationarity is the basis of the Box–Jenkins model-building methodology. A non-stationary time-series being analysed must first be transformed into one that is stationary in order to identify a model.[2–4] The tools and methods used to achieve and determine the stationarity were described in the paper.

The residual diagnostics are used to explain the general behaviour of the model, determine model adequacy and determine how to improve the model. Statistics and analysis used in SSS1 to assess the model residuals are i) the residual plot, ii) residual mean, iii) the autocorrelation function (ACF) and the partial autocorrelation function (PACF) of the residual, iv) model fit, v) Q statistics to test the hypothesis of the model adequacy, and vi) closeness of fit statistics.[2–4] The results from these methods supported the adequacy of all our models. For example, the residual means of all models were close to zero, the t-values of the residual means were small and the numbers of negative and positive residuals as well as zero crossings were similar *(table 1)*. The χ^2 probabilities from Q statistics in all models were much larger than 0.05, which indicated that the model residuals were random (uncorrelated) and the models were adequate.

We had tried and compared many different models for each series before we decided to choose the reported models. As is known, more than one form of ARIMA model may adequately fit a given time-series. For example, for the time-series for Sweden, all ages and Sweden, ≥40 years, the forecast pattern was similar from the two models ARIMA (0,1,1)*(0,1,1) and ARIMA (0,0,0)*(0,1,1). The latter was a pure

Table 1 Residual diagnostics for ARIMA models constructed on the time-series of Guillain–Barré syndrome

Residual diagnostics	Sweden All ages	Sweden ≥40 years	Sweden <40 years	Stockholm All ages
Residual mean	–0.004284	–0.018057	–0.005359	–0.016823
SE of the residual mean	0.017118	0.025243	0.24642	0.050819
t-value of the residual mean	–0.250238	–0.715318	–0.217491	–0.331033
Number of negative residuals	79.0	74.0	79.0	53.0
Number of positive residuals	76.0	81.0	76.0	60.0
Number of zero crossings	85.0	80.0	83.0	52.0