

Notes for 751-752

Sections 21

Martin Lindquist*

November 24, 2017

*Thanks to Brian Caffo and Ingo Ruczinski for supplying their notes, upon which much of this document is based.

1 Linear Models with Autocorrelated Errors

When observations are taken serially at equally spaced time periods and the error terms from adjacent time points are correlated, we say that the error term is serially correlated or autocorrelated. A major cause of autocorrelated error is the omission of one or more key explanatory variable in the model. If the time-ordered effects of the missing variable are positively correlated, then the error terms also tend to be positively correlated. This violates the usual assumption of independent errors made in the linear model. Serial correlation will not affect the unbiasedness or consistency of OLS estimators, but it does affect their efficiency. With positive serial correlation, the OLS estimates of the standard errors will be smaller than the true standard errors. This will lead to the erroneous conclusion that the parameter estimates are more precise than they actually are. Therefore there will be a tendency to reject the null hypothesis when it should not be rejected.

To properly understand autocorrelated errors we begin with discussing methods from time series analysis. A time series is a sequence of observations measured sequentially in time. Data from a large number of research areas (e.g., finance, social sciences and medicine) appear as time series. An intrinsic feature of a time series is that observations tend to be dependent and studying the nature of this dependency is of interest.

1.1 Time series analysis

If a random variable X is indexed in time, the observations $\{X_t, t \in T\}$ is called a time series. A time series $X_t, t \in T$ can be regarded as a realization of a stochastic process. We are in particular interested in discrete equally spaced time series. A complete description of a time series, observed as a collection of N random variables, is provided by its joint distribution function.

Often these functions are not easily expressed and instead we concentrate on studying second order properties of the time series. These include:

1. The mean function of X_t is $\mu_X(t) = E(X_t)$.
2. The variance function of X_t is $\sigma_X^2(t) = E(X_t - \mu_X(t))^2$.
3. The autocovariance function of X_t is defined as:

$$\gamma(r, s) = \text{cov}(X_r, X_s) = E((X_r - E(X_r))(X_s - E(X_s)))$$

for $s, t \in T$.

4. The autocorrelation function of X_t is defined as:

$$\rho(r, s) = \frac{\gamma(r, s)}{\sqrt{\gamma(r, r)\gamma(s, s)}}.$$

1.1.1 Stationarity

A central feature in the development of time series models is the assumption of some form of regularity over time in the time series. Often we assume that the behavior of a time series in an interval $[t, t + h)$ closely resembles that in the interval $[s, s + h)$. This implies there is a temporal homogeneity in the behavior of the series, called stationarity.

A time series X_t is strictly stationary if all of its finite dimensional distributions are time invariant. This implies that

$$P(X_{i_1} < \alpha_1, \dots, X_{i_k} < \alpha_k) = P(X_{h+i_1} < \alpha_1, \dots, X_{h+i_k} < \alpha_k)$$

for any collection of time points $i_1 \dots i_k$, for any positive integer k and integer-valued lag h .

A time series X_t is weakly stationary if

1. $E|X_t|^2 < \infty$ for all t .
2. The mean function $\mu_X(t)$ does not depend on t .
3. The covariance function $\gamma_X(t, t + h)$ is independent of t for all h .

If X_t is a weakly stationary time series then the autocovariance function (ACVF) at lag h can be written:

$$\gamma(h) = \text{cov}(X_{t+h}, X_t)$$

When $h = 0$, we have that $\gamma_X(0) = \text{var}(X_t)$. Hence, the autocorrelation function (ACF) of X_t at lag h can be written:

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)}.$$

A strictly stationary process with finite second moment is also weakly stationary. However, the converse is not necessarily true. An exception is that if X_t is a weakly stationary Gaussian process then it is also strictly stationary. Stationary processes play an important roll in the analysis of time series. In reality, many observed time series are non-stationary. They evolve over time and have trends in their mean or variance. Transformations such as differencing or detrending can often be used to make them stationary.

1.1.2 Time series models

There exist a number of commonly used time series models. These include autoregressive (AR) processes, moving-average (MA) processes, and autoregressive moving average (ARMA) processes, We begin by discussing white noise processes which are the building block of the processes mentioned above.

A sequence of uncorrelated random variables Z_t , each with mean 0 and variance σ^2 , is called white noise. This is written: $Z_t \sim WN(0, \sigma^2)$. In Gaussian white noise, the Z_t are independent identically distributed

(i.i.d.) normal random variables with mean 0 and variance σ^2 . A white noise process Z_t has the following properties:

$$E(Z_t) = 0 \quad \forall t,$$

$$Var(Z_t) = \sigma^2 \quad \forall t$$

and

$$\rho(h) = \begin{cases} 1, & \text{if } h = 0 \\ 0, & \text{if } h \neq 0 \end{cases}$$

A white noise process is weakly stationary (but not necessarily strictly stationary, since the Z_i 's are not necessarily identically distributed). An i.i.d. sequence of random variables is white noise if it has a finite variance. White noise processes are the fundamental building blocks of weakly stationary processes and play a crucial role in time series analysis.

A time series X_t is an autoregressive process of order p , written AR(p):

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + Z_t$$

where $Z_t \sim WN(0, \sigma^2)$ and $\phi_1, \phi_2, \dots, \phi_p$ are constants. An AR(p) process regresses X_t on its past values.

A special case is the AR(1) process:

$$X_t = \phi X_{t-1} + Z_t$$

where $Z_t \sim WN(0, \sigma^2)$ and $|\phi| < 1$. The AR(1) process has the following properties:

$$E(X_t) = 0 \quad \forall t,$$

$$Var(X_t) = \frac{\sigma^2}{1 - \phi^2} \quad \forall t$$

and

$$\rho(h) = \begin{cases} 1, & \text{if } h = 0 \\ \phi^{|h|}, & \text{if } h \neq 0 \end{cases}$$

A time series is a moving-average process of order q , written MA(q), if

$$X_t = Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \cdots + \theta_q Z_{t-q}$$

where $Z_t \sim WN(0, \sigma^2)$ and $\theta_1, \theta_2, \dots, \theta_q$ are constants.

A special case is the MA(1) process:

$$X_t = Z_t + \theta Z_{t-1}$$

where $Z_t \sim WN(0, \sigma^2)$ and θ are constants. The MA(1) process has the following properties:

$$E(X_t) = 0 \quad \forall t,$$

$$Var(X_t) = (1 + \theta^2)\sigma^2 \quad \forall t$$

and

$$\rho(h) = \begin{cases} 1, & \text{if } h = 0 \\ \frac{\theta}{1+\theta^2}, & \text{if } h = \pm 1 \\ 0 & \text{if } |h| > 2 \end{cases}$$

A time series is an autoregressive moving-average process, written ARMA(p,q), if

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + Z_t + \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + \dots + \theta_q Z_{t-q}$$

where $Z_t \sim WN(0, \sigma^2)$ and $\phi_1, \phi_2, \dots, \phi_p, \theta_1, \theta_2, \dots, \theta_q$ are constants.

1.1.3 Analyzing time series

Given a set of observations from a stationary time series, the goal of time series analysis is to find an appropriate model to represent the observed data. Important issues involve: (i) model selection; (ii) order selection; and (iii) estimation of the model parameters.

Candidate models can be identified studying the ACF and other related statistics. Model and order selection can be performed using information criteria methods. The parameters of an AR(p) model can be estimated using Yule-Walker estimates (method of moments). The parameters of an ARMA(p,q) model can be estimated using maximum likelihood or restricted maximum likelihood methods.

1.1.4 Estimating time series

From the observations $\{X_1, X_2, \dots, X_n\}$ of a stationary time series X_t we often seek to estimate the autocovariance function $\gamma(\cdot)$. The sample autocovariance function is defined by

$$\hat{\gamma}(h) = \frac{1}{n} \sum_{j=1}^{n-h} (x_{j+h} - \bar{x})(x_j - \bar{x})$$

for $0 \leq h \leq n$.

Let us illustrate the method of moments approach by assuming an AR(2) process:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + Z_t$$

We seek to estimate the parameters ϕ_1, ϕ_2, σ^2 using the method of moments.

Begin by multiplying the process by X_{t-k} for $k = 1, 2$ and take the expectation. Thereafter divide by $\gamma(0)$. This gives the following set of equations:

$$\begin{aligned}\rho(1) &= \phi_1 + \phi_2\rho(1) \\ \rho(2) &= \phi_1\rho(1) + \phi_2\end{aligned}$$

These are the Yule-Walker estimates.

Now equate the sample and population moments and solve these equations. This gives the following estimates:

$$\begin{aligned}\hat{\phi}_1 &= \frac{\hat{\rho}(1)(1 - \hat{\rho}(2))}{1 - \hat{\rho}(1)^2} \\ \hat{\phi}_2 &= \frac{\hat{\rho}(2) - \hat{\rho}(1)^2}{1 - \hat{\rho}(1)^2}.\end{aligned}$$

To estimate σ^2 , multiply the process by X_t and take the expectation. This gives the following:

$$\gamma(0) = \phi_1\gamma(1) + \phi_2\gamma(2) + \sigma^2$$

Solving for σ^2 we obtain the following estimate:

$$\hat{\sigma}^2 = \hat{\gamma}(0)(1 - \hat{\phi}_1\hat{\rho}(1) - \hat{\phi}_2\hat{\rho}(2)).$$

1.2 Time series regression

There are a number of problems that may arise if serial correlation is ignored.

1. The estimated regression coefficients will still be unbiased, but no longer minimum variance.
2. Estimates of σ^2 will be biased.
3. The variance of β will be underestimated and resulting t-statistics will be inflated.
4. Tests using the t and F distributions may not be applicable.

When working with time series data, we typically use the index t to indicate the temporal ordering of observations. Throughout, we assume observations are measured at equally spaced time periods. A simple linear regression model for time series data is given by:

$$y_t = \beta_0 + \beta_1 X_t + \epsilon_t \quad t = 1, \dots, n$$

We need to construct a model for ϵ_t that can account for autocorrelation. Commonly used models include autoregressive (AR), moving average (MA) and autoregressive-moving average (ARMA) models. In the continuation we use a first order autoregressive model, but it is important to note that the methods can be applied more generally.

Consider the model: $\mathbf{y} = \mathbf{X}\beta + \epsilon$ with $\epsilon \sim N(\mathbf{0}, \Sigma)$. Here we can write:

$$\Sigma = \frac{\sigma^2}{1 - \phi^2} \begin{pmatrix} 1 & \phi & \phi^2 & \dots & \phi^n \\ \phi & 1 & \phi & \dots & \vdots \\ \phi^2 & \phi & 1 & \dots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \phi \\ \phi^n & \phi^{n-1} & \dots & \phi & 1 \end{pmatrix}$$

When Σ is known we can estimate β using generalized least-squares. In general the form of the variance-covariance matrix is unknown, which means it has to be estimated. Estimating Σ depends on knowing β , and estimating β depends on knowing Σ . We need to use an iterative procedure, such as the Cochrane-Orcutt Procedure.

Cochrane-Orcutt Procedure:

1. Assume that $\Sigma = \mathbf{I}\sigma^2$ and calculate the OLS solution.
2. Estimate the parameters of the time series model from the residuals.
3. Re-estimate the β values using the estimated covariance matrix from step 2.
4. Iterate until convergence.

1.3 Durbin-Watson test

The Durbin-Watson test allows us to test whether first order autocorrelation is present in the residuals. Because most regression problems involving time series data exhibit positive autocorrelation, the hypotheses that are usually considered are

H_0 : No residual correlation (i.e. $\rho = 0$)

H_a : Positive residual correlation (i.e. $\rho > 0$)

The test proceeds as follows. First, compute the OLS solution and obtain the residuals. Second, compute the Durbin-Watson statistic:

$$d = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2}$$

where $e_i = y_i - \hat{y}_i$.

Small values (between 0 and 1) indicate that we can reject H_0 . It can be shown that the Durbin-Watson statistic can be written: $d \approx 2(1 - r)$, where r is the sample autocorrelation of the residuals. Hence, d takes values between 0 and 4. Values of d around 2 indicate no correlation. Small values indicate positive correlation, while large values indicate negative correlation.