An Appreciation of Professor George C. Tiao from across The Ocean

Presented by Howell Tong

Department of Statistics

LSE

last part based on joint work with Kung-Sik Chan, Dong Li and Shiqing Ling

17 December 2013

Overview

George

1976

- Bayesian statistics, intervention analysis, seasonal adjustment, adaptive prediction, aggregation, causality, outlier detection, EACF, multiple time series, canonical analysis, environmetrics,......
- The ET Interview by Ngai Hang Chan, 1999.

身無彩鳳雙飛翼
心有靈犀一點通

- Institute of Statistics Annual Conference, King's College, Cambridge.
- A young 43-year old vs a younger 32-year old.
- Multiple time series.

Multiple time series

- time-domain approach vs frequency-domain approach using

$$\mathbf{X}(\omega_j) = (2\pi T)^{-1} \sum_{t=1}^{T} \mathbf{X}_t e^{i\omega_j t}$$

- Frequency-domain: Treat $\mathbf{X}(\omega_j)$s as if they were iid so that we can perform conventional PCA, canonical correlation analysis, factor analysis etc.-Priestley, Subba Rao and myself. Convenient but yielding no exciting insights, lost in generality!
- Time-domain: Box and Tiao (1977, Biometrika), canonical analysis led to a major breakthrough!

Box and Tiao, 1977

Canonical correlation analysis wrt vector time series and its predictor.

Figure: US hog data:hog supply; hog price; corn price; corn supply; farm wages

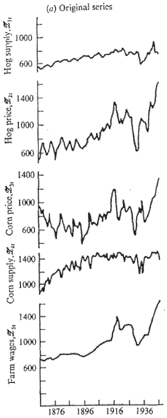


Figure: US hog data:hog supplyZ_{1t}; hog priceZ_{2t}; corn priceZ_{3t}; corn supplyZ_{4t}; farm wagesZ_{5t}

Pre-dating cointegration by 10 years

- $$\frac{H_p H_s}{(R_p R_s)^{0.75} W^{0.50}}$$

  is approximately independently distributed about a fixed mean. *That is these 5 time series co-move/co-integrate!!!*
  {return to the farmer} / {farmer's expenditure} follows a stable economic law!

- $$H_s(R_p/H_p)$$

  is approximately independently distributed about a fixed mean.

  {hog supply} x {price ratio} follows a stable economic law.

- Box and Tiao (1977) pre-dated Engle and Granger (1987) by 10 years.
- The two Georges were victims at the hands of those whose motto is *"What I understand is mine."*
- Speaking as a fellow victim of similar acts committed by a few econometricians, I offer them my deepest sympathy.

Multi-step predictions

- If the model is true and known up to some tuning parameters, we can estimate the unknown parameters by minimizing a functional of one-step prediction errors. Use the fitted model as if it were the true model and obtain the k-step prediction in the usual way.
- If the model is NOT true, what do we do?

George Tiao and Xu (1993) focus on prediction.
They wear a pair of multi-pocket trousers, using the model fitted in the ℓ-th pocket for ℓ-step prediction. Different pockets contain different fitted models.

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

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

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

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

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 using different η

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

- Focus is on fitting THE wrong model rather than prediction.
- Pretending that the wrong model is true, fitting it with conventional methods and then performing diagnostic checks is useful but smacks of

  自圓其說

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

  Observed Data: $\{y_1, y_2, \cdots, y_T\}$
- For expositional simplicity, let $p = 1$.

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

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

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

Catch-all

Standing on the shoulders of David Cox, George Tiao etc. while tunnelling through their pockets!

$$\begin{aligned} & E\left[\{x_1(\theta_0), \ldots, x_m(\theta_0)\} | x_0(\theta_0) = y_0\right] \\ = \quad & E\left[\{y_1, \ldots, y_m\} | y_0\right]. \end{aligned}$$

We set θ_0 to minimize the difference of the predictions,

$$E\left\{ \left\| E\left[\{x_1(\theta), \ldots, x_m(\theta)\} | x_0(\theta) = y_0\right] - E\left[\{y_1, \ldots, y_m\} | y_0\right] \right\|^2 \right\}.$$

Nicholson's blowfly

- 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.

- Single species animal population discrete model as suggested by Gurney *et al.* (1980)

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

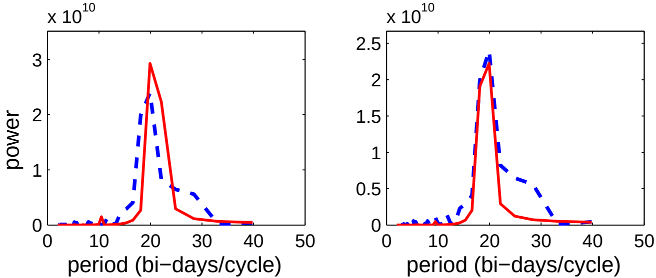
where we take $\tau = 8$ (bi-days) corresponding to the time taken for an egg to develop into an adult.

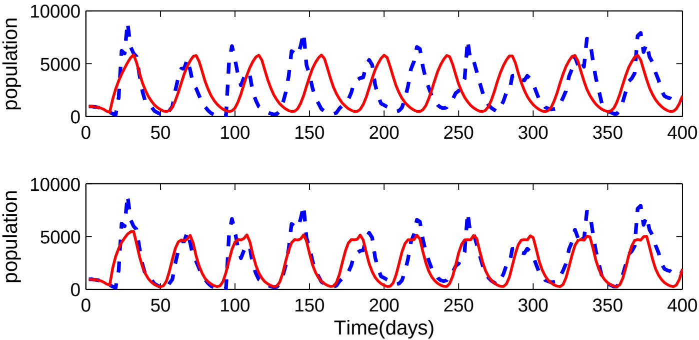
- 3 parameters: c, N_0 and ν.
- MLE estimates for the parameters are

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

- Catch-all method gives

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


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


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

Bayesian Statistics

- George entered time series analysis through the Bayesian door.
- Threshold models in time series have Bayesian underpinnings.
- Time series $\{X_t : t = 0, \pm 1, \pm 2, \cdots\}$
- 'True' model

$$E(X_t | X_{t-1} = x) = \mu(x)x,$$

  μ a 'smooth' function.

- Approximate it by a Bayesian linear model

$$E(X_t | X_{t-1} = x) = \theta x,$$

$$\theta \sim N(c, V).$$

- The closeness of approximation measured by the loss function

$$L(\theta) = h[1 - \exp\{-\frac{1}{2k}(\theta - \mu)^2\}],$$

where $h > 0, k > 0$. Conjugate to the Gaussian belief.

- Introduce decision space D, s.t. $\delta \in D$ moves c to $c + \delta$.
- Since we would not expect to have to make *drastic* adjustments to the value of θ for a 'smooth' function μ,

$$V(\delta) = \alpha + \beta|\delta|,$$

$\alpha > 0, \beta > 0$

- Define expected loss function by

$$E_V(\delta) = \int_{-\infty}^{\infty} L(\theta) dF_V(\theta|\delta), \quad \delta \in D,$$

- $F_V(\theta|\delta)$ is $N(c + \delta, V)$
-

$$E_V(\delta) = h[1 - (\frac{k}{k + V})^{\frac{1}{2}} \exp\{-[2(k + V)]^{-1}(\delta - \mu + c)^2\}].$$

- the minimizer of $E_V(\delta)$ with respect to δ, the Bayes decision, is uniformly zero, meaning that no adjustment is needed for $0 < \mu(x) - c < \{(1 + \gamma^2)^{\frac{1}{2}} - 1\}\gamma^{-1}$, where $\gamma = \beta(k + \alpha)^{-\frac{1}{2}}$.

TAR models

- $$X_t = a_0^{(J_t)} + a_1^{(J_t)} X_{t-1} + \ldots + a_p^{(J_t)} X_{t-p} + \sigma_{(J_t)} \varepsilon_t,$$
  where J_t is a stochastic process taking positive integer values.
- Example: $J_t = j$ if $X_{t-d} \in R_j$ for some positive integer d, where $(-\infty, \infty) = R_1 \cup R_2 \cup \ldots \cup R_k$, for some k.
  $R_i = (r_{i-1}, r_i], \quad r_0 = -\infty, r_k = \infty.$
  self-exciting threshold AR models–SETAR model order (k;p,p,...,p).
- It seems that the TAR models have passed the Tiao's test.

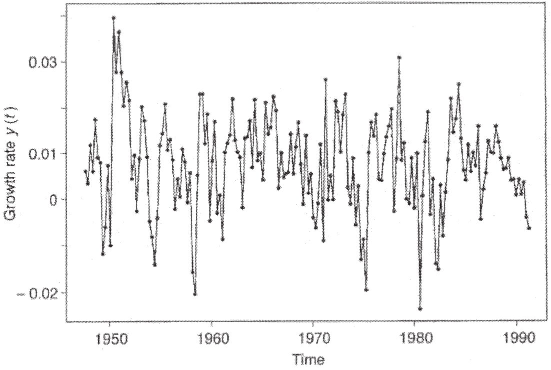


Figure 1. Time plot of growth of US quarterly real GNP: 1947.2–1991.1

Figure: X_t = growth of quarterly US real GNP (in 1982 dollars) from 1947:1 to 1991:1 (177 observations)

$$X_t = log(Y_t) - log(Y_{t-1}).$$

- Tiao and Tsay (1994) fitted a SETAR(2;2,2) model with $\hat{r}_1 = 0.0$ and estimates of $a_i^{(j)}, \sigma_j; i = 0, 1, 2; j = 1, 2$ as follows:

$j = 1$	$-0.0039(0.0033)$	$0.44(0.18)$	$-0.79(0.33)$	0.0120
$j = 2$	$0.0038(0.0014)$	$0.31(0.08)$	$0.20(0.11)$	0.0087

- The SETAR model shows that after 'contraction' the economy behaves cyclically but after 'expansion' it tends to decay exponentially to some mean level.

Motivation

- ARCH and GARCH model conditional heteroscedascity in white noise (martingale difference sequence).
- strong conditions on the coefficients for ergodicity and inference.
- simple alternatives?

Faithful geyser

Figure: Old Faithful Geyser

Waiting times between the starts of two consecutive eruptions of the old faithful geyser, collected from August 1–15, 1985.

- Waiting times between the starts of two consecutive eruptions of the old faithful geyser, collected from August 1–15, 1985.
- The waiting time is strongly associated with lag 1 values but much less so with values at higher lags.
- The presence of conditional heteroscedasticity.
- A number of outliers.

- Preliminary analysis: AR(1) model plus additive outliers for the mean structure;
  AR(1) coefficient estimate $-0.571(0.0497)$,
  mean 4.248 (0.00650) and adjusted for five
  outliers at epochs 22, 37, 172, 237 and 266.
  The residuals of the preceding model appear to be white, but are highly conditionally heteroscedastic.
- Tried to fit a GARCH(1,1) model to the AR(1) errors as suggested by the sample EACF of the absolute residuals, but without success due to convergence problem.
  $X_t = \sigma_t \eta_t$, $\sigma_t^2 = \alpha_0 + \alpha X_{t-1}^2 + \beta \sigma_{t-1}^2$, where $\alpha > 0$, $\beta > 0$ and $\alpha + \beta < 1$.

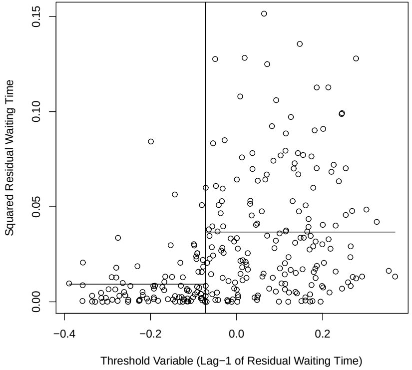
T-CHARM

- T-CHARM:

$$X_t = \sigma(X_{t-1})\eta_t, \qquad (2)$$

  $\{\eta_t\}$ IID variables of zero mean and unit variance, η_t independent $\{X_s : s < t\}$, and $\sigma(\cdot)$ is a piecewise constant function.
- Assume $\sigma(x) = \sigma_i > 0$ for $x \in R_i$, σ_i's distinct, and the regimes $\{R_i, i = 1, \ldots, m\}$ is a partition of R.

The residuals of the preceding model appear to be white, but are highly conditionally heteroscedastic.



- We fit a two-regime T-CHARM with the lag 1 of the AR(1) errors as the threshold variable, giving the following parameter estimates: $\hat{\sigma}_1^2 = 0.0093(0.0011), \hat{\sigma}_2^2 = 0.037(0.0031)$, and $\hat{r} = -0.072$ (95% confidence interval: $(-0.088, -0.058)$), which is about the 34 percentile.
- The existence of the threshold is supported by the LR test with p-value $< 10^{-5}$. And there is no need for more thresholds, based on the LR test.
- The standardized residuals from the T-CHARM are no longer conditionally heteroscedastic, based on the McLeod-Li test up to lag 50. However, the standardized residuals seems to be somewhat heavy tailed based on the quantile-quantile normal score plot.

自圓其說

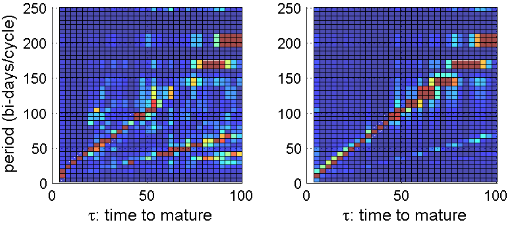
天合之作

TWO BEAUTIES HAVE JOINED ME IN WISHING YOU A HAPPY 80TH BIRTHDAY, GEORGE!

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.



Ergodicity

- $$X_t = \sigma(W_{t-1})\eta_t, \tag{3}$$
- Define the regime process $\{S_t\}$ for which $S_t = i$ if and only if $X_t \in R_i, i = 1, \ldots, m$.
- If $W_{t-1} = H(X_{t-1}, \cdots, X_{t-p})$, then $\{X_t\}$ is a p-th order Markov chain, i.e. $\{(X_t, \cdots, X_{t-p+1})^\top\}$ is a markov chain, which is (Lebesgue-)irreducible if the pdf of η_t is positive everywhere. If $\{(X_t, \cdots, X_{t-p+1})^\top\}$ is irreducible, it is uniformly ergodic, i.e. it admits a unique stationary probability measure to which the marginal distribution converges uniformly in total variation norm, for any initial distribution.

Theorem

Let $\{X_t\}$ be defined by the T-CHARM and the transition probability matrix P of the associated regime process $\{S_t\}$ be an irreducible $m \times m$ matrix. Let $Y_t = h(X_t)$, where h is a continuous function. Assume that $\{Y_t\}$ admits finite second moments and $\mathbb{E}(h(\eta_t)) \neq 0$. Let $\gamma_k = \gamma_{k,Y}$ be the kth lag auto-covariance of $\{Y_t\}$. Then $\{\gamma_k\}$ satisfies the Yule-Walker equation

$$\gamma_k = c_1\gamma_{k-1} + \ldots + c_{m-1}\gamma_{k-m+1} \tag{4}$$

for $k \geq m$.

- That $\{\gamma_k\}$ satisfies the Yule-Walker equation means that the ACF of $\{Y_t = h(X_t)\}$ is exactly the same as that of some ARMA$(m-1, m-1)$ process.
- Any instantaneous nonlinear transformation of $\{X_t\}$ is generically an ARMA$(m-1, m-1)$ process. Estimate the number of regimes m by adding the estimated AR order plus 1.

Comparison of the ACFs of T-CHARM vs. GARCH models

- Under general conditions, for $m = 2$ and for $k \geq 0$
  1. $\mathrm{cov}(\sigma^2(X_t), \sigma^2(X_{t-k})) = (\sigma_2^2 - \sigma_1^2)^2 \delta(1-\delta)\{\mathbb{P}(\sigma_2\eta_t \in R_2) - \mathbb{P}(\sigma_1\eta_t \in R_2)\}^k$, where $\delta = \mathbb{P}(\sigma_1\eta_t \in R_2)/\{1 - \mathbb{P}(\sigma_2\eta_t \in R_2) + \mathbb{P}(\sigma_1\eta_t \in R_2)\}$;
  2. $\rho_k = \{\mathbb{P}(\sigma_2\eta_t \in R_2) - \mathbb{P}(\sigma_1\eta_t \in R_2)\}^k$.
- Consider GARCH (1,1) model, namely $\sigma_t^2 = \alpha_0 + (\alpha\eta_t^2 + \beta)\sigma_{t-1}^2$, where $\alpha > 0$ and $\beta > 0$. The corresponding ACF $\{\rho_{gk}\}$ is

$$\rho_{gk} = (\alpha + \beta)^k,$$

  where $\alpha + \beta < 1$.
- Estimation of ρ_{gk} requires a finite fourth moment condition, i.e., $2\alpha^2 + (\alpha+\beta)^2 < 1$, which is rarely satisfied in practice, but no such restriction for T-CHARM.
- The T-CHARM can capture the heavy-tailed property (via Cauchy-Schwartz)

$$\frac{\mathbb{E}X_t^4}{(\mathbb{E}X_t^2)^2} = (\mathbb{E}\eta_t^4)\frac{\sum_{i=1}^m \sigma_i^4 \mathbb{P}(X_{t-1} \in R_i)}{\{\sum_{i=1}^m \sigma_i^2 \mathbb{P}(X_{t-1} \in R_i)\}^2} \geq \mathbb{E}\eta_t^4$$

- A practical parametrized T-CHARM:

$$X_t = \sigma(W_{t-1})\eta_t,$$
$$\sigma(W_{t-1}) = \sum_{i=1}^{m} \sigma_i I\{r_{i-1} < W_{t-1} \le r_i\}, \tag{5}$$

  where $-\infty = r_0 < r_1 < \cdots < r_{m-1} < r_m = \infty$.
- Let $\theta = (\sigma_1^2, ..., \sigma_m^2)^\top$ and $\mathbf{r} = (r_1, ..., r_{m-1})^\top$.
- Quasi-likelihood function

$$L_n(\theta, \mathbf{r}) = -\frac{1}{2}\sum_{t=1}^{n}\sum_{i=1}^{m}\left(\log \sigma_i^2 + \frac{X_t^2}{\sigma_i^2}\right) I_{it}, \tag{6}$$

  where $I_{it} = I\{r_{i-1} < W_{t-1} \le r_i\}$. For each $\mathbf{r}$, $L_n(\theta, \mathbf{r})$ is maximized at $\hat{\theta}_n(\mathbf{r}) \equiv (\hat{\sigma}_{1n}^2(\mathbf{r}), ..., \hat{\sigma}_{mn}^2(\mathbf{r}))^\top$ with $\hat{\sigma}_{in}^2(\mathbf{r}) = \frac{\sum_{t=1}^n X_t^2 I_{it}}{\sum_{t=1}^n I_{it}}, \quad i = 1, ..., m.$
- The quasi-maximum likelihood estimator (QMLE):

$$\hat{\mathbf{r}}_n = \arg\max_{\mathbf{r}} L_n(\hat{\theta}_n(\mathbf{r}), \mathbf{r})$$
$$\hat{\theta}_n = \hat{\theta}_n(\hat{\mathbf{r}}_n)$$

Assumption

The density $f(\cdot)$ of η_t is continuous and positive on R, $\mathbb{E}\eta_t = 0$ and $\mathbb{E}\eta_t^2 = 1$.

Assumption

The density $f_w(\cdot)$ of W_t is continuous and $f_w(r_{j0}) > 0$ for $j = 1, ..., m-1$.

Theorem

If (i) Assumptions 1 and 2 hold, and (ii) $\sigma_{i0}^2 \neq \sigma_{i+1,0}^2$ for $i = 1, ..., m-1$, then, $(\hat{\theta}_n, \hat{\mathbf{r}}_n) \to (\theta_0, \mathbf{r}_0)$ a.s. as $n \to \infty$.

Theorem

Under the conditions of the preceding theorem, if $\sup_{x\in R}\{(1+|x|)f(x)\} < \infty$ *and* $\kappa_4 \equiv E\eta_t^4 < \infty$, *then*

$$(a) \quad n(\hat{\mathbf{r}}_n - \mathbf{r}_0) = O_p(1);$$

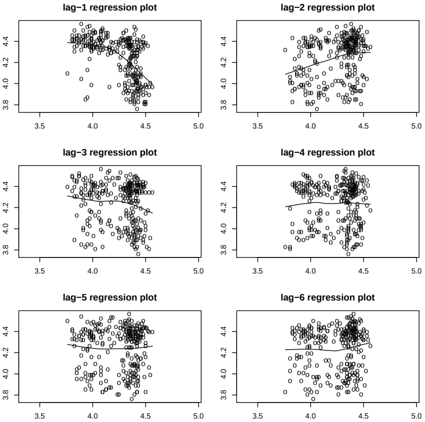
$$(b) \quad \sqrt{n} \sup_{\|\mathbf{r}-\mathbf{r}_0\| \le B/n} |\hat{\sigma}_{in}^2(\mathbf{r}) - \hat{\sigma}_{in}^2(\mathbf{r}_0)| = o_p(1) \quad \text{for any fixed } B \in (0,\infty).$$

Furthermore,

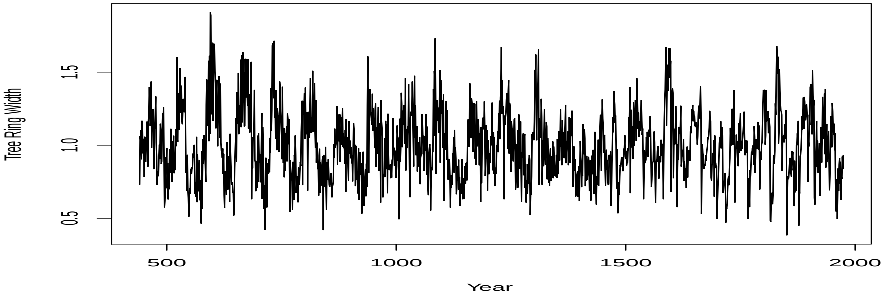
$$\sqrt{n}(\hat{\sigma}_{in}^2(\mathbf{r}_0) - \sigma_{i0}^2) \Longrightarrow \mathcal{N}\left(0, \ \frac{(\kappa_4 - 1)\sigma_{i0}^4}{F_w(r_{i0}) - F_w(r_{i-1,0})}\right), \ i = 1, \dots, m,$$

and all the normalized estimators are asymptotically independent, where $F_w(x)$ *is the cumulative distribution function of* W_t, *and henceforth the symbol* $\Longrightarrow$ *indicates weak convergence.*

- the waiting time is strongly associated with lag 1 values but much less so with values at higher lags;
- the presence of conditional heteroscedasticity;
- a number of outliers.

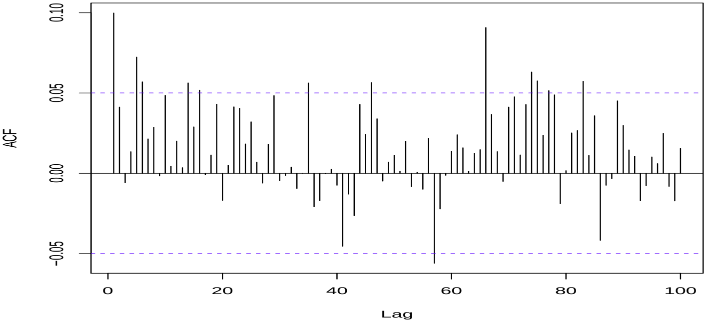


A long time series of annual tree ring width

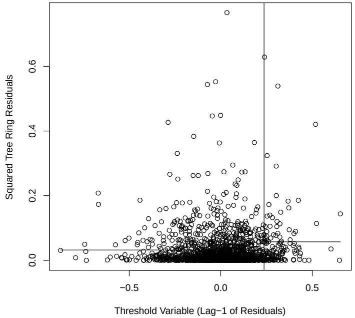


with the measurements taken from a tree in a location at high altitude in Argentina. The time series spans over the period from year 441 to 1974 and it was contributed by J. Boninsegna to the NOAA Paleoclimatology database.

An IMA(1,1) model is initially identified and fitted to the data with the MA coefficient given by -0.6110 with standard error 0.0216. The residuals of the fitted IMA(1,1) model appear to be white noise but not independent as the absolute residuals appear to be correlated.



We fit a two-regime T-CHARM to the residuals to account for the conditional heteroscedasticity, with lag 1 of the IMA(1,1) error as the threshold variable. The following parameter estimates are obtained: $\hat{\sigma}_1^2 = 0.03205(0.00153)$, $\hat{\sigma}_2^2 = 0.05729(0.00893)$, $\hat{r} = 0.2366$ (95% confidence interval: (0.1619, 0.3049)), which is approximately the 91 percentile.



- The threshold structure is supported by the LR test for T-CHARM with p-value $p_0 = 0.005$.
- The first regime contains 1402 observations while the second regime 131 observations.
- There are no residual ARCH effects in the standardized residuals from the fitted T-CHARM, by reference to the McLeod-Li test up to 100 lags.
- *The fitted T-CHARM suggests that during fast-growing years, tree growth is much more variable, with a variance that almost doubles that during non-fast-growing years.*

- In comparison, the IMA(1,1) residuals may be fitted by a GARCH(1,1) model whose conditional variance equals $h_t = \beta h_{t-1} + \alpha_0 + \alpha_1 X_{t-1}^2$ where X_t stands for the IMA(1,1) errors, as this model passes the McLeod-Li test but the simpler ARCH(1) model does not. The GARCH estimates, with their standard errors enclosed in parentheses, are $\hat{\alpha}_0 = 0.0284(0.00656)$, $\hat{\alpha}_1 = 0.0987(0.0288)$ and $\hat{\beta} = 0.0747(0.198)$.
- The simpler ARCH(1) model did not pass the McLeod-Li test.
- It is unclear to us as to how to interpret the fitted GARCH(1,1) model. Finally, both fitted models involve three parameters each with comparable quasi-likelihoods.