Basics of Time Series
A Summary*

Andrea Luksch, Verena Peternell, Simon Weidenholzer

April 2004

Abstract

The aim of this summary is to provide a basic overview of time series analysis. In a first step, some properties of time series and time series operators will be discussed. Secondly, the concepts of moving average processes and autocorrelated processes will be introduced and reviewed. A section about theoretic concepts of time series forecasting, such as forecastebility , unpredicatability or informativness, concludes.

1 Lag Operators

A time series operator transforms one group of time series into a new time series. An input series such as $\{x_t\}_{t=-\infty}^{\infty}$ or a group of input series such as $(\{x_t\}_{t=-\infty}^{\infty}, ..., \{w_t\}_{t=-\infty}^{\infty})$ is transferred into a new sequence $\{y_t\}_{t=-\infty}^{\infty}$. Examples of time series operators include the multiplication operator, represented as

$$y_t = \beta x_t$$

or the addition operator

$$y_t = x_t + w_t.$$

*Written in partial fullfillment of the requirements of the course "406347 UK Ökonometrische Prognose". The first six chapters follow Hamilton (1994) and Dinardo and Johntson (1997). The last chapter follows Clements and Hendry (1998).

Note that it does not matter if one applies the multiplication operator first and the addition multiplicator second or vice versa, i.e.

$$\beta x_t + \beta w_t = \beta(x_t + w_t).$$

A very useful time series operator is the lag operator. The lag operator creates a new series such that the values of the new series $\{y_t\}_{t=-\infty}^{\infty}$ at t are equal to the values of $\{x_t\}_{t=-\infty}^{\infty}$ at $t-1$. This operation is usually represented by the symbol L.

$$Lx_t = x_{t-1}.$$

If the lag operator is applied twice then the values of the new series $\{y_t\}_{t=-\infty}^{\infty}$ at t are equal to the values of $\{x_t\}_{t=-\infty}^{\infty}$ at $t-2$. Such a double application of the lag operator is often indicated by L^2 :

$$L^2 x_t = L(Lx_t) = Lx_{t-1} = x_{t-2}.$$

More generally, for any integer k , L^k gives the values of the new series $\{y_t\}_{t=-\infty}^{\infty}$ at t , which are equal to the values of $\{x_t\}_{t=-\infty}^{\infty}$ at $t-k$:

$$L^k x_t = x_{t-k}.$$

Note that the lag operator is commutative, i.e. it does not matter if we apply the lag operator first and then the multiplication operator or if we apply the multiplication operator first and then the lag operator. In other words

$$\beta(Lx_t) = L(\beta x_t).$$

An additional property of the lag operator is that it is distributive over the addition operator, i.e.:

$$L(x_t + y_t) = Lx_t + Ly_t.$$

Hence the lag operator follows the same algebraic rules as the multiplication operator. For this reason the expression "multiply y_t by L" instead of "operate $\{y_t\}_{t=-\infty}^{\infty}$ on L" is very often used.

2 Difference equations

2.1 First order difference equations

A first order difference equation is a dynamic equation relating the values y takes on at date t to another variable w_t and to the value y took in the previous period $t-1$.

$$y_t = \phi y_{t-1} + w_t.$$

This equation can be rewritten using the lag operator:

$$y_t = \phi L y_t + w_t.$$

One can rearrange terms such that

$$y_t(1 - \phi L) = w_t. \tag{1}$$

If one multiplies both sides of (1) by $(1 + \phi L + \phi^2 L^2 + ... + \phi^t L^t)$ it is obtained that

$$(1 + \phi L + \phi^2 L^2 + ... + \phi^t L^t) y_t (1 - \phi L) = (1 + \phi L + \phi^2 L^2 + ... + \phi^t L^t) w_t. \tag{2}$$

Now the left side of (2) can be rewritten as

$$[(1 + \phi L + \phi^2 L^2 + ... + \phi^t L^t) - (1 + \phi L + \phi^2 L^2 + ... + \phi^t L^t)\phi L] y_t$$

or equivalently

$$\begin{aligned} &[(1 - \phi^{t+1} L^{t+1}) - (\phi L + \phi^2 L^2 + ... + \phi^{t+1} L^{t+1})] y_t \\ &= (1 - \phi^{t+1} L^{t+1}) y_t. \end{aligned} \tag{3}$$

Substituting (3) into (2) yields:

$$(1 - \phi^{t+1} L^{t+1}) y_t = (1 + \phi L + \phi^2 L^2 + ... + \phi^t L^t) w_t,$$

which in turn can written out explicitly:

$$y_t - \phi^{t+1} y_{t-t-1} = w_t + \phi w_{t-1} + \phi^2 w_{t-2} ... + \phi^t w_{t-t}$$

or

$$y_t = \phi^{t+1} y_{-1} + w_t + \phi w_{t-1} + \phi^2 w_{t-2} ... + \phi^t w_0. \tag{4}$$

Thus the value that y takes on at date t can be described as a function of its initial value y_{-1} and the history of w between date 0 and date t. Another way of obtaining this result is by "recursive substitution" where one substitutes y_{-1} into the equations for $y_0, y_1, ...y_t$ and recursively solves for y_t.

If $|\phi| < 1$ and y_{-1} is a finite number the expression $\phi^{t+1}y_{-1}$ will become negligeable as t becomes large. Hence from (3) and the left side of (1) it is obtained that:

$$(1 + \phi L + \phi L^2 + ...\phi L^t)(1 - \phi L)y_t \cong y_t \text{ for } t \text{ large.}$$

In particular this implies that the operator $(1 + \phi L + \phi L^2 + ... + \phi L^j)$ can be approximated by the inverse of the operator $(1 - \phi L)$. This approximation can be made arbitrary accurate by choosing j sufficiently large, i.e.

$$(1 - \phi L)^{-1} = \lim_{j\to\infty} (1 + \phi L + \phi L^2 + ... + \phi L^j).$$

Note that this operator has the property

$$((1 - \phi L)^{-1}(1 - \phi L) = 1.$$

where "1" refers to the identity operator. If $|\phi| < 1$ and the sequence both sides of (1) can be "divided" by $(1 - \phi L)$ in order to obtain

$$y_t = (1 - \phi L)^{-1} w_t$$

or equivalently

$$y_t = w_t + \phi w_{t-1} + \phi^2 w_{t-2} + \phi^3 w_{t-3} + ... \tag{5}$$

Hence in the long run the y_t will only depend on the sequence w.

2.2 Dynamic multipliers

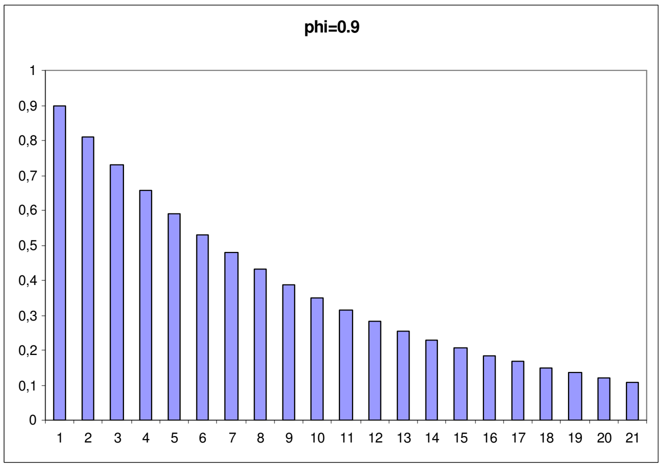
Looking at (4) it is easy to ascertain the effect of an increase in w_0 on y_t:

$$\frac{\partial y_t}{\partial w_0} = \phi^j.$$

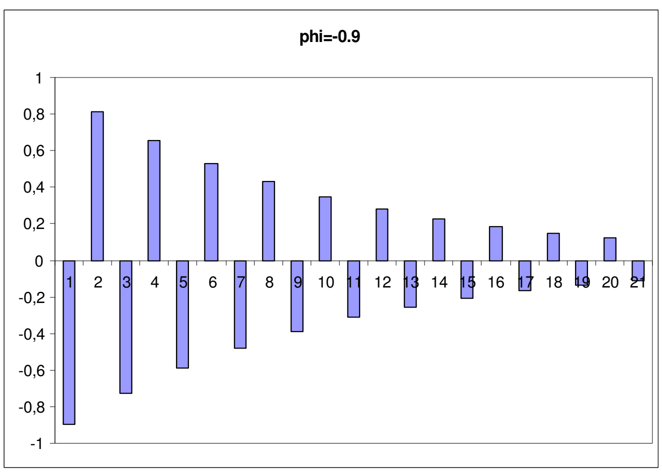
Likewise the effect of w_t on y_{t+j} is given by:

$$\frac{\partial y_{t+j}}{\partial w_t} = \phi^j. \tag{6}$$

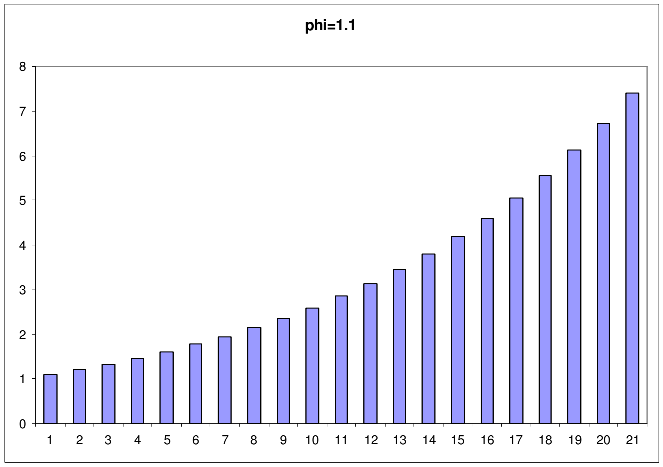
Thus the dynamic multiplier only depends on j, the length of time separating the disturbance to the input w_t and the observed value of the output y_t Different values of ϕ can cause different dynamic responses of y to w. If $0 < \phi < 1$ the multiplier in (6) decays geometrically towards zero.



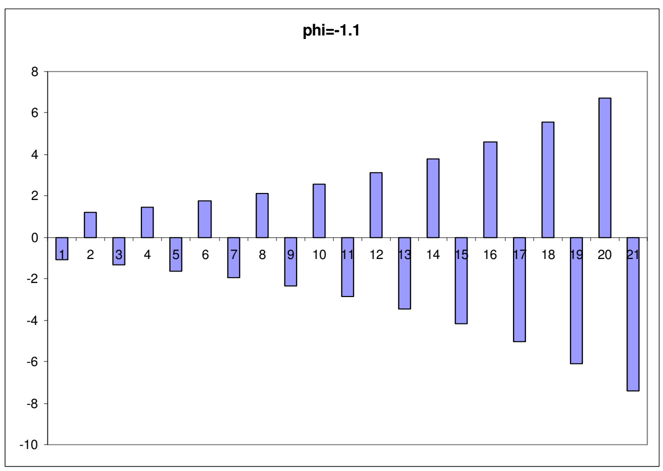
If $-1 < \phi < 0$ the multiplier will alternate in sign. However the absolute value of the effect will decrease towards zero.



If however $\phi > 1$ the multiplier will exponentially increase over time.



Finally, $\phi < -1$ implies that the multiplier will exhibit explosive oscillation.



Thus, if $|\phi| < 1$, the system is stable, i.e. any given change of w will die out. Whereas if $|\phi| > 1$ the system is explosive.

2.3 pth order difference equations

Consider now a difference equation of pth order. A pth order difference equation is a dynamic equation relating the values y takes on at date t to another variable w_t and to the value y took in the previous periods $t-p$.

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_p y_{t-p} + w_t. \tag{7}$$

The aim now is to find an expression similar to (5) such that y_t only depends on the sequence w. Note that (7) can be rewritten using lag operators:

$$(1 - \phi_1 L - \phi_2 L^2 - ... - \phi_p L^p) y_t = w_t.$$

Consider now the operator on the left side. This operator can be factored such that:

$$(1 - \phi_1 L - \phi_2 L^2 - ... - \phi_p L^p) = (1 - \lambda_1 L)(1 - \lambda_2 L)...(1 - \lambda_p L).$$

Since the lag operator follows the same algebraic rules as the multiplication operator this is the same task as finding the values of $(\lambda_1, \lambda_2, ..., \lambda_p)$ such that the polynomials are the same for all z:

$$(1 - \phi_1 z - \phi_2 z^2 - ... - \phi_p z^p) = (1 - \lambda_1 z)(1 - \lambda_2 z)...(1 - \lambda_p z).$$

Now we multiply this equation by z^{-p} :

$$(z^{-p} - \phi_1 z^{1-p} - \phi_2 z^{2-p} - ... - \phi_p z^{p-p}) = (1 - \lambda_1 z)(1 - \lambda_2 z)...(1 - \lambda_p z) z^{-p}.$$

Upon setting $\lambda \equiv z^{-1}$ it is obtained:

$$(\lambda^p - \phi_1 \lambda^{p-1} - \phi_2 \lambda^{p-1} - ... - \phi_p) = (\lambda - \lambda_1)(\lambda - \lambda_2)...(\lambda - \lambda_p).$$

Note that for $\lambda = \lambda_1, \lambda_2, ...,$ or λ_p the right side of this expression becomes zero. Thus the values of $(\lambda_1, \lambda_2, ..., \lambda_p)$ must be such that the right side of this expression is zero as well:

$$\lambda^p - \phi_1 \lambda^{p-1} - \phi_2 \lambda^{p-1} - ... - \phi_p = 0. \tag{8}$$

If the roots of (8) lie inside the unit circle the difference equation will in fact turn out to be stable. Furthermore, provided that the values $(\lambda_1, \lambda_2, ..., \lambda_p)$ are all distinct, this allows us to write equation (7) such that y_t only depends on the sequence w:

$$y_t = \psi_0 w_t + \psi_1 w_{t-1} + \psi_2 w_{t-1}... \tag{9}$$

where

$$\psi_j = [c_1\lambda_1^j + c_1\lambda_2^j + ... + c_k\lambda_k^j + ... + c_p\lambda_p^j]$$

and

$$c_k = \frac{\lambda_k^{p-1}}{\prod\limits_{i\neq k}(\lambda_k - \lambda_i)}.$$

Again, the dynamic multipliers can be directly read off (9):

$$\frac{\partial y_{t+j}}{\partial w_j} = \psi_j = [c_1\lambda_1^j + c_1\lambda_2^j + ... + c_k\lambda_k^j + ... + c_p\lambda_p^j].$$

3 Expectation, Stationarity, and Ergodicity

3.1 Expectations

Consider a battery of I computers separately generating realizations of a time series $\{y_t^{(1)}\}_{t=-\infty}^{\infty}, \{y_t^{(2)}\}_{t=-\infty}^{\infty}, ..., \{y_t^{(I)}\}_{t=-\infty}^{\infty}$. Now one can select from each sequence the observation associated with time t.

$$\{y_t^{(1)}, y_t^{(2)}, ..., y_t^{(I)}\}.$$

This describes a sample of I realizations of the random variable Y_t. The expectation of the tth observation of a time series refers to the mean this probability distribution, provided that it exists:

$$E(Y_t) = \int\limits_{-\infty}^{\infty} y_t f_{y_t}(y_t) dy_t.$$

Sometimes the expectation $E(Y_t)$ is referred to as unconditional mean of Y_t, which is denoted by μ_t. One can also think of $E(Y_t)$ as the limit probability of the ensemble average:

$$E(Y_t) = p \lim_{I\to\infty} \frac{1}{I}\sum_{i=1}^{I} Y_t^{(i)}.$$

The variance of the random variable Y_t which is denoted by γ_{0t} is defined as:

$$\gamma_{0t} = E(Y_t - \mu_t)^2 = \int\limits_{-\infty}^{\infty} (y_t - \mu_t) f_{y_t}(y_t) dy_t.$$

3.2 Autocovariances

Given a realization of the a time series process $\{y_t^{(1)}\}_{t=-\infty}^{\infty}$, one can construct a vector $x_t^{(1)}$ associated with date t which collects the $[j+1]$ most recent observations:

$$x_t^{(1)} \equiv \begin{pmatrix} y_t^{(1)} \\ y_{t-1}^{(1)} \\ \\ y_{t-j}^{(1)} \end{pmatrix}$$

Each realization of $\{y_t\}_{t=-\infty}^{\infty}$ creates one particular vector x_t. The distribution of this vector $x_t^{(i)}$ across realizations i is called the joint distribution of $(Y_t, Y_{t-1}, ...Y_{t-j})$ and can be used to calculate the jth autocovariance of Y_t, denoted γ_{jt}.

$$\gamma_{jt} = E(Y_t - \mu_t)(Y_{t-j} - \mu_{t-j}). \tag{10}$$

Note that (10) has the form of a covariance between two variables. Hence the autocovariance γ_{jt} is the covariance of Y_t with its own lagged value. Again the jth autcovariance can be thought of as the probability limit of an ensemble average:

$$E(Y_t) = p\lim_{I\to\infty} \frac{1}{I} \sum_{i=1}^{I} (Y_t^{(i)} - \mu_t)(Y_{t-j}^{(i)} - \mu_{t-j})$$

3.3 Stationarity

If neither the mean μ_t nor the autocovariances γ_{it} of a process depend on the date t then the process for Y_t is said to be covariance-stationary or weakly stationary.

$$\begin{aligned} E(Y_t) &= \mu \text{ for all } t \\ E(Y_t - \mu)(Y_{t-j} - \mu) &= \gamma_j \text{for all } t \text{ and all } j. \end{aligned}$$

A different concept of stationarity is that of strict stationarity. A process is said to be strictly stationary if for any values of $j_i, j_2, ..., j_n$ the joint distribution of $(Y_t, Y_{t+j_1}, Y_{t+j_2}, ..., Y_{t+j_{t+n}})$ depend only on the intervals separating the dates $(j_i, j_2, ..., j_n)$ and not the date t itself. To see why the concept of strict stationarity is stronger than the concept of covariance-stationary, note that it can be the case that, although the autocovariances and the mean are independent of the date t , higher moments of the distribution depend on t.

3.4 Ergodicity

Assume that we have observed a time series $(y_1^{(1)}, y_2^{(1)}, ..., y_T^{(1)})$. From these observations one can calculate the sample mean $\overline{y}$:

$$\overline{y} = \frac{1}{T}\sum_{t=1}^{T} y_t^{(1)}$$

In this context, a covariance-stationary process is said to be ergodic for the mean if $\overline{y}$ converges in probability to $E(Y_t)$ as $T \to \infty$.

4 Moving Average Processes

4.1 The First-Order Moving Average Process

Consider a process

$$Y_t = \mu + \varepsilon_t + \theta\varepsilon_{t-1}$$

Where μ and θ could be any constants. It is called a "first-order moving average process" or short $MA(1)$ and it is constructed from a weighted sum of the two most recent values of ε (white noise).

The expectation or mean of Y_t is

$$E(Y_t) = E(\mu + \varepsilon_t + \theta\varepsilon_{t-1}) = \mu + E(\varepsilon_t) + \theta E(\varepsilon_{t-1}) = \mu$$

The variance is given by:

$$\gamma_0 = E(Y_t - \mu)^2 = E(\varepsilon_t + \theta\varepsilon_{t-1})^2 = (1 + \theta^2)\sigma^2$$

and the first autocovariance is

$$\gamma_1 = E(Y_t - \mu)(Y_{t-1} - \mu) = \sigma^2\theta.$$

All higher autocovariances γ_j are zero for j>1 and hence the variance and the autocovariance are not functions of time which implies that a $MA(1)$ process is covariance-stationary regardless of the value of θ. Furthermore, the $MA(1)$ is ergodic for all moments since $\{\varepsilon_t\}$ is white noise.

The jth autocorrelation is defined as

$$corr(Y_t, Y_{t-j}) = \rho_j \equiv \frac{\gamma_j}{\gamma_0}.$$

For the first autocorrelation, this means

$$\rho_1 = \frac{\theta\sigma^2}{(1+\theta^2)\sigma^2} = \frac{\theta}{(1-\theta^2)}$$

This indicates that different values of θ have different effects on the correlation type. If θ >0, it induces positive correlation, meaning that a high Y_t is likely to be followed by a very high Y_{t+1}. The opposite is correct for negative values of θ.

4.2 The qth-order Moving Average Process

A $MA(q)$ process is characterized by

$$Y_t = \mu + \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + ... + \theta_q\varepsilon_{t-q}$$

Where $\{\varepsilon_t\}$ is white noise and $(\theta_1, ..., \theta_q)$ are any real numbers.

The expectation or mean is given by

$$E(Y_t) = \mu.$$

The variance of a $MA(q)$ is

$$\gamma_0 = E(Y_t - \mu)^2 = \sigma^2 + \theta_1^2\sigma^2 + ... + \theta_q^2\sigma^2 = (1 + \theta_1^2 + ... + \theta_q^2)\sigma^2$$

For $j = 1, 2, ..., q$, the autocovariance is

$$\begin{aligned}
\gamma_j &= E\left[(\varepsilon_t + \theta_1\varepsilon_{t-1} + ... + \theta_q\varepsilon_{t-q})\text{x}(\varepsilon_{t-j} + \theta_1\varepsilon_{t-j-1} + ... + \theta_q\varepsilon_{t-j-q})\right] \\
&= E\left[\theta_j\varepsilon_{t-j}^2 + \theta_{j+1}\theta_1\varepsilon_{t-j-1}^2 + ... + \theta_q\theta_{q-j}\varepsilon_{t-q}^2\right].
\end{aligned}$$

Terms involving $\varepsilon`s$ at different points in time drop out because their products are zero.

$$\gamma_j = \left[\theta_j + \theta_{j+1}\theta_1 + ... + \theta_q\theta_{q-j}\right]\sigma^2 \ for\ j = 1, ..., q$$
$$0\ for\ j > q.$$

The $MA(q)$ process is covariance-stationary and ergodic for all moments.

4.3 The Infinite-Order Moving Average Process

A $MA(q)$ process can be rewritten:

$$Y_t = \mu + \sum_{j=0}^{q} \theta_j \varepsilon_{t-j}$$

If q goes to infinity we obtain a $MA(\infty)$ process:

$$Y_t = \mu + \sum_{j=0}^{\infty} \varphi_j \varepsilon_{t-j} = \mu + \varphi_0 \varepsilon_t + \varphi_1 \varepsilon_{t-1} + ...$$

This infinite sequence is a covariance-stationary process if it is square summable:

$$\sum_{j=0}^{\infty} \varphi_j^2 < \infty$$

or absolutely summable (stronger condition)

$$\sum_{j=0}^{\infty} \left|\varphi_j\right| < \infty$$

The mean of a $MA(\infty)$ process is given by:

$$E(Y_t) = \lim_{T \to \infty} E(\mu + \varphi_0 \varepsilon_t + \varphi_1 \varepsilon_{t-1} + ... + \varphi_{t-T}) = \mu.$$

The variance is given by:

$$\gamma_0 = E(Y_t - \mu)^2 = \lim_{T \to \infty} (\varphi_0^2 + \varphi_1^2 + ... + \varphi_T^2)\sigma^2,$$

and the autocovariances are given by:

$$\gamma_j = E(Y_t - \mu)(Y_{t-j} - \mu) = \sigma^2(\varphi_j \varphi_0 + \varphi_{j+1}\varphi_1 + \varphi_{j+2}\varphi_2 + ...)$$

If a $MA(\infty)$ process is absolutely summable, it is ergodic for the mean and ergodic for all moments when $\{\varepsilon_t\}$ is white noise.

5 Autoregressive Processes

5.1 The First-Order Autoregressive Process

An $AR(1)$ satisfies the following difference equation with the input variable $w_t = c + \varepsilon_t$:

$$Y_t = c + \phi Y_{t-1} + \varepsilon_t$$

If $|\phi| \geq 1$, the consequences of the ε's for Y accumulate rather than die out over time. This implies that there is a covariance-stationary process for Yt when $|\phi| < 1$. There are two possibilities of deriving the moments for an $AR(1)$. One is by viewing it as an $MA(\infty)$. The second alternative is the assumption that the process is covariance-stationary and we can calculate the moments directly from the difference equation. I will consider the first proposal.

$$\begin{aligned} Y_t &= (c + \varepsilon_t) + \phi(c + \varepsilon_{t-1}) + \phi^2(c + \varepsilon_{t-2}) + ... \\ Y_t &= \left[\frac{c}{(1-\phi)}\right] + \varepsilon_t + \phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + ... \end{aligned}$$

Now, take the expectation to get the mean

$$E(Y_t) = \frac{c}{(1-\phi)} + 0 + 0 + ... = \frac{c}{(1-\phi)} = \mu$$

The variance is

$$\gamma_0 = E(Y - \mu)^2 = E(\varepsilon_t + \phi\varepsilon_{t-1} + ...)^2 = (1 + \phi^2 + \phi^4 + ...)\sigma^2$$

$$\gamma_0 = \frac{\sigma^2}{(1-\phi^2)}$$

The jth autocovariance is

$$\gamma_j = E(Y_t - \mu)(Y_{t-1} - \mu) = \left[\frac{\phi^j}{(1-\phi^2)}\right]\sigma^2$$

It follows that autocorrelation looks like

$$\rho_j = \frac{\gamma_j}{\gamma_0} = \phi^j$$

This is equivalent to the dynamic multiplier or impulse-response function mentioned earlier. The effect of a one-unit increase in on Y_{t+j} is equal to the correlation between Y_t and Y_{t+j}. If $\phi > 0$, we have positive correlation. On the other hand, if $\phi < 0$, this indicates negative first-order but positive second-order autocorrelation.

5.2 The Second-Order Autoregressive Process

An $AR(2)$ process satisfies:

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t.$$

Expressed in Lag operator notation:

$$(1 - \phi_1 L - \phi_2 L^2) Y_t = c + \varepsilon_t.$$

The difference equation is stable if $|\phi|<1$ and the roots lie outside the unit circle. When this is satisfied, the $AR(2)$ process is covariance-stationary and the operator is invertible:

$$\Psi(L) = (1 - \phi_1 L - \phi_2 L^2)^{-1} = \Psi_0 + \Psi_1 L + \Psi_2 L^2 + \Psi_3 L^3 + ...$$

Now, multiply the Lag operator by $\Psi(L)$:

$$\begin{aligned} Y_t &= \Psi(L)c + \Psi(L)\varepsilon_t \\ &\Rightarrow \Psi(L)c = \frac{c}{(1 - \phi_1 - \phi_2)} \\ &\Rightarrow \sum_{j=0}^{\infty} |\Psi_j| < \infty \end{aligned}$$

We obtain a $MA(\infty)$ process, so the mean is given by the constant term

$$\mu = \frac{c}{(1 - \phi_1 - \phi_2)}$$

Following, rewrite the difference equation in order to find second moments:

$$Y_t = \mu(1 - \phi_1 - \phi_2) + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t$$

Multiply both sides by $(Y_{t-j} - \mu)$ and take expectations to get the autocovariances:

$$\gamma_j = \phi_1 \gamma_{j-1} + \phi_2 \gamma_{j-2}$$

To obtain the autocorrelations, divide this expression by the variance γ_0:

$$\rho_j = \phi_1 \rho_{j-1} + \phi_2 \rho_{j-2}$$

The variance is

$$\begin{aligned}
E(Y_t-\mu)^2 &= \phi_1 E(Y_{t-1}-\mu)(Y_t-\mu)+\phi_2(Y_{t-2}-\mu)(Y_t-\mu)+E(\varepsilon_t)(Y_t-\mu) \\
\gamma_0 &= \phi_1\gamma_1+\phi_2\gamma_2+\sigma^2 \\
\gamma_0 &= \phi_1\rho_1\gamma_0+\phi_2\rho_2\gamma_0+\sigma^2 \\
\gamma_0 &= \left[\frac{\phi_1^2}{(1-\phi_2)}+\frac{\phi_2\phi_1^2}{(1-\phi_2)}+\phi_2^2\right]\gamma_0+\sigma^2
\end{aligned}$$

or

$$\gamma_0=\frac{(1-\phi_2)\sigma^2}{(1+\phi_2)\left[(1-\phi_2)^2-\phi_1^2\right]}.$$

5.3 The pth-Order Autoregressive Process

An $AR(p)$ process satisfies

$$Y_t=c+\phi_1 Y_{t-1}+\phi_2 Y_{t-2}+...+\phi_p Y_{t-p}+\varepsilon_t$$

The mean is

$$\begin{aligned}
EY_t &= \mu=c+\phi_1\mu+\phi_2\mu+...+\phi_p\mu \\
\mu &= \frac{c}{(1-\phi_1-\phi_2-...-\phi_p)}
\end{aligned}$$

Rewriting the $AR(p)$ process by substracting μ gives

$$Y_t-\mu=\phi_1(Y_{t-1}-\mu)+\phi_2(Y_{t-2}-\mu)+...+\phi_p(Y_{t-p}-\mu)+\varepsilon_t$$

Then multiply this expression by $(Y_{t-j}-\mu)$ to find autocovariances:

$$\gamma_j=\left\{\begin{array}{l}\phi_1\gamma_{j-1}+...+\phi_p\gamma_{j-p}\ for\ j=1,2,.... \\ \phi_1\gamma_1+\phi_2\gamma_2+...+\phi_p\gamma_p+\sigma^2\ for\ j=0\end{array}\right.$$

To get the autocorrelations, just divide the γ_j -function by the variance γ_0 (Yule-Walker-equation):

$$\frac{\gamma_j}{\gamma_0}=\rho_j=\phi_1\rho_{j-1}+\phi_2\rho_{j-2}+...+\phi_p\rho_{j-p}\ for\ j=1,2,...$$

We can see that ρ_j and γ_j follow the same pth-order difference equation as the process itself.

6 Mixed Autoregressive Moving Average Processes

This $ARMA(p,q)$ processes contain terms of AR and MA processes and have the following shape:

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + ... + \theta_q \varepsilon_{t-q}$$

or in Lag operator form,

$$(1 - \phi_1 L - \phi_2 L^2 - ... - \phi_p L^p) Y_t = c + (1 + \theta_1 L + \theta_2 L^2 + ... + \theta_q L^q)\varepsilon_t$$

By dividing both sides by $(1 - \phi_1 L - \phi_2 L^2 - ... - \phi_p L^p)$ we obtain

$$Y_t = \mu + \Psi(L)\varepsilon_t$$

with

$$\Psi(L) = \frac{(1 + \theta_1 L + \theta_2 L^2 + ... + \theta_q L^q)}{(1 - \phi_1 L - \phi_2 L^2 - ... - \phi_p L^p)}$$

and

$$\sum_{j=0}^{\infty} \left|\varphi_j\right| < \infty$$

and a mean of

$$\mu = \frac{c}{(1 - \phi_1 - \phi_2 - ... - \phi_p)}$$

Now we see that the stationarity of an $ARMA$-process only depends on the AR-parameters $(\phi_1, \phi_2, ..., \phi_p)$ not on the MA-parameters.

In order to express the autocovariances, we substract the mean from the Y_t-equation and multiply it by the term $(Y_{t-j} - \mu)$. For $j > q$, the resulting equations are

$$\gamma_j = \phi_1 \gamma_{j-1} + \phi_2 \gamma_{j-2} + ... + \phi_p \gamma_{j-p} \; for \; j = q+1, q+2, ...$$

From this it follows that after q lags the autocovariance follows the pth-order different equation determined by $\phi_1, \phi_2, ..., \phi_p$.

7 Principles of Forecasting

In this section the basic concepts of unpredictability and forecastability are defined and the notions of loss functions and optimal predictors are addressed..

The variables to be forecast are an n-dimensional discrete-time stochastic process $\{x_t\}$, where $x_t^{'} = (x_{1,t} \ ... \ x_{n,t})$ for $t = 1, ..., T + H$. The joint density of x_t at time t, given the history of the process, is assumed to exists and is $D_{x_t}(x_t \,|X_{t-1}, \theta)$ where the vector of k parameters $\theta \in \Theta \subset R^k$. Thus, $D_{x_t}(\cdot)$ is a function of past information: $X_{t-1} = (... \ x_1 \ ... \ x_{t-1})$. We assume the existence of any necessary pre-sample information, X_0 and use the notation $X_{t-1} = (X_0, X_{t-1}^1)$ so $X_{t-1}^1 = (x_1 \ ... \ x_{t-1})$.

The period up to T is deemed to have occurred, and the forecast interval is $T+1$, ... , $T+H$. A statistical forecast $\widetilde{x}_{T+h}$for period $T+h$, conditional on information up to period $T+h$, is given by $\widetilde{x}_{T+h}= f_h(X_T)$, where the prior estimate of θ may be needed in order to compute the forecast. We assume θ to be the same over $t = 1$, ... ,T, but all elements of θ need not be relevant at all points in time. For simplification we take $n = 1$ and we refer to a scalar element of x_t as y_t.

The best known theorem is that, when the first two moments exist, forecasts calculated as the conditional expectation $\hat{y}_{T+H}= E\ [(y_{T+h}\,|Y_T]$ are unbiased, and no other predictor conditional on Y_T alone has a smaller mean-square forecast error (MSFE):

$$M_h[\hat{y}_{T+H}\ |Y_T] = E\ [(y_{T+h} - \ \hat{y}_{T+H})^2\,|Y_T]\,.$$

However, MSFE is not necessarily the most desirable criterion, and under other loss functions, the conditional expectation may not be optimal. For example, one may be concerned with the 'profitability' resulting from a forecast, or the costs of forecast errors may not be symmetric. If loss is asymmetric so that there are proportionately greater costs attached to, say, under-prediction than over-prediction, an optimal predictor would on average over-predict. While asymmetries are clearly important for individual agents's decisions, their importance at the macro-level is less obvious.

Now we want to define the notion of an unpredictable process, and note some of its implications. Unpredictability is the property of a random variable in relation to an information set, so is well defined. The same cannot be said of the next two concepts, namely informativeness, which is a relationship between information and a pre-existing knowledge state, and forecasting, which is a process of producing statements about future events.

7.1 Unpredictability

The definition of the unpredictability is equivalent to the statistical independence of an m-dimensional stochastic variable ν_t for an information set denoted T_{t-1}. So ν_t, is unpredictable with respect to T_{t-1} if the conditional and unconditional distribution coincide:

$$D_{\nu_t}(\nu_t \,|T_{t-1}) = D_{\nu_t}(\nu_t)$$

The knowledge of T_{t-1} does not improve prediction or reduce any aspect of the uncertainty about ν_t.A simple example is tossing of a unbiased coin: the next outcome of a head or a tail is unpredictable however many previous trials are observed, but one of the two outcomes is certain.
Unpredictability is invariant under non-singular contemporaneous transforms, and conversely it is not invariant under intertemporal transforms since if $u_t = \nu_t + Af(T_{t-1})$ then

$$D_{u_t}(u_t \,|T_{t-1}) = D_{u_t}(u_t)$$

when $A \neq 0$ since $E\,[u_t \,|T_{t-1}] = Af(T_{t-1})$. This concept helps to remove a possible 'paradox': the change in the log of real equity prices seam to be unpredictable in mean but the level is predictable. When the time series x_t is of interest, we assume the information set includes at least the history of x_t. Thus, a necessary condition for x_t to be unpredictable is that it is an innovation, and (weak) white noise (when its variances exists). Unpredictability is relative to the information set used; when $I_{t-1} \subset T_{t-1}$ it is possible that

$$D_{u_t}(u_t \,|I_{t-1}) = D_{u_t}(u_t)$$

yet

$$D_{u_t}(u_t \,|T_{t-1}) \neq D_{u_t}(u_t)$$

Any joint density $D_x(X_T^1 \,|X_0, \theta)$ can be sequentially factorized as

$$D_x(X_T^1 \,|X_0, \theta) = \prod_{t=1}^{T} D_{x_t}(x_t \,|X_{t-1}, \theta)$$

and if we define the 1-step error as:

$$e_t = x_t - E\,[x_t \,|X_{t-1}]$$

then

$$E\,[e_t \,|X_{t-1}] = 0 = E\,[e_t] = 0$$

which is unpredictable by construction. Consequently, predictability requires combinations of the unpredictable with T. In an odd sense, we can only predict what has already happened, or less paradoxically, the 'cause' must already be in train. For forecasting, a view has to be taken concerning both the relevant information set T_{t-1} and the form of the conditional density that relates the quantity to be forecast to the information set.

7.2 Informativeness

Informativeness is a relative concept, dependant on the pre-existing state of the recipient's information set I_{t-1}. If a variable is unpredictable, then the forecasts of it would seem to be uninformative, depending on the background information. For example knowing the form of $D_{\nu_t}(\nu_t)$, may be highly informative relative to not knowing it.

One might take one forecast to be more informative than another if its forecasting error has a more concentrated probability distribution given I. For example, over a horizon of H periods for two forecasts f_{T+h} and g_{T+h}, consider an ε-neighborhood $N_\varepsilon(\cdot)$, that

$$P\left(F_{T+H}^{T+1} \in N_\varepsilon(X_{T+H}^{T+1} \,|I_t)\right) > P\left(G_{T+H}^{T+1} \in N_\varepsilon(X_{T+H}^{T+1} \,|I_t)\right)$$

where $F_{T+H}^{T+1} = (f_{T+1} \ldots f_{T+H})$, then f is more informative.

7.3 Forecastability

A forecasting rule is any systematic operational procedure for making statements about the future. It is extremely difficult to define forecastable and unforecastable. One could perhaps define events as forecastable relative to a loss measure if the relevant procedure produced a lower expected loss than the unconditional mean.

For a weakly stationary, scalar time series, we define the limit of forecastability horizon H to be such that

$$V\ [y_{T+H}\ |T_T] > (1-\alpha)V\ [y_{T+H}]$$

α may be 0.05, 0.01 or else, so the forecast error variance is 100a percent of the unconditional variance. The choice of a will depend on the context and objectives of forecasting, and the value of H, where the limit is reached, will depend on the dynamics of the time series. If the time series is non-stationary, suitable differencing might reduce it to stationarity so the formula above could be applied to the differenced series. If the series is inherently positive, one could use instead:

$$\sqrt{V\ [y_{T+H}\ |T_T]} > \kappa\ \bar{y}$$

where κ may be 0.25, 0.5 or else, or alternatively, scaling could be relative to y_t. If y_t is in logs, there is no need to scale by the sample mean or initial forecast $(\bar{y})$.

7.4 Implications

These concepts have a number of important implications applicable to most statistical forecasting methods.

1. Since the conditional mean of an unpredictable process is its unconditional mean, predictability is necessary for forecastability. However it is not sufficient, since the relevant information set may be unknown in practice. T_{t-1} denotes the conditioning set of relevant events, this events are, or are not, predictable relative to T_{t-1}. But as an action by humans forecasting also requires knowledge of how T_{t-1} enters the conditional density.

2. If the occurrence of large ex-ante shocks (such as earth quakes, or oil crisis) induced their inclusion in later information sets the past will be more explicable then the future is forecastable.

3. From $D_x(X_T^1 | X_0, \theta) = \prod_{t=1}^{T} D_{x_t}(x_t | X_{t-1}, \theta)$ intertemporal transforms affect predictability, so no unique measure of predictability, and hence of forecast accuracy, exists.

4. Because new unpredictable components can enter in each period, forecast uncertainty could increase or decrease over increasing horizons from any given T. For integrated processes $V[y_{T+h} | T_T]$ is nondecreasing in h when the innovation distribution is homoscedastic. If the distribution is heteroscedastic forecasting uncertainty may increase or decrease in h.

5. If the true T_{t-1} is unknown, one cannot prove the genuinely relevant information is needed to forecast.

References

[1] Clements, Michael P. and David F. Hendry, (1998), "Forecasting Economic Time Series", Cambridge University Press.

[2] Dinardo, John and Jack Johnston, (1997), "Econometric Methods", McGraw-Hill.

[3] Hamilton, James D., (1994), "Time Series Analysis", Princeton University Press.