TMA4285 TIME SERIES December 2018

① a) A stochastic process is a random variable U_t for each t in an index set I.
A random variable is a function $X : \Omega \to \mathbb{R}$ such that $(X \leq x) = \{\omega \mid X(\omega) \leq x\}$ is an event for all real x. The law of U is determined by the joint CDF $F(u) = P(U_{t_1} \leq u_1, \dots, U_{t_k} \leq u_k)$ for all choices of $t_1 < \dots < t_k$.

b) Let $t_1, \dots, t_n$ be the observation times and assume that there is a joint density $f = F'$. The density is :

$$f(u_1, \dots, u_n) = f_n(u_n \mid u_1, \dots, u_{n-1}) \cdot f(u_1, \dots, u_{n-1})$$

$$\overset{\text{Induction}}{=} \prod_{i=1}^{n} f_i(u_i \mid u_1, \dots, u_{i-1})$$

with $f_1(u_1 \mid \cdot) = f_1(u_1)$. This proves the claim since $L = f$, and all the unknown conditional densities are determined by the model parameter.

1 c) $L = \prod_{i=1}^{n} f_i(u_i)$

d) $f_i(u_i \mid u_1, \cdots, u_{i-1}) = \frac{1}{\sqrt{2\pi}\,\hat{\sigma}_i} e^{-\frac{1}{2}\left(\frac{u_i - \hat{u}_i}{\hat{\sigma}_i}\right)^2}$

where $\hat{u}_i = E(U_{t_i} \mid u_1, \cdots, u_{i-1})$ and $\hat{\sigma}_i^2 = Var(U_{t_i} \mid u_1, \cdots, u_{i-1})$.

e) The maximum likelihood estimate $\hat{\theta}$ is determined by $L(\hat{\theta}) \geq L(\theta)$ $\forall \theta$ if it exists.

f) The conditional expectation is the best L^2 approximation by any (measurable) function which is always as good or better than approximation by a linear function.

g) Let $\hat{V} = \hat{E}(V \mid U_1, \cdots, U_n) = c_1 U_1 + \cdots + c_n U_n$

The projection theorem gives $V - \hat{V} \perp U_i$

or $\langle V, U_i \rangle = \langle \hat{V}, U_i \rangle = \sum_j c_j \langle U_j, U_i \rangle$

or

$$c = \left[Cov(U_i, U_j)\right]^{-1} \cdot \left[Cov(V, U_i)\right]$$

The inverse is any generalized

inverse, $V = U_{t+h}$, and it has been assumed that $E(U_t) = E(V) = 0$. The latter is not reasonable for the given data, and the formula is adjusted to:

$$(*) \quad \hat{E}(U_{n+h} \mid U_1, \dots, U_n) = E(U_{n+h}) + \sum_{i=1}^{n} c_i \cdot (U_i - E(U_i))$$

The previous calculation and linearity of $\hat{E}$ can be used to verify $(*)$.

b) A specific model must be chosen with a corresponding θ. The prototypical example in classical time series is that θ is given by the coefficients in the SARIMA polynomials together with the noise variance. Eq. (1) can then be used to find an estimate θ^*, and hence a corresponding model density

$$f^*(u_1, \dots, u_n, u_*)$$

where u_* is the wine sales in January 1992. This determines the marginal conditional

$$f^*(u_* \mid u_1, \dots, u_n)$$

and $\hat{u}^* = E^*(U_* \mid u_1, \dots, u_n)$ can be

used as a forecast for u^*. The E^* can be replaced by $\hat{E}^*$ for computational simplicity if needed.
The uncertainty can be reported as $\hat{\sigma}^*$ where $\hat{\sigma}_*^2 = Var(U_* \mid u_1, \dots, u_n)$, but this excludes the uncertainty from the estimate θ^*.

(2)

a) The variance increases with time in Fig. 1, and a log transformation will tend to stabilize the variance.

b) Differencing (one or more times) with lag s will tend to remove seasonality with period s. In this case the 12 month seasonality is removed by one differencing at lag 12 as can be seen from figure 2.

c) $\varphi(B) X_t = \theta(B) Z_t \quad (**)$

where $Z_t \sim WN(0, \sigma^2)$ and φ & θ are polynomials of degree p & q respectively with no common roots, $\varphi(0) = \theta(0) = 1$ and $\varphi(z) \neq 0$ on $|z| = 1$.

d) The first step is to compute $\gamma(t-s) = E(X_t X_s)$ from φ, θ & σ. This can be done by establising a finite difference equation for γ from $(**)$ by $E(X_{t-h} \cdot (**))$. From this the joint covariance matrix of $(X_1, \dots, X_n)$ is determined, and hence the likelihood.

The calculation can be simplified by a calculation of the conditional densities $f_i(x_i \mid x_1, \dots, x_{i-1})$ given by the means $\hat{x}_i = \hat{E}(X_i \mid x_1, \dots, x_{i-1})$ and variances $\hat{\sigma}_i^2 = \hat{Var}(X_i \mid x_1, \dots, x_{i-1})$. These can be calculated effectively by the innovations algorithm.

Causality, $\varphi(z) \neq 0$ on $|z| \leq 1$ is assumed here.

e) & f)

$$X_t = \varphi_1 X_{t-1} + \dots + \varphi_p X_{t-p} + Z_t$$

$$\Rightarrow \hat{E}(X_t \mid x_1, \dots, x_{t-1}) \Big\} Z_t \perp X_s = \textstyle\sum_{k \leq s} \psi_k Z_k, \quad s < t$$

$$= \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \dots + \varphi_p X_{t-p}$$

for $t > p$,

and $\alpha(h) = \phi_{hh} = 0$ for $h > p$. This is as in Figure 3 with $p = 12$.

② g) Let every forecast be defined as the best linear prediction based on the observed y_t and x_t values up to T corresponding to October 1991. Forecasts at $T+h$ for $h=1$ (Nov), $h=2$ (Dec), and $h=3$ (Jan 1992) can be computed from:

$$\hat{u}_{T+3} = \exp(\hat{y}_{T+3}) \quad \Big\} \quad e^{\hat{y}_{T+3} + \frac{1}{2}\hat{\sigma}^2_{T+3}}$$

is better; log-normal mean.

$$y_{T+1} = x_{T+1} + c + y_{T-11} \quad ; \quad c = 0{,}0681$$

$$\hat{y}_{T+1} = \hat{x}_{T+1} + c + y_{T-11}$$

$$y_{T+2} = x_{T+2} + c + y_{T-10}$$

$$\hat{y}_{T+2} = \hat{x}_{T+2} + c + y_{T-10}$$

$$\hat{y}_{T+3} = \hat{x}_{T+3} + c + y_{T-9}$$

$$x_{T+1} = \sum_{i=1}^{12} \varphi_i x_{T+1-i} + z_{T+1} \; ; \; \varphi_1 = 0{,}27 ; \; \varphi_5 = 0{,}224 ;$$

$\varphi_8 = 0{,}149$; $\varphi_{11} = -0{,}099$

$\varphi_{12} = -0{,}353$

$\varphi = 0$ otherwise

$$\hat{x}_{T+1} = \sum_i \varphi_i x_{T+1-i}$$

$$\hat{x}_{T+2} = \sum_i \varphi_i \hat{x}_{T+2-i}$$

$$\hat{x}_{T+3} = \sum_i \varphi_i \hat{x}_{T+3-i}$$

$\hat{x}_j = x_j$ if $j \leq T$

The forecast depends hence only on the observed sales from Nov 1989 given by the last two years.
(The estimated φ's depend however on the complete time series.)

(1) a) Causal : Z_t depends only on e_s for $s \leq t$.

Stationary : The joint law of $(Z_{t_1}, \dots, Z_{t_n}) \sim (Z_{t_1+h}, \dots, Z_{t_n+h})$ for all $h \in \mathbb{Z}$ and $t_1, \dots, t_n \in \mathbb{Z}$.

Weakly stationary if :

$\operatorname{Var} Z_t < \infty$

$\operatorname{Cov}(Z_t, Z_s) = \gamma(t-s)$

$E(Z_t) = \mu$

c) $E(Z_t) = E(\sqrt{h_t} \cdot e_t) = E(\sqrt{h_t}\, E(e_t \mid e_s, s<t))$

$= E(\sqrt{h_t} \cdot 0) = 0$

Or : $E(Z_t) = E(\sqrt{h_t}\, e_t) \underset{\text{Independence}}{=} E(\sqrt{h_t}) \cdot E(e_t) = 0$

$E(Z_{t+h} \cdot Z_t) = E(Z_t \sqrt{h_{t+h}}\, E(e_{t+h} \mid e_s, s<t+h))$

$= 0 \quad h > 0$

$\sigma_Z^2 = E(Z_t^2) = \dfrac{\alpha_0}{1-\alpha_1-\beta_1}$ from $E(h_t)$

b) $E(Z_t^2) = \alpha_0 + (\alpha_1 + \beta_1) E(Z_{t-1}^2)$

and stationarity.

③

b) $E(Z_t^2 | Z_s^2, s<t) = E(h_t e_t^2 | e_s, s<t)$

$= h_t \cdot E(e_t^2 | e_s, s<t)$

$= h_t \cdot E(e_t^2) = h_t$

d) From b) it follows that h_{t+1} is the conditional variance of Z_{t+1} given the history up to time t.

e) Serial dependence without correlation (Z ~ WN!) volatility clustering (h_t is ARMA) and tail heaviness ($E(Z_t^k) = \infty$ for some k).