Journal of Forecasting, Vol. 13, 133–140 (1994)

Discussion
Second-generation Time-series Models: A Comment on 'Some Advances in Non-linear and Adaptive Modelling in Time-series Analysis', by Tiao and Tsay

DANIEL PEÑA
Department of Statistics and Econometrics, Universidad Carlos III de Madrid

INTRODUCTION

This is an excellent paper and I have enjoyed reading it very much. It blends together many of the components of what can be considered the second generation of time-series models: non-linearity, Bayesian inference of changing parameters using the Gibbs sampling, outlier models, long-memory series and adaptive forecasting. These techniques are illustrated by means of a thought-provoking analysis of three economic time series. I draw three main conclusions from this paper: (1) it is very useful to compare models by looking at their out-of-sample forecast performance for different lags; (2) Bayesian inference using Gibbs sampling is a very promising tool for time-series analysis; (3) we should learn more about the relationship between outlier analysis and non-linear time-series modelling. Also, each example suggests some specific comments, as follows.

THE GNP DATA

The authors fit a linear AR(2) to the rate of growth of real GNP. Then they find that a four-regime TAR model produces better forecast and also leads to a better understanding of the data.

In order to check if the parameters of the linear AR(2) model are changing over time, I split the sample into two periods: the first from observation 1 to 85, and the second from 86 to 177. Then I estimated an AR(2) model for each period and built an analysis of covariance-type F-statistics. The F-value is 1.34 and therefore there is no evidence of parameter changing over time. Then I checked the data robustness of the AR(2) model by computing the univariate influence measure suggested by Peña (1990). The most influential points are the 13th ($D = 0,36$) and 19th ($D = 0,29$), and given the small values of the statistic the model seems to be robust to the data. The outlier identification approach by Chen *et al.* (1991) identifies point 13 as an innovational outlier and point 134 as an additive outlier. In order to check the effect of the 13th point I interpolated it assuming that it is an innovational outlier by using intervention

CCC 0277–6693/94/020133–08

analysis (Peña and Maravall (1991)). The estimated model was

$$(1 - 0.37B - 0.12B^2)y_t = 0.0037 + 0.0390 I_t^{(13)} + a_t \tag{1}$$

where $I_t^{(13)} = 1$ for $t = 13$ and $I_t^{13} = 0$ otherwise. When checking the residuals from this model points 97 and 134 now appeared as additive outliers. When the three points are assumed to be missing and interpolated as additive outliers, the second AR coefficient became not significant ($t = 0.42$) and the following model was obtained:

$$\tilde{y}_t = 0.0041 + 0.46\tilde{y}_{t-1} + a_t \tag{2}$$

where $\tilde{y}(13) = y(13) - 0.0311$, $\tilde{y}(97) = y(97) - 0.0281$; $\tilde{y}(134) = y(134) + 0.0298$. As model (2) is very similar to the one obtained by the authors for Regimes III and IV in the TAR model, we conclude that these three points determine the order of the AR linear model for the complete period. Also, this analysis shows the relationship between outliers and non-linear modelling: the three points that are outliers in the linear model are accommodated in the TAR model, with a corresponding reduction in the residual sum of squares.

The TAR model presents two interesting features: (1) it is non-stationary, because regime I has an explosive root, (2) it offers an interesting interpretation of the asymmetric behaviour of the real GNP during recession and expansion periods. On the other hand, the linear AR(2) model leads to the following forecasting equation for $x_t = \log \mathrm{GNP}_t$:

$$\hat{x}_t(l) = A_0^{(t)} + A_1^{(t)}(0.56)^l + A_2^{(t)}(-0.23)^l + 0{\cdot}0076l \tag{3}$$

where $A_0^{(t)}$, $A_1^{(t)}$, $A_2^{(t)}$ are constants that depend on the origin of the forecast, 0.56 and -0.23 are the roots of the AR(2) polynomial and 0.0076 is the slope, that is, very approximately the mean of the stationary series $y_t = \nabla x_t$, and therefore it is estimated by

$$\bar{y} = \frac{1}{n-1}\sum_{2}^{n} y_i = \frac{x_n - x_1}{n-1} = 0.007 \tag{4}$$

Model (2) for the corrected data leads to a similar forecasting function:

$$\hat{x}_t(l) = B_0^{(t)} + B_1^{(t)}(0.43)l + 0.0075l \tag{5}$$

where, again, apart from some transitory effects, the forecast will be a deterministic linear trend with a slope very approximately estimated by (4).

One wonders if a more adaptive forecasting equation such as

$$\hat{x}_t(l) = A_0^{(t)} + \hat{\beta}_t l + T_t$$

where T_t is some transitory component and $\hat{\beta}_t$ depends on the forecast origin would result in a smaller forecast error. For instance, the model

$$\phi(B)\,\nabla^2 x_t = (1 - \theta B)a_t$$

will forecast with an adaptive slope given by

$$\hat{\beta}_t = (1-\theta)\,[\nabla x_t + \theta\nabla x_{t-1} + \theta^2\nabla x_{t-2} + \cdots]$$

and the slope will be estimated giving more weight to the last observed growth rate. Taking a second difference, the following model is obtained:

$$\nabla\tilde{y}_t = (1 - 0.48B)a_t \tag{6}$$

and although this is similar to model (2) it has a larger residual variance.

Table I shows the mean squared errors of forecast for different lags with model (2) and (6) when applying a similar post-sample forecast exercise as described in the paper with the difference that here the same parameters are used in all the subseries instead of re-estimating them as in the second Section. That is, the models have been estimated in the period (1, 103) and out-of sample forecast of one to 12 steps ahead are computed with origin 104, ..., 163. Table I shows the result of this exercise. It can be seen that the AR(2) provides a better forecast and, therefore, the non-stationary and more adaptive structure of the IMA(1,1) does not help in this case. Note also that the difference between both models increases with the lag, which suggests that this type of comparation could be useful to discriminate between competitive models.

A procedure that may be useful to identify non-linear relationships is to plot y_t versus y_{t-k} and try to find what pattern will indicate TAR behaviour. First, note that a Gaussian autoregressive process will display linear behaviour, that is, $E(y_t \mid y_{t-k})$ is a linear function of y_{t-k}. Let us consider this expectation for a simple TAR(2,2,2) model with two regimes and threshold lag y_{t-2}. Then

$$E(y_t \mid y_{t-1}) = P(y_{t-1})\,[\phi_0^{(1)} + \phi_1^{(1)} y_{t-1} + \phi_2^{(1)} E(y_{t-2} \mid y_{t-1})] \\ + (1 - P(y_{t-1}))\,[\phi_0^{(2)} + \phi_1^{(2)} y_{t-1} + \phi_2^{(2)} E(y_{t-2} \mid y_{t-1})] \quad (7)$$

where $P(y_{t-1}) = P(y_{t-2} < r \mid y_{t-1})$ and r is the threshold. Now, when y_{t-1} takes values near the mean of the marginal distribution $f(y_{t-1})$, then $P(y_{t-1})$ is roughly constant and equal to $P(y_{t-2} < r)$. However, when y_{t-1} goes towards $-\infty$ then $P(y_{t-2} < r \mid y_{t-1})$ goes to one (assuming a positive correlation between y_{t-1} and y_{t-2}) and when $y_{t-1} \to \infty$ then $P(y_{t-2} < r \mid y_{t-1})$ goes to zero. Taking this into account and assuming that $E(y_{t-2} \mid y_{t-1})$ is roughly a linear function of y_{t-1} it is clear from model (7) that $E(y_t \mid y_{t-1})$ will be a linear function of y_{t-1} in the central values of y_{t-1} and will display a different slope when y_{t-1} is very small or very large. The slope in the extremes will be determined mainly by the coefficients of both regimes.

On the other hand, when we reach the threshold lag

$$E(y_t \mid y_{t-2}) = \phi_0^{(1)} + \phi_1^{(1)} E(y_{t-1} \mid y_{t-2}) + \phi_2^{(1)} y_{t-2}, \text{ if } y_{t-2} < r$$

and

$$E(y_t \mid y_{t-2}) = \phi_0^{(2)} + \phi_1^{(2)} E(y_{t-1} \mid y_{t-2}) + \phi_2^{(2)} y_{t-2}, \text{ if } y_{t-2} \geqslant r$$

Now according to the previous result, $E(y_{t-1} \mid y_{t-2})$ is a linear function of y_{t-2} in the central part and it will show different slopes in the extremes. Therefore it is clear that $E(y_t \mid y_{t-2})$ will have a breaking point around the value $y_{t-2} = r$, and will be approximately, apart from end effects, piecewise linear.

These result suggest a simple identification tool for TAR models. Let us estimate $E(y_t \mid y_{t-k})$ for $k = 1, 2, \ldots$ and look for non-linear patterns. If $E(y_t \mid y_{t-1})$ displays some non-linear behaviour in the extremes and is linear in the central part this suggests TAR behaviour. The threshold variable y_{t-d} can be identified by searching for a clear break in the slope of the curve $E(y_t \mid y_{t-d})$, that it is not present for $k < d$ and remains for $k \geqslant d$.

Table I. 10^{-4} mean squared errors of forecast

	1	2	3	4	5	6	7	8	9	10	11	12
AR(2)	1.03	1.09	1.13	1.13	1.10	1.10	1.04	1.00	0.88	0.88	0.87	0.88
IMA(1,1)	1.23	1.51	1.73	1.83	1.83	1.93	1.92	1.98	1.54	1.53	1.56	1.86

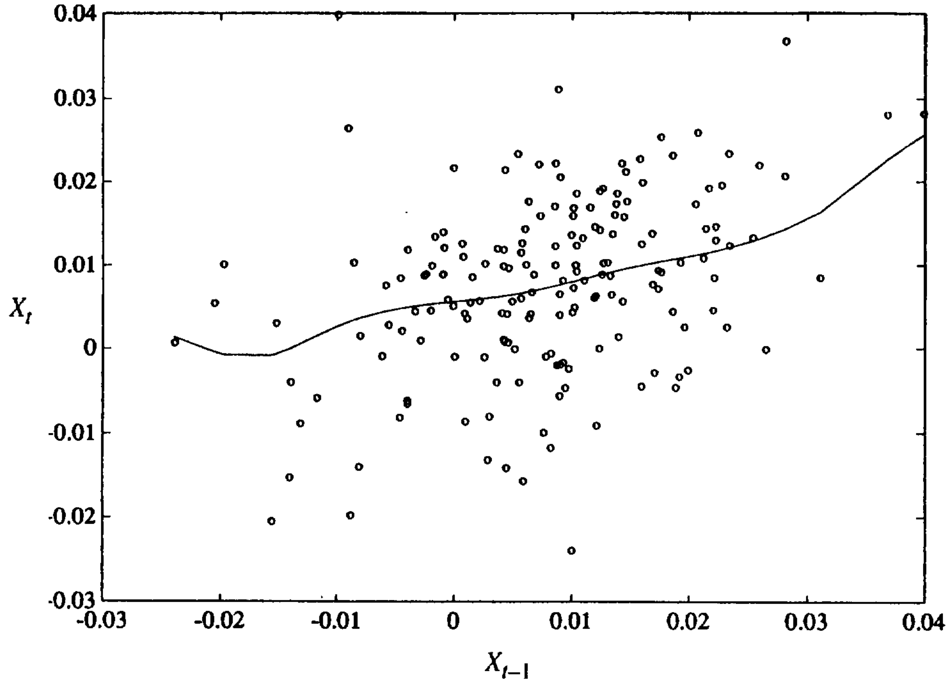

Figure 1. Non-parametric estimator of $E(X_t \mid X_{t-1})$ for the GNP data. The bandwidth is 0.005 ($n = 0.05$)

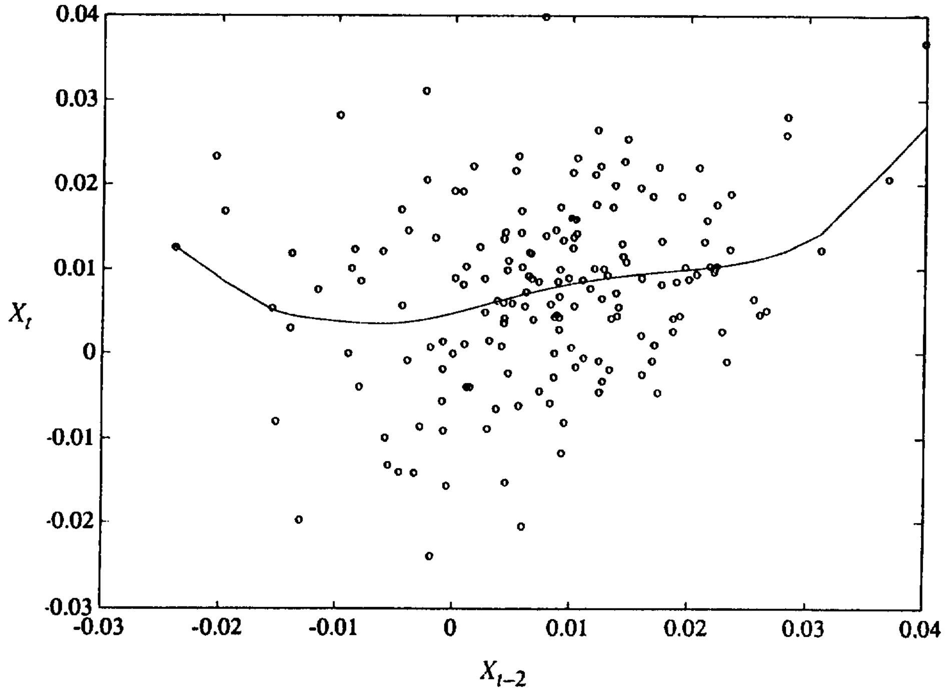

Figure 2. Non-parametric estimator of $E(X_t \mid X_{t-2})$ for the GNP data. The bandwidth is 0.005 ($n = 0.05$)

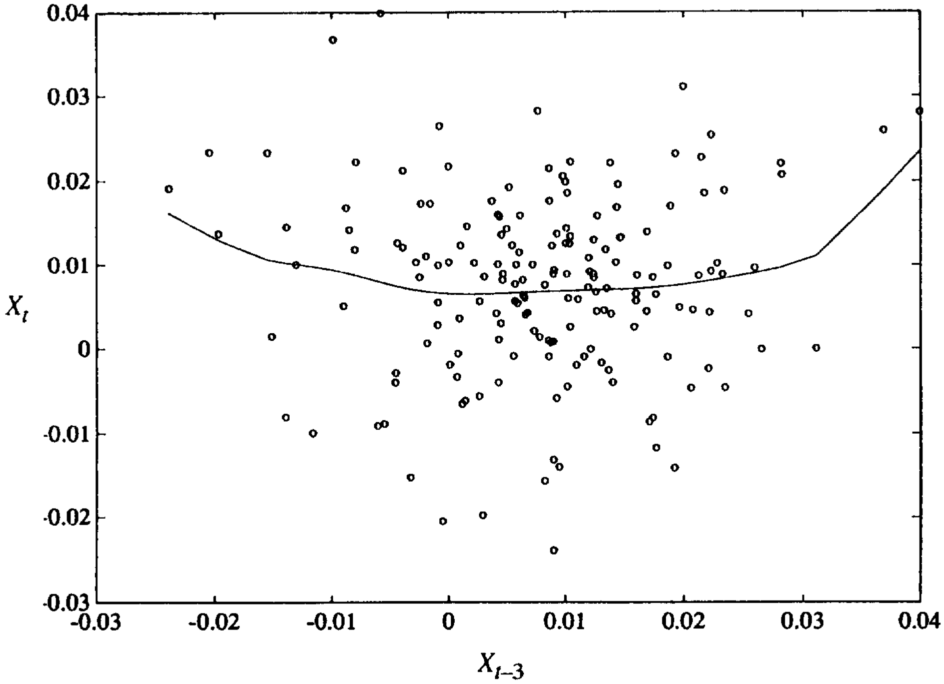

Figure 3. Non-parametric estimator of $E(X_t \mid X_{t-3})$ for the GNP data. The bandwidth is 0.005 ($n = 0.05$)

Figure 1 shows a non-parametric estimate of $\hat{E}[y_t \mid y_{t-1}]$ for the GNP data. This curve has been estimated using a normal kernel $K(x - x_i/h)$ and bandwith $h = 0.005$. That is,

$$\hat{E}[y_t \mid y_{t-1}] = \Sigma y_t \frac{K[(x - y_{t-1})/h]}{\Sigma K[(x - y_{t-1})/h]}. \tag{9}$$

It can be seen that the curve is approximately linear in the middle, and displays some non-linearity in the extremes. This indicates TAR behaviour.

Figure 2 shows $\hat{E}[y_t \mid y_{t-2}]$. The graph suggests a breaking point in the slope around zero. This implies that y_{t-2} is a threshold variable and zero the threshold value. The behaviour of $\hat{E}[y_t \mid y_{t-3}]$ is in agreement with these assumptions (Figure 3) as well as the curves $\hat{E}[y_t \mid y_{t-k}]$ for $k > 4$ (not shown to save space).

In summary, this analysis confirms that the TAR model seems to capture the relevant features of the data. Also, as shown earlier, it accommodates the outliers found with the linear modelling, stressing the need for further research into the relationship between outliers and non-linear behaviour.

THE MONTHLY RETAIL PRICE FOR UNLEADED GASOLINE

The analysis of influential observations and outliers in this set of data shows several innovational outliers. This suggests the random variance-shift model as an interesting alternative. The procedure seems to work very well.

In my experience with Gibbs sampling applied to outlier problems I have found that when there is strong masking, a very large number of iterations are needed in order to identify the outliers. The problem seems to be that when the parameter space is large and the joint distribution is multimodal, as is usually the case in outlier problems with a strong masking effect, if the starting point is very close to a local maximum and far from the global maximum of the distribution, the probability of moving towards the global maximum is very low. This behaviour has been shown by Justel and Peña (1994) in regression problems with strong masking and, although it does not seem to be a problem in this set of data, it may occur in all masking problems.

Also, the authors suggest several alternative approaches for analying this set of data. More research is needed on the theoretical relationship among these approaches and on efficient procedures to discriminate among them.

ADAPTIVE FORECASTING

Adaptive forecasting is a very brilliant and challenging idea: as any fitted model may be wrong, we could reduce the cost of fitting an incorrect model by using a different set of parameter values for each forecasting horizon. The parameters to be used for forecasting l periods ahead are obtained by minimizing the l periods ahead residual sum of squares.

It is easy to find situations in which adaptive forecasting will work. For instance, suppose that the first difference, y_t, of some observed time series, X_t, follows an AR(1) model, but the series has an additive outlier at $t = T$, that is,

$$\begin{aligned} X_t &= Z_t + \omega \\ X_t &= Z_t \qquad t \neq T \end{aligned} \tag{10}$$

where Z_t follows an ARI(1,1) model. Then, after differencing, stating $y_t = \nabla X_t$, $z_t = \nabla Z_t$, we will have

$$\begin{aligned} y_t &= z_t \qquad t \neq T, T+1 \\ y_T &= z_T + \omega \\ y_{T+1} &= z_{T+1} - \omega \end{aligned} \tag{11}$$

where z_t follows an AR(1) model. The model identified for y_t will also be AR(1) and the parameter ϕ will be estimated by the first autocorrelation coefficient of the observed y_t.

Suppose that adaptive forecasting is used to estimate the parameter for different lags. Then the function to be minimized is $\Sigma e_t^2(l)$, where

$$e_t(l) = y_{t+l} - \hat{y}_t(l) = a_{t+l} + \phi a_{t+l-1} + \cdots + \phi^{l-1} a_{t-1}$$

and as

$$\Sigma e_t^2(l) = \Sigma (y_{t-l} - \phi^l y_t)^2$$

the estimator obtained by minimizing this function, $\hat{\phi}(l)$, will be

$$\hat{\phi}(l) = \left(\frac{\Sigma y_{t+l} y_t}{\Sigma y_t^2} \right)^{1/l} = r_l^{1/l}$$

Now, according to model (11) this estimator can be written as

$$\hat{\phi}(l) = \left(\frac{\Sigma z_{t+l} z_t + \omega(z_{T+l} + z_{T-l} - z_{T+1-l} - z_{T+1+l})}{\Sigma z_t^2 + 2\omega(z_T - z_{T+1}) + 2\omega^2}\right)^{1/l} \qquad l > 1$$

$$\hat{\phi}(1) = \left(\frac{\Sigma z_t z_{t+1} + \omega(z_{T-1} + z_{T+1} - z_T - z_{T-2}) - \omega^2}{\Sigma z_t^2 + 2\omega(z_T - z_{T+1}) + 2\omega^2}\right)$$

and $\hat{\phi}(1)$ will be more affected by the outlier than $\hat{\phi}(l)$, $(l > 1)$. Therefore $\hat{\phi}(l)$, for $l > 2$ will be a more robust estimator of $\hat{\phi}$ than $\hat{\phi}(1)$. For instance, Table II shows 100 simulations of this estimator for $l = 2$, different values of ω and ϕ and a sample size of 100.

An additional usefulness of adaptive estimation is as a diagnostic tool to identify problems with the model. For instance, Table III shows the value of $\hat{\phi}(l)$ as a function of l and the size of the outlier. When $\omega = 0$, the no outliers case, similar values are obtained for all lags and no clear pattern is found. On the other hand, as the size of the outlier increases, better results are obtained by using a larger lag.

These results suggest that plotting $\hat{\phi}(l)$ on l could be an interesting diagnostic tool. A clear trend in the estimation of the parameter will indicate the need for a closer examination of the data to look for non-linearities, outliers and other sources of model inadequacy.

As a final example, let us assume that we want to discriminate between the two models

$$y_t = \phi y_{t-1} + \mu + a_t \tag{12}$$

$$\nabla y_t = (1 - \theta B) a_t \tag{13}$$

For large θ and $\phi = 1 - \theta$, both models display a similar behaviour and it is difficult to discriminate between them. Suppose that we choose model (12) when model (13) is correct. Then $\hat{\phi}(l)$ is expected to increase with l, showing the non-stationary behavior, that will be more obvious for lags greater than the first. Again, as Table IV shows, studying the evolution of the parameters for different lags could be a useful diagnostic device.

Table II. Mean and standard deviation of two estimates of the AR(1) parameter in 100 simulations. In all cases additive outliners of size has been added at positions 50 and 51 in a sample of size 100

	$\phi = 0.5,\ \omega = 5$		$\phi = 0.5,\ \omega = 3$		$\phi = 0.8,\ \omega = 5$		$\phi = 0.8,\ \omega = 3$	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std
$\hat{\phi}(1)$	0.23	0.09	0.38	0.09	0.59	0.10	0.68	0.09
$\hat{\phi}(2)$	0.53	0.20	0.55	0.20	0.74	0.14	0.75	0.11

Table III. Mean values of $\hat{\phi}(l)$ in 100 simulations as a function of l and ω, the size of the additive outlier. The true value of ϕ is 0.8

		l				
		1	2	3	4	5
ω	0	0.78	0.78	0.77	0.76	0.76
	5	0.59	0.73	0.74	0.74	0.73
	7	0.43	0.67	0.70	0.71	0.71
	10	0.25	0.60	0.69	0.69	0.70

Table IV. Mean values of $\hat{\phi}(l)$ in 100 simulations from an IMA(1,1) with several θ values. The sample size is 100

	l				
θ	1	2	3	4	5
0.8	0.38	0.58	0.69	0.74	0.75
0.6	0.70	0.81	0.85	0.87	0.88
0.4	0.87	0.91	0.92	0.93	0.93

CONCLUSIONS

I have learned a lot reading this paper. It is a fine example of the useful interaction between theory and applications that can be found in the most important statistical advances. The authors are to be congratulated for showing us new methods for time-series analysis and for introducing useful new ideas that are improving our understanding of statistical models and data analysis.

REFERENCES

Justel, A. and Peña, D., 'Improving Gibbs sampling in Outlier Detection problems', 1994 (unpublished).

Peña, D., 'Influential observations in time series' *Journal of Business and Economic Statistics*, **8** (1990), 2, 235–241.

Peña, D. and Maravall, A., 'Interpolation, Outliers and Inverse Autocorrelations' *Communication in Statistics-Theory and Methods*, **20** (1991), 3175–3186.

Author's biography:
Daniel Peña is Professor of Statistics at the Universidad Carlos III de Madrid. He has held visiting positions at the University of Wisconsin-Madison and the University of Chicago. His papers have appeared in *JASA*, *Technometrics*, *JBES*, *Biometrika* and *JRSSB*, among other journals. His research interests are time series, Bayesian statistics and robust methods.

Author's address:
Professor Daniel Peña Department of Statistics and Econometrics, Universidad Carlos III de Madrid, 28903 Getafe, Madrid, Spain.