How to fit a time series model when you know it is wrong?

Howell Tong

London School of Economics

Y. Xia and H. Tong, Statistical Science 2011

March 25, 2014

Outline

Are Conventional Methods of Estimation Fit for Purpose?

Conventional Methods of Parametric Modelling

1. Assume that there is a true model;
2. Evaluate the efficacy of the estimation as if the postulated model is true.

Are Conventional Methods of Estimation Fit for Purpose?

Let

$$\begin{aligned} y &= \text{the observed/real data} \\ x &= \text{the data generated by the postulated model.} \end{aligned}$$

Are Conventional Methods of Estimation Fit for Purpose?

Postulated Model

$$x_t = g_\theta(x_{t-1}, ..., x_{t-p}) + \varepsilon_t, \quad (1)$$

where ε_t is the innovation, and the function $g_\theta(.)$ is known up to parameters θ.

Are Conventional Methods of Estimation Fit for Purpose?

Observed Data: $\{y_1, y_2, \cdots, y_T\}$

Conventional methods

Typically minimize a loss function

$$L(\theta) = (T-p)^{-1} \sum_{t=p}^{T-1} \{y_{t+1} - g_\theta(y_t, ..., y_{t-p+1})\}^2,$$

where, here and elsewhere, T denotes the sample size.

Are Conventional Methods of Estimation Fit for Purpose?

Efficient **IF** the postulated model

$$x_t = g_\theta(x_{t-1}, ..., x_{t-p}) + \varepsilon_t,$$

is true.

Are Conventional Methods of Estimation Fit for Purpose?

Match observations $\{y_1, y_2, ..., y_T\}$, by postulating g.
Postulate linear g_θ, Gaussian ε_t, zero mean, finite var(ε_t).
Let

$$\{C(j), j = 0, 1, 2, ..., T-1\}$$

denote the sample autocovariance function of the y-data.

Minimizing $L(\theta)$ yields well-known estimates of θ that are functions of

$$\mathbf{S} = \{C(0), C(1), ..., C(p)\}.$$

Are Conventional Methods of Estimation Fit for Purpose?

1. Postulated model is right
   - $\mathbf{S}$ is a minimal set of sufficient statistics (ignoring boundary effects)
2. Postulated model is wrong
   - It is unlikely that $\mathbf{S}$ is a minimal set of sufficient statistics.

Multi-step predictions

ARIMA(0,1,1) model leads to EWMA predictor:

$$(1 - B)X_t = (1 - \eta B)e_t$$

$$\hat{X}_t(1) = (1 - \eta)X_t + \eta \hat{X}_{t-1}(1).$$

$$\hat{X}_t(\ell) = \hat{X}_t(\ell - 1); \quad \ell = 2, 3, \ldots$$

SERIES A "Uncontrolled" Concentration, Two-Hourly Readings: Chemical Process

Figure: Box and Jenkins Series A-Chemical Process Concentration Readings: every two hours

Table 1. *Values of* Ave $(\eta(l))$ *and* $\hat{R}(l)$ *for Box–Jenkins' Series* A

l	Ave $(\eta(l))$	$\hat{R}(l)$	l	Ave $(\eta(l))$	$\hat{R}(l)$
1	0·730	1·000	21	1·000	1·429
2	0·747	0·994	22	1·000	1·437
3	0·789	0·991	23	1·000	1·455
4	0·781	0·984	24	1·000	1·407
5	0·761	0·984	25	1·000	1·322
6	0·704	0·973	26	1·000	1·305
7	0·605	0·942	27	1·000	1·339
8	0·776	1·049	28	1·000	1·374
9	0·882	1·103	29	1·000	1·387
10	0·956	1·303	30	1·000	1·460
11	0·971	1·192	31	1·000	1·444
12	0·981	1·419	32	1·000	1·445
13	0·983	1·587	33	1·000	1·450
14	0·992	1·713	34	1·000	1·477
15	1·000	1·546	35	1·000	1·476
16	1·000	1·487	36	1·000	1·546
17	1·000	1·426	37	1·000	1·499
18	1·000	1·475	38	1·000	1·461
19	1·000	1·478	39	1·000	1·502
20	1·000	1·411	40	1·000	1·545

Figure: $Ave(\eta(\ell))$ refers to model in pocket ℓ. $\hat{R}(\ell)$ is ratio of ave squared prediction errors using η estimated in 1st pocket to that in the

George Box

1. Essentially, all models are wrong, but some are useful.
2. Consistency of estimation is predicated on the model being correct/true.
3. Post-modelling reconciliation of Box's dictum : diagnostic checks, goodness of fit tests, etc.

Challenge

Recognising Box's dictum right at the modelling stage.

Catch-all Approach

Our objective

- Good matching of salient features of the observed time series.
- Short-term prediction is secondary.

Catch-all Approach

For expositional simplicity, let $p = 1$.
We want

$$P\{x_1(\theta_0) < u_1, ..., x_n(\theta_0) < u_n | x_0(\theta_0) = y_0\}$$

to be as close as possible to

$$\equiv P\{y_1 < u_1, ..., y_n < u_n | y_0\}$$

almost surely for some θ_0 and any n and real values $u_1, u_2, \ldots, u_n$.

First Weaker Form

First weaker form

Generalize the loss function to

$$L(\theta) = (T-p)^{-1} \sum_{t=p}^{T-1} \sum_{m=1}^{Q} w_m \{y_{t+m} - g_\theta^{[m]}(y_t, ..., y_{t-p+1}), \}^2,$$

where

$$g_\theta^{[m]}(y_t, ..., y_{t-p+1}) = E(x_{t+m} | x_t = y_t, \ldots, x_{t-p+1} = y_{t-p+1}),$$

and $w_m \geq 0$ and $\sum w_m = 1$.

Second Weaker Form

Suppose that observable $\{y_t\}$ and generated $\{x_t(\theta)\}$ are both second-order stationary.

$$D_C(y_t, x_t(\theta)) = \sup_{\{w_m\}} \sum_{m=0}^{\infty} w_m \{\gamma_{x(\theta)}(m) - \gamma_y(m)\}^2.$$

$$f_y(\omega) = \frac{1}{2\pi}\gamma(0) + \frac{1}{\pi}\sum_{k=1}^{\infty} \gamma_y(k)\cos(k\omega).$$

$$D_F(y_t, x_t(\theta)) = \int_{-\pi}^{\pi} \Big\{\frac{f_y(\omega)}{f_x(\omega)} + \log(\frac{f_x(\omega)}{f_y(\omega)}) - 1\Big\} d\omega.$$

Second Weaker Form

Whittle's likelihood

$$\hat{\theta} = \min_{\theta} \sum_{j=1}^{T} \left\{ \frac{I(\omega_j)}{f_\theta(\omega_j)} + \log(f_\theta(\omega_j)) \right\},$$

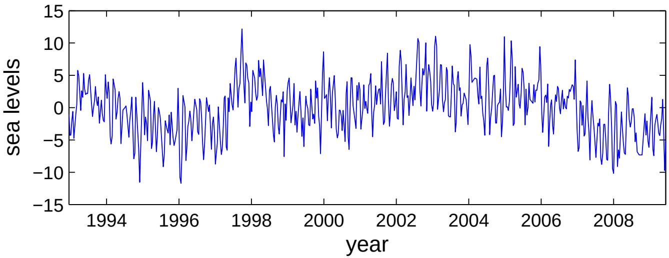
where

$$I(\omega) = \frac{1}{2\pi T} \left| \sum_{t=1}^{T} y_t \exp(-\iota \omega t) \right|^2$$

and $\omega_j = 2\pi j/T$.

- Whittle's likelihood matches the second-order moments by using a natural sample version of $D_F(y_t, x_t(\theta))$ up to a constant.

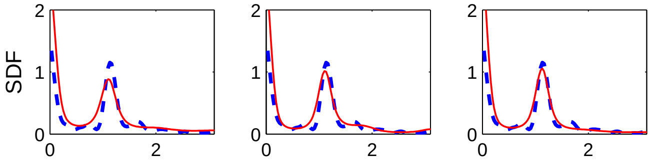
Example R1 [Sea levels]



Example R1 [Sea levels]

Postulated model= AR(6), using AIC for order determination.

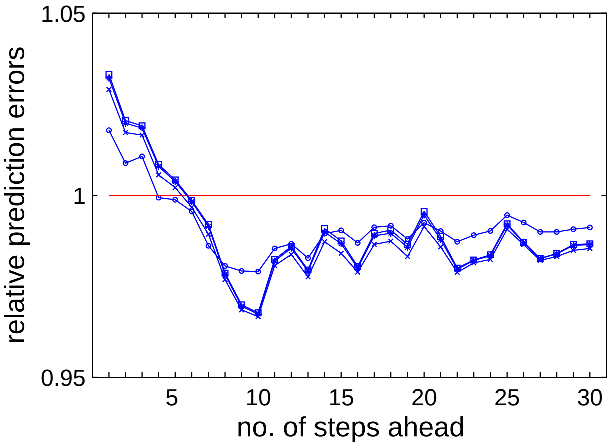
MLE; Whittle; APE(≤ 20)



Example R1 [Sea levels]

relative averaged multi-step ahead prediction errors by taking those of the one-step method as 1 unit.

o = APE(≤ 10); $\times$ =APE(≤ 20); $\star$ = APE(≤ 30); $\diamond$ = APE(≤ 50)



Example R2: Nicholson's Blowflies

- Total number of blowflies (Lucilia cuprina) under controlled laboratory conditions.
- Counts for every second day.
- The developmental delay (from egg to adult) is between 14-15 days.
- Nicholson obtained 361 bi-daily recordings over a 2-year period (722 days).
- A major transition appears to have occurred around day 400.
- Following Tong (1990), we consider the first part of the time series (to day 400, thus T=200), for which the population has a 19-bi-day cycle; see figure.

We postulate the single species animal population discrete model as suggested by Gurney *et al.* (1980), and thus

$$x_t = Poisson(cx_{t-\tau}\exp(-x_{t-\tau}/N_0)x_{t-\tau} + \nu x_{t-1}),$$

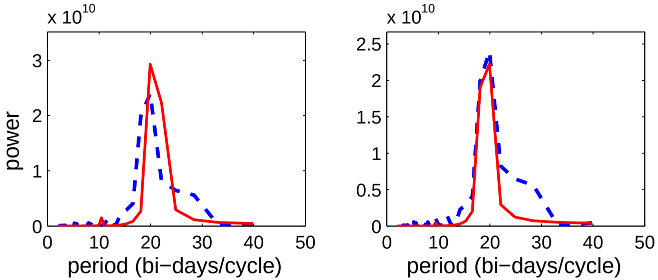
where we take $\tau = 8$ (bi-days) corresponding to the time taken for an egg to develop into an adult.

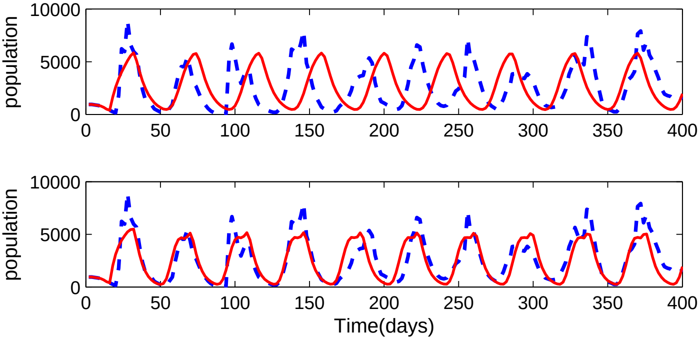
- 3 parameters: c, N_0 and ν.
- MLE estimates for the parameters are

$$\hat{c} = 8.49, \hat{N}_0 = 528.23, \hat{\nu} = 0.77.$$

- Catch-all method gives

$$\hat{c} = 8.82, \hat{N}_0 = 604.98, \hat{\nu} = 0.67.$$

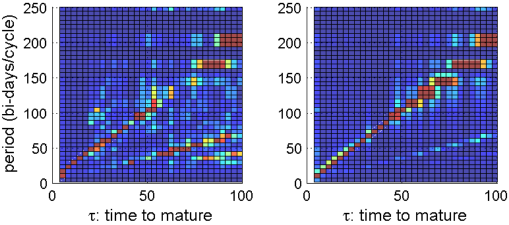

x 10^10
power
0
1
2
3
0 10 20 30 40 50
period (bi−days/cycle)
x 10^10
0
0.5
1
1.5
2
2.5
0 10 20 30 40 50
period (bi−days/cycle)


population
10000
5000
0
0 50 100 150 200 250 300 350 400
population
10000
5000
0
0 50 100 150 200 250 300 350 400
Time(days)

How do the cycles change with the time needed by the fly to grow to maturity?

We vary the time τ from 4 to 100 bi-days. The corresponding cycles (in bi-days) are shown in the figure.

- APE($\leq T$) shows a clear linearly increasing trend in the cycle-periods as τ increases.
- APE(≤ 1) shows strange excursions that are difficult to interpret.



Example R3: [Measles in London]

Postulated SIR model for the transmission of measles [infected (I); susceptible (S); birth (B)]:

$$I_{t+1} = \exp(\delta_{t,k}\beta_k)S_t I_t, \quad S_{t+1} = S_t + B_t - I_{t+1} = S_0 + \sum_{\tau=0}^{t} B_\tau - \sum_{\tau=1}^{t+1} I_\tau,$$

I_t cannot be observed; can observe y_t that has mean I_t. For observable y_t, we postulate model

$$x_t = Poi(I_t),$$

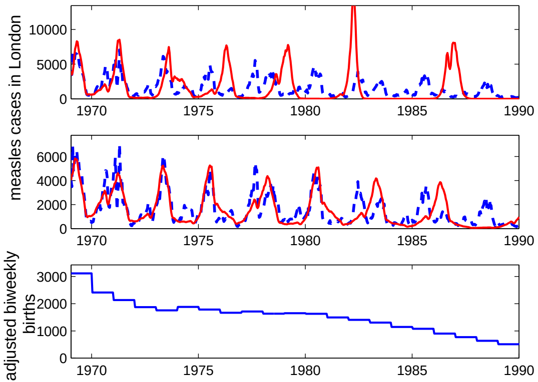
where $(\delta_{t,k}\beta_k)$ is employed to indicate the seasonality force, with $\delta_{t,k} = 1$ if time t is at the kth season, 0 otherwise.

- For measles, the time unit for t is bi-weekly, based on the infection procedure of measle.

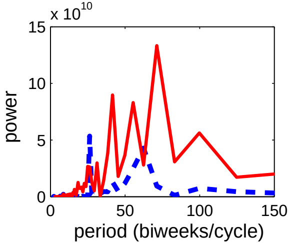
Table: Estimated parameters in the transmission model

method	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8
APE(≤ 1)	-11.92	-12.00	-11.88	-11.99	-11.89	-11.81	-11.89	-11.97
APE($\leq T$)	-11.95	-12.00	-11.93	-11.99	-11.93	-11.89	-11.93	-11.98
	β_9	β_{10}	β_{11}	β_{12}	β_{13}	β_{14}	β_{15}	β_{16}
APE(≤ 1)	-11.92	-11.99	-12.05	-12.01	-11.93	-11.96	-11.98	-12.04
APE($\leq T$)	-11.95	-11.99	-12.03	-12.00	-11.96	-11.98	-11.99	-12.02
	β_{17}	β_{18}	β_{19}	β_{20}	β_{21}	β_{22}	β_{23}	β_{24}
APE(≤ 1)	-11.95	-12.15	-12.28	-12.40	-12.21	-11.99	-11.79	-11.87
APE($\leq T$)	-11.97	-12.08	-12.16	-12.23	-12.12	-11.99	-11.87	-11.92
	β_{25}	β_{26}	S_0					
APE(≤ 1)	-11.99	-11.98	178280					
APE($\leq T$)	-11.99	-11.98	168190					

The skeletons based on APE(≤ 1) and APE($\leq T$) are shown in solid red line in panel 1 and panel 2. APE($\leq T$) much better match than APE(≤ 1) in terms of outbreak scale and cycle period.



The periodogram is also much better matched by APE($\leq T$) than by APE(≤ 1).



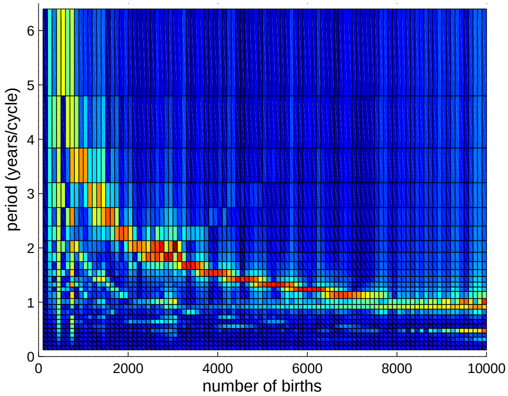


Figure shows clearly that when the birth rate is high (from about 5000 and above) the cycle is annual, but when the birth rate is medium at about 3000 to 4000, the cycles become two-year cycles.

As the birth rate gets lower, the model shows that cycle become three-year cycles or even five-year cycles. Thus, the postulated model has thrown some light on the effect of the birth rate on the frequency of outbreaks of measles.

Our conclusion agrees with the changing cycle-periods observer over different time durations and at different places as reported by the epidemiologists Earn *et al. Nature, 2000*.

Conclusions

(1) Truer to Box than Box!

(2) Ways to improve the ability of a postulated parametric model to match features (e.g. second-order moments, cycles, and others) of the observed time series that are deemed important;

Conclusions

(3) Methods based on just the one-step-ahead prediction often found unfit for purpose, in many situations:

- the absence of a true model
- short data sets
- observation errors
- highly cyclical data
- and others

(4) Deeper understanding of Whittle's likelihood as an extended likelihood of a model; W-likelihood is a precursor of XT-likelihood.

TONY, ENJOY YOUR PERMANENT SABBATICAL LEAVE!

Optimal Parameter

For some positive integer m (which may be infinite), we define the *optimal parameter* by

$$\vartheta_{m,\mathbf{w}} = \arg\min_{\theta} \sum_{k=1}^{m} w_k E\left[y_{t+k} - E\{x_{t+k}(\theta) | X_t(\theta) = Y_t\}\right]^2,$$

where $X_t(\theta) = (x_t(\theta), x_{t-1}(\theta), \ldots, x_{t+p-1}(\theta))$, and $\{w_k\}$ define the weight function, typically positive and summing to unity.

Optimal Parameter

For ease of exposition, we assume that the solution to the above minimization is unique. Then we have immediately

Theorem

Under general conditions

$$\tilde{\theta}_{\{m\}} \to \vartheta_{m,\mathbf{w}}$$

in probability as the sample size $T \to \infty$.