In Brief: Maximum Likelihood Estimation in Time Series

Review of Time Series: Time series is an ordered sequence of observations through time or space, and generally is stochastic (only partly determined by previous values rather than wholly deterministic) which makes exact predictions impossible. Instead the future values in a time series take on a probability distribution conditioned upon knowledge of past values. Since it is the nature of time series for observations to be dependent/correlated and observation order is important, statistical techniques that rely on the assumption of independence are not valid for this type of data.

Introduction to ARMA in Time Series: General time series models contain the idea that an observed time series $X_1, X_2, \dots, X_T$, driven by a stochastic process. For an ARMA (auto regressive moving average mixed model) process the relationship can be described with the equation:

$$\varphi_p(B)X_t = \theta_q(B)a_t \qquad (1)^*$$

where

$$\varphi_p(B) = 1 - \varphi_1 B + \varphi_2 B^2 + \cdots + \varphi_p B^{p\dagger}$$

$$\theta_q(B) = 1 - \theta_1 B + \theta_2 B^2 + \cdots + \theta_p B^{p\ddagger}.$$

For this process to be invertible the roots of $\theta_q(B) = 0$ are required to lie outside the unit circle, whereas for the function to be stationary the roots of $\varphi_p(B) = 0$ also would have to lie outside the unit circle. We also assume that the two equations have no common roots. Equation (1) can be re-written as:

$$X_t = \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \cdots + \varphi_p X_{t-p} + a_t - \theta_1 X_{t-1} a_{t-1} - \theta_2 X_{t-2} a_{t-2} - \cdots - \theta_q X_{t-q} a_{t-q}. \quad (2)$$

Assuming that the model is known for time series, ML estimation is an alternative to least squares estimation. In time series, we can calculate conditional or unconditional ML estimates.

Introduction to Conditional ML Estimation in Time Series[§]: For the ARMA(p,q)[**] model presented in equation (2), where $\dot{X}_t = X_t - \mu$ and $\{a_t\}\ iid\ N(0, \sigma_a^2)$ white noise[††], we can write a joint probability density of $\boldsymbol{a} = (a_1, a_2, \dots, a_n)'$ as:

$$P(\boldsymbol{a}|\boldsymbol{\varphi}, \mu, \boldsymbol{\theta}, \sigma_a^2) = (2\pi\sigma_a^2)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma_a^2}\sum_{t=1}^{n} a_t^2\right]. \qquad (3)$$

Rewritten as

$$a_t = \theta_1 a_{t-1} + \theta_2 a_{t-2} + \cdots + \theta_q a_{t-q} + \dot{X}_t - \varphi_1 \dot{X}_{t-1} - \varphi_2 \dot{X}_{t-2} - \cdots - \varphi_p \dot{X}_{t-p}. \qquad (4)$$

---

[*] The process is also referred to as an ARMA(p,q) process or model n which p and q refer to the orders of the auto regressive and moving average polynomials.

[†] Auto regressive polynomial

[‡] Moving average polynomial

[§] The majority of this information was taken from *Time Series Analysis: Univariate and Multivariate Methods* by William Wei.

[**] Required that ARMA(p,q) be the general, stationary, process.

[††] White noise is another way of saying "error" or unexplained deviations.

Letting $\boldsymbol{X} = (X_1, X_2, \dots, X_n)'$ and assuming initial conditions of $\boldsymbol{X}_* = (X_{1-p}, \dots, X_{-1}, X_0)'$ and $\boldsymbol{a}_* = (a_{1-q}, \dots, a_{-1}, a_0)'$ are known we then have the conditional log likelihood function

$$\ln[L_*(\boldsymbol{\varphi}, \mu, \boldsymbol{\theta}, \sigma_a^2)] = -\frac{n}{2}\ln[2\pi\sigma_a^2] - \frac{S_*(\boldsymbol{\varphi}, \mu, \boldsymbol{\theta})}{2\sigma_a^2} \qquad (5)$$

where

$$S_*(\boldsymbol{\varphi}, \mu, \boldsymbol{\theta}) = \sum_{t=1}^{n} a_t^2\,(\boldsymbol{\varphi}, \mu, \boldsymbol{\theta} | \boldsymbol{X}_*, \boldsymbol{a}_*, \boldsymbol{X}). \qquad (6)^{‡‡}$$

We have a couple alternatives for specifying our initial equations that we conditioned upon. Based on the assumption that $\{X_t\}$ and $\{a_t\}$ is iid $N(0, \sigma_a^2)$ random variables we can replace our unknown X_t with the sufficient statistic $\bar{X}$and the unknown a_t with $E(a_t) = 0$. In model (2) we could also assume that $a_p = a_{p-1} = \cdots = a_{p-(q-1)} = 0$ and calculate a_t for $t \geq (p+1)$ using equation (4). Then the conditional sum of squares S_* reduces to:

$$S_*(\boldsymbol{\varphi}, \mu, \boldsymbol{\theta}) = \sum_{t=1}^{n} a_t^2\,(\boldsymbol{\varphi}, \mu, \boldsymbol{\theta} | \boldsymbol{X})^{§§}$$

Using normal ML estimation we then take the derivatives with respect to $\boldsymbol{\varphi}, \mu, \boldsymbol{\theta}$ and solve for them to get parameter estimates $\widehat{\boldsymbol{\varphi}}, \hat{\mu}, \widehat{\boldsymbol{\theta}}$. To get the estimate $\hat{\sigma}_a^2$ of σ_a^2 we need to calculate

$$\hat{\sigma}_a^2 = \frac{S_*\left(\widehat{\boldsymbol{\varphi}}, \hat{\mu}, \widehat{\boldsymbol{\theta}}\right)}{df} \qquad (7)$$

where $df^{***} = n - m$, where n is the number of terms in the sum of $S_*\left(\widehat{\boldsymbol{\varphi}}, \hat{\mu}, \widehat{\boldsymbol{\theta}}\right)$ and m is the number of parameters estimated.

Unconditional ML estimation in Time Series and the Backcasting Method: For unconditional ML estimation and the backcasting method for time series please refer to *Time Series Analysis: Univariate and Multivariate Methods* by William Wei[†††].

Referenced

Chatfield, Chris. *The Analysis of Time Series: An Introduction*. 6th ed. Boca Raton: CRC, 2004. Print.
Fox, John. *Applied Regression Analysis and Generalized Linear Models.* Los Angeles: Sage, 2008. Print.
Lütkepohl, Helmut. *A New Introduction to Multiple Time Series Analysis*. New York: Springer, 2005. Print.
Stulajter, Frankisek. *Predictions in Time Series Using Regression Model.* New York: Springer, 2002. Print.
Wei, William. *Time Series Analysis: Univariate and Multivariate Methods*. 2th ed. New York: Pearson, 2006. Print.

[‡‡] Equation (6) is the conditional sum of squares function.
[§§] This is the form used in most computer programs.
[***] If we are using equation (7) to calculate the sum of squares $df = (n-p) - (p+q+1) = n - (2p+q+1)$.
[†††] Dr. William Wei also discusses the exact likelihood functions, in the unconditional, conditional, and backcasting methods we use an approximated likelihood function.