Basics of Time Series

Greg Strabel

January 31, 2023

1 Preliminaries

Definition: A time series $\{\epsilon_t\}$ is *White Noise*, $WN(0, \sigma_2)$, if it has finite first and second moments satisfying:

$$\mathbb{E}\epsilon_t = 0$$

$$\mathbb{E}\epsilon_t\epsilon_s = \begin{cases} \sigma^2 & t = s \\ 0 & t \neq s \end{cases} \tag{1}$$

Definition: Given a time series $\{Y_t\}$ the *lag operator* L is defined by the relation

$$LY_t = Y_{t-1}$$

With a slight, but common, abuse of notation, given vectors $\phi \in \mathbb{C}^p$ and $\theta \in \mathbb{C}^q$, the autoregressive and moving average lag polynomials $\phi(L)$ and $\theta(L)$ are defined as

$$\phi(L) = 1 - \sum_{i=1}^{p} \phi_i L^i \qquad \theta(L) = 1 + \sum_{i=1}^{q} \theta_i L^i \tag{2}$$

respectively.

Definition: A time series $\{Y_t\}$ is *covariance stationary* or *weakly stationary* if it has finite first and second moments satisfying:

$$\mathbb{E}Y_t = \mu \qquad \forall t \in \mathbb{Z}$$

$$\mathbb{E}(Y_t - \mu)(Y_{t-k} - \mu) = \gamma_k \qquad \forall t, k \in \mathbb{Z}$$

Given a time series $\{X_t\}$ satisfying $\mathbb{E}|X_t| < M$ and $\mathbb{E}|X_tX_s| < M$ and $|\phi| < 1$, we can form another time series

$$Y_t = \sum_{i=0}^{\infty} \phi^i X_{t-i}$$

which exists as an L^2 limit since

$$\mathbb{E}\left[\sum_{i=0}^{k}\phi^i X_{t-i} - \sum_{j=0}^{m}\phi^j X_{t-j}\right]^2 = \mathbb{E}\left[\sum_{i=k\wedge m+1}^{k\vee m}\phi^i X_{t-i}\right]^2$$

$$= \sum_{j=k\wedge m+1}^{k\vee m}\sum_{i=k\wedge m+1}^{k\vee m}\phi^j\phi^i \mathbb{E}X_{t-j}X_{t-i}$$

$$\leq M\phi^{k\wedge m+1}\sum_{i=0}^{(k\vee m)-(k\wedge m)-1}\sum_{j=0}^{(k\vee m)-(k\wedge m)-1}\phi^i\phi^j \qquad (3)$$

$$= \phi^{k\wedge m+1}M\left(\frac{1-\phi^{(k\vee m)-(k\wedge m)}}{1-\phi}\right)^2$$

$$\to 0 \quad \text{as} \quad k\wedge m\to\infty$$

Notice that $X_t = (1-\phi L)Y_t$ so that we have effectively defined an inverse to $1-\phi L$:

$$(1-\phi L)^{-1}:\{X_t\}\mapsto\left\{\sum_{i=0}^{\infty}\phi^i X_{t-i}\right\} \qquad (4)$$

This mapping is the workhorse of covariance stationary ARMA models.

Since L^2 is a pseudo-Hilbert space, we can also show that:

$$\mathbb{E}Y_t = \sum_{i=0}^{\infty}\phi^i\mathbb{E}X_{t-i}$$

and

$$\mathbb{E}Y_tY_s = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\phi^{i+j}\mathbb{E}X_{t-i}X_{s-j}$$

In particular, this implies that if $\{X_t\}$ is covariance stationary, then so is $\{Y_t\}$.

Definition: A covariance stationary process $\{Y_t\}$ is *causal* if there is a white noise process $\{\epsilon_t\}$ and a sequence of constants $\{\psi_i\}_{i\geq 0}$ such that $\sum_{i=0}^{\infty}|\psi_i|<\infty$ and

$$Y_t = \sum_{i=0}^{\infty}\psi_i\epsilon_{t-i} \qquad (5)$$

Definition: If $\{Y_t\}$ is covariance stationary with autocovariance $\{\gamma_k\}$, then its Autocovariance Generating Function is defined by

$$G(z) = \sum_{k=-\infty}^{k=\infty}\gamma_k z^k \qquad (6)$$

2

provided that the series converges for all z in some annulus $r^{-1} < |z| < r$ with $r > 1$.
If $\{\epsilon_t\} \sim WN\left(0, \sigma^2\right)$ and $Y_t = \sum_{i=-\infty}^{\infty} \psi_i \epsilon_{t-i}$ and there exists $r > 1$ such that $\sum_{i=-\infty}^{\infty} |\psi_i| z^i < \infty$ for $r^{-1} < |z| < r$ then

$$
\begin{aligned}
G\left(z\right) &= \sigma^2 \sum_{k=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \psi_j \psi_{j+|k|} z^k \\
&= \sigma^2 \left[\sum_{j=-\infty}^{\infty} \psi_j^2 + \sum_{k=1}^{\infty} \sum_{j=-\infty}^{\infty} \psi_j \psi_{j+k} \left(z^k + z^{-k} \right) \right] \\
&= \sigma^2 \left(\sum_{j=-\infty}^{\infty} \psi_j z^j \right) \left(\sum_{k=-\infty}^{\infty} \psi_k z^{-k} \right)
\end{aligned}
\tag{7}
$$

Definition: Let $(\Omega, \mathfrak{F}, \mathbb{P})$ be a probability space. The measurable transformation $T : \Omega \to \Omega$ is $measure-preserving$ if $\mathbb{P}T^{-1}A = \mathbb{P}A \ \forall A \in \mathfrak{F}$. The measure-preserving transformation $T : \Omega \to \Omega$ is $ergodic$ if $\forall A \in \mathfrak{F}, T^{-1}A = A \Rightarrow \mathbb{P}A \in \{0, 1\}$. A measurable function f is $T-invariant$ if $f\left(\omega\right) = f\left(T\omega\right) \ \forall \omega \in \Omega$.

The Ergodic Theorem: Let $T : \Omega \to \Omega$ be a measure-preserving transformation on $(\Omega, \mathfrak{F}, \mathbb{P})$ and f a measurable and integrable function. Then

$$
\widehat{f}\left(\omega\right) = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} f\left(T^i \omega\right)
\tag{8}
$$

converges with probability 1 and $\widehat{f}$ is $T-invariant$, integrable and $\mathbb{E}\widehat{f} = \mathbb{E}f$. If T is ergodic then $\widehat{f} = \mathbb{E}f$ with probability 1.

Definition: Given a covariance stationary time series $\{Y_t\}$, the partial autocorrelation function is:
$$
\alpha\left(1\right) = Corr\left(Y_2, Y_1\right)
\tag{9}
$$
and
$$
\alpha\left(k\right) = Corr\left(Y_{k+1} - P_{\overline{sp}\{1, Y_2, \ldots, Y_k\}} Y_{k+1}, Y_1 - P_{\overline{sp}\{1, Y_2, \ldots, Y_k\}} Y_1\right)
\tag{10}
$$
where $P_{\overline{sp}\{1, Y_2, \ldots, Y_k\}}$ is the operator of L^2 projection onto the subspace spanned by $\{1, Y_2, \ldots, Y_k\}$.

It is not hard to show that $\alpha\left(k\right) = \phi_{kk}$ where

$$
\begin{bmatrix}
\rho_0 & \rho_1 & \rho_2 & \cdots & \rho_{k-1} \\
\rho_1 & \rho_0 & \rho_1 & \cdots & \rho_{k-2} \\
\vdots & & \ddots & & \vdots \\
\rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & \rho_0
\end{bmatrix}
\begin{bmatrix}
\phi_{k1} \\
\phi_{k2} \\
\vdots \\
\phi_{kk}
\end{bmatrix}
=
\begin{bmatrix}
\rho_1 \\
\rho_2 \\
\vdots \\
\rho_k
\end{bmatrix}
\tag{11}
$$

and $\rho_j = \gamma_j / \gamma_0$ is the correlation between Y_{t+j} and Y_t.

3

2 Moving Average Processes and Invertibility

Suppose $\{\epsilon_t\}_{t=-\infty}^{\infty}$ is $WN\left(0, \sigma^2\right)$ and let

$$Y_t = \mu + \epsilon_t + \sum_{i=1}^{q} \theta_i \epsilon_{t-i} = \theta\left(L\right)\epsilon_t \tag{12}$$

Taking expectations on both sides of (12) shows that μ is the mean of the process. First note that

$$\gamma_k := \mathbb{E}\left(Y_t - \mu\right)\left(Y_{t-k} - \mu\right) = \begin{cases} \sigma^2 \sum_{i=0}^{q} \theta_i \theta_{i+|k|} & |k| \leq q \\ 0 & |k| > q \end{cases} \tag{13}$$

where we set $\theta_i = 0 \ \forall i > q$. Suppose that the polynomial $\theta\left(L\right)$ factors as $\prod_{i=1}^{q}\left(1 + \lambda_i L\right)$. If $|\lambda_i| < 1 \ \forall i$ then $\theta\left(L\right)$ is *invertible* and we can recover ϵ_t as

$$\epsilon_t = \left[\prod_{i=1}^{q}\left(1 + \lambda_i L\right)^{-i}\right]\left(Y_t - \mu\right)$$

However, even if $\theta\left(L\right)$ is not invertible, we can still find an invertible MA(q) representation for Y_t as long as $|\lambda_i| \neq 1 \ \forall i$. Define

$$\widetilde{\lambda}_i = \begin{cases} \lambda_i & |\lambda_i| < 1 \\ \frac{1}{\lambda_i} & |\lambda_i| > 1 \end{cases} \tag{14}$$

Then define

$$\widetilde{\epsilon}_t = \prod_{i=1}^{q}\left(1 + \widetilde{\lambda}_i L\right)^{-1}\left(Y_t - \mu\right) = \prod_{i=1}^{q}\left(1 + \widetilde{\lambda}_i L\right)^{-1}\left(1 + \lambda_i L\right)\epsilon_t \tag{15}$$

which exists as an L^2 limit. Moreover, the Autocovariance Generating Function for $\widetilde{\epsilon}_t$ is:

$$\begin{aligned} G_{\widetilde{\epsilon}_t}\left(z\right) &= \sigma_2 \prod_{i=1}^{q}\left(1 + \widetilde{\lambda}_i z\right)^{-1}\left(1 + \widetilde{\lambda}_i z^{-1}\right)^{-1}\left(1 + \lambda_i z\right)\left(1 + \lambda_i z^{-1}\right) \\ &= \sigma^2 \prod_{i:|\lambda_i|>1} \frac{\left(1 + \lambda_i z\right)\left(1 + \lambda_i z^{-1}\right)}{\left(1 + \frac{1}{\lambda_i}z\right)\left(1 + \frac{1}{\lambda_i}z^{-1}\right)} \\ &= \sigma^2 \prod_{i:|\lambda_i|>1} |\lambda_i|^2 \end{aligned} \tag{16}$$

It follows that $\{\widetilde{\epsilon}_t\} \sim WN\left(0, \sigma^2 \prod_{i:|\lambda_i|>1}|\lambda_i|^2\right)$ and $Y_t - \mu = \prod_{i=1}^{q}\left(1 + \widetilde{\lambda}_i L\right)\widetilde{\epsilon}_t$. We call $\{\widetilde{\epsilon}_t\}$ the *fundamental innovation* for $\{Y_t\}$. What we have shown is that even if we are given an MA(q) representation that is NOT invertible (i.e. some roots of the MA

4

polynomial lie within the unit circle), we can find an MA(q) representation of the process that is invertible.

For an invertible MA(q) process,

$$\epsilon_t = \theta\left(L\right)^{-1}\left(Y_t - \mu\right) = \sum_{k=0}^{\infty} \sum_{\alpha \in \mathbb{N}_0^q : |\alpha| = k} \left[\prod_{i=1}^{q} (-\lambda_i)^{\alpha_i} \right] (Y_{t-k} - \mu) \in \text{span}\left(\{Y_j : j \le t\} \cup \{1\}\right)$$

(17)

Letting P_t denote the L^2 projection operator onto $\text{span}\left(\{Y_j : j \le t\} \cup \{1\}\right)$, it follows that

$$P_t Y_{t+s} = \begin{cases} \mu & s > q \\ \mu + \sum_{i=s}^{q} \theta_i \epsilon_{t+s-i} & 1 \le s \le q \end{cases}$$

(18)

and therefore that

$$\mathbb{E}\left(Y_{t+s} - P_t Y_{t+s}\right)^2 = \begin{cases} \sigma^2 \left(1 + \sum_{i=1}^{q} \theta_i^2\right) & s > q \\ \sigma^2 \left(1 + \sum_{i=1}^{s-1} \theta_i^2\right) & 1 \le s \le q \end{cases}$$

(19)

Of course, we generally do not have infinitely many lagged values of Y_t to calculate ϵ_t.

Although the proof is long, it is possible to show using (17) that the partial autocorrelations of Y_t eventually decay asymptotically to zero as the sum of a geometric series.

3 Autoregressive Processes

3.1 Properties of Autoregressive Processes

An AR(p) process is a time series $\{Y_t\}$ satisfying:

$$Y_t = c + \sum_{i=1}^{p} \phi_i Y_{t-i} + \epsilon_t$$

(20)

If $\{Y_t\}$ is stationary then

$$\mu = c + \sum_{i=1}^{p} \phi_i \mu$$

so that $c = \mu\left(1 - \sum_{i=1}^{p} \phi_i\right)$ and $\mu = c\left(1 - \sum_{i=1}^{p} \phi_i\right)^{-1}$ and hence we can rewrite (20) as

$$\left(Y_t - \mu\right) = \sum_{i=1}^{p} \phi_i\left(Y_{t-i} - \mu\right) + \epsilon_t$$

(21)

Therefore, the recentered process obeys the same law of motion as the original process. It follows that

$$\gamma_k = \begin{cases} \sum_{i=1}^p \phi_i \gamma_i + \sigma^2 & \text{for } k = 0 \\ \sum_{i=1}^p \phi_i \gamma_{k-i} & \text{for } k \geq 1 \end{cases} \tag{22}$$

These are called the Yule-Walker equations and, if γ_k is either known or can be estimated, can be used to construct estimators of $\phi_1, ..., \phi_p$. We can write the process in matrix notation as:

$$\mathbf{Y}_t = \begin{bmatrix} Y_t \\ Y_{t-1} \\ \vdots \\ Y_{t-p+1} \end{bmatrix} \qquad \mathbf{\Xi}_t = \begin{bmatrix} \epsilon_t \\ 0 \\ \vdots \\ 0 \end{bmatrix} \qquad \overrightarrow{\mu} = \begin{bmatrix} \mu \\ \mu \\ \vdots \\ \mu \end{bmatrix} \tag{23}$$

$$\Phi = \begin{bmatrix} \phi_1 & \phi_2 & \cdots & \phi_p \\ 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \cdots & 0 \\ 0 & \cdots & 1 & 0 \end{bmatrix} \tag{24}$$

so that

$$\begin{aligned} \mathbf{Y}_t - \overrightarrow{\mu} &= \Phi \left(\mathbf{Y}_{t-1} - \overrightarrow{\mu} \right) + \mathbf{\Xi}_t \\ &= \Phi^{(2)} \left(\mathbf{Y}_{t-2} - \overrightarrow{\mu} \right) + \Phi \mathbf{\Xi}_{t-1} + \mathbf{\Xi}_t \\ &= \Phi^{(k+1)} \left(\mathbf{Y}_{t-k-1} - \overrightarrow{\mu} \right) + \sum_{i=0}^k \Phi^{(i)} \mathbf{\Xi}_{t-i} \end{aligned} \tag{25}$$

Let Γ be the $p \times p$ matrix where $\Gamma_{ij} = \gamma_{|i-j|}$ and let Υ be the $p \times p$ matrix where the only non-zero entry is $\Upsilon_{11} = \sigma^2$. Then

$$\Gamma = \mathbf{\Phi} \Gamma \mathbf{\Phi}^{-1} + \Upsilon \tag{26}$$

so that

$$vec\left(\Gamma\right) = \left(\left(\Phi^{-1}\right)^T \bigotimes \Phi \right) vec\left(\Gamma\right) + vec\left(\Upsilon\right) \tag{27}$$

from which it follows that

$$vec\left(\Gamma\right) = \left[I_{p^2} - \left(\Phi^{-1}\right)^T \bigotimes \Phi \right]^{-1} vec\left(\Upsilon\right) \tag{28}$$

Letting P_t denote the L^2 projection operator onto $\text{span}\left(\{Y_j : j \leq t\} \cup \{1\}\right)$, it follows that

$$P_t Y_{t+s} = \mu + \sum_{i=1}^p \Phi_{1i}^{(s)} \left(Y_{t+1-i} - \mu \right) \tag{29}$$

6

so that

$$Y_{t+s} - P_t Y_{t+s} = \sum_{i=0}^{s-1} \Phi_{11}^{(i)} \epsilon_{t+s-i} \tag{30}$$

and therefore

$$\mathbb{E}\left(Y_{t+s} - P_t Y_{t+s}\right)^2 = \sigma^2 \sum_{i=1}^{s-1} \left(\Phi_{11}^{(i)}\right)^2 \tag{31}$$

In particular, this implies that the partial autocorrelation of an AR(p) process satisfies $\alpha\left(k\right) = 0 \ \forall k > p$.

3.2 Estimating an Autoregressive Model with OLS

Assume we have data from an AR(p) process of the form (20). If we have a total of $T + p$ observations indexed from $1 - p$ to T and define

$$X_t = \begin{bmatrix} 1 \\ Y_{t-1} \\ Y_{t-2} \\ \vdots \\ Y_{t-p} \end{bmatrix} \tag{32}$$

then the OLS estimator $\widehat{\beta}$ of $\beta = \begin{bmatrix} \mu & \phi_1 & \dots & \phi_p \end{bmatrix}'$ is

$$\widehat{\beta} = \left[\sum_{t=1}^{T} X_t X_t'\right]^{-1} \sum_{t=1}^{T} X_t Y_t \tag{33}$$

If we define the $(p+1) \times (p+1)$ matrix Ψ by

$$\Psi_{ij} = \begin{cases} 1 & i = j = 1 \\ \mu & i = 1, j \geq 2 \\ \mu & i \geq 2, j = 1 \\ \gamma_{|i-j|} + \mu^2 & i \geq 2, j \geq 2 \end{cases} \tag{34}$$

then by the law of large numbers,

$$\left[\frac{1}{T}\sum_{t=1}^{T} X_t X_t'\right]^{-1} \xrightarrow{P} \Psi^{-1} \tag{35}$$

Likewise, by the Central Limit Theorem for Martingale Difference Sequences

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} X_t \epsilon_t \rightsquigarrow \mathbf{N}\left(0, \sigma^2 \Psi\right) \tag{36}$$

7

It follows that

$$\sqrt{T}\left(\widehat{\beta}-\beta\right) \rightsquigarrow \mathbf{N}\left(0, \sigma^2 \Phi^{-1}\right) \tag{37}$$

In particular, for an AR(1) process,

$$\Psi = \begin{bmatrix} 1 & \mu \\ \mu & \frac{\sigma^2}{1-\phi_1^2}+\mu^2 \end{bmatrix} \qquad \Psi^{-1} = \frac{1}{\sigma^2}\begin{bmatrix} \sigma^2 + \mu^2\left(1-\phi_1^2\right) & -\mu\left(1-\phi_1^2\right) \\ -\mu\left(1-\phi_1^2\right) & 1-\phi_1^2 \end{bmatrix} \tag{38}$$

4 Autoregressive Moving Average (ARMA)

An $ARMA\left(p,q\right)$ process $\{Y_t\}$ is a stochastic process that follows the evolution:

$$Y_t = c + \sum_{i=1}^{p}\phi_i Y_{t-i} + \epsilon_t + \sum_{i=1}^{q}\theta_i \epsilon_{t-i} \tag{39}$$

Covariance stationarity of $\{Y_t\}$ implies that $\mu = c + \mu\sum_{i=1}^{p}\phi_i$ so that

$$Y_t - \mu = \sum_{i=1}^{p}\phi_i\left(Y_{t-i}-\mu\right) + \epsilon_t + \sum_{i=1}^{q}\theta_i \epsilon_{t-i} \tag{40}$$

or, using lag polynomials,

$$\phi\left(L\right)\left(Y_t - \mu\right) = \theta\left(L\right)\epsilon_t \tag{41}$$

A causal, stationary solution for (39) has the representation $Y_t - \mu = \sum_{i=0}^{\infty}\psi_i \epsilon_{t-i}$, from which we can recover $\{\psi_i\}$ from matching coefficients in the relation $\phi\left(L\right)\psi\left(L\right) = \theta\left(L\right)$. It follows that

$$\gamma_k = \begin{cases} \sum_{i=1}^{p}\phi_i\gamma_i + \sigma^2\left(1+\sum_{i=1}^{q}\theta_i\psi_i\right) & k=0 \\ \sum_{i=1}^{p}\phi_i\gamma_{|k-i|} + \sigma^2\sum_{i=k}^{q}\theta_i\psi_{i-k} & 1 \le k \le q \\ \sum_{i=1}^{p}\phi_i\gamma_{|k-i|} & k > q \end{cases} \tag{42}$$

Notice that at sufficiently large lags, the autocovariance decays as the same sum of a geometric series that we saw for autoregressive models.

Without loss of generality, we may assume that $\theta\left(L\right)$ is invertible so that $\{\epsilon_t\}$ are the fundamental innovations for the process. Then

$$\epsilon_t = \frac{\phi\left(L\right)}{\theta\left(L\right)}\left(Y_t - \mu\right) \in \operatorname{span}\left(\{Y_j : j \le t\} \cup \{1\}\right) \tag{43}$$

It is also common to write this in matrix notation as:

$$\mathbf{Y}_t = \begin{bmatrix} Y_t \\ Y_{t-1} \\ \vdots \\ Y_{t-p+1} \end{bmatrix} \qquad \mathbf{\Xi}_t = \begin{bmatrix} \epsilon_t \\ \epsilon_{t-1} \\ \vdots \\ \epsilon_{t-q} \end{bmatrix} \qquad \overrightarrow{\mu} = \begin{bmatrix} \mu \\ \mu \\ \vdots \\ \mu \end{bmatrix} \tag{44}$$

8

$$\Phi = \begin{bmatrix} \phi_1 & \phi_2 & \dots & \phi_p \\ 1 & 0 & \dots & 0 \\ \vdots & \ddots & \dots & 0 \\ 0 & \dots & 1 & 0 \end{bmatrix} \qquad \Theta = \begin{bmatrix} 1 & \theta_1 & \dots & \theta_q \\ 0 & 0 & \dots & 0 \\ \vdots & \ddots & \dots & 0 \\ 0 & 0 & \dots & 0 \end{bmatrix} \qquad (45)$$

so that

$$\mathbf{Y}_t - \overrightarrow{\mu} = \Phi\left(\mathbf{Y}_{t-1} - \overrightarrow{\mu}\right) + \Theta\Xi_t \qquad (46)$$

Iterating on (46), we have:

$$
\begin{aligned}
\mathbf{Y}_t - \overrightarrow{\mu} &= \Phi\left(\mathbf{Y}_{t-1} - \overrightarrow{\mu}\right) + \Theta\epsilon_t \\
&= \Phi\left(\Phi\left(\mathbf{Y}_{t-2} - \overrightarrow{\mu}\right) + \Theta\Xi_{t-1}\right) + \Xi_t \\
&= \sum_{i=0}^{k} \Phi^{(i)}\Xi_{t-i} + \Phi^{(k+1)}\left(\mathbf{Y}_{t-k-1} - \overrightarrow{\mu}\right)
\end{aligned}
\qquad (47)
$$

It follows that covariance stationarity requires that the eigenvalues of Φ lie strictly inside the unit circle.

Similar to the case of an MA(q) process, it is possible to show that the autocorrelation of an ARMA(p,q) eventually decays towards 0 as the sum of geometric series.

5 Durbin-Levinson and Innovations Algorithms

Suppose we have a zero mean process $\{X_t\}_{t\geq 1}$ in L^2 and we wish to calculate the L^2 projection of X_{n+1} onto the linear span of $\{X_t\}_{t=1}^n$. The projection, $\widehat{X}_{n+1}$, can be written

$$\widehat{X}_{n+1} = \sum_{i=1}^{n} \phi_{ni} X_{n+1-i} \qquad (48)$$

Let

$$\nu_n = \mathbb{E}\left(X_{n+1} - \widehat{X}_{n+1}\right)^2 \qquad (49)$$

be the mean square error of the projection.

The Durbin-Levinson Algorithm is a recursive method for calculating the coefficients of (48) and the mean squared errors in (49) for covariance stationary $\{X_t\}$.

Durbin-Levinson Algorithm (Proposition 5.2.1 in Brockwell and Davis): Assume $\{X_t\}$ is a mean-zero covariance stationary process with autocovariances $\{\gamma_k\}$ where $\gamma_0 > 0$ and $\gamma_k \to 0$ as $k \to \infty$. Then the coefficients ϕ_{ni} and mean squared errors ν_n in (48) and (49) satisfy $\phi_{11} = \gamma_1/\gamma_0$, $\nu_0 = \gamma_0$,

$$\phi_{nn} = \left[\gamma_n - \sum_{i=1}^{n-1} \phi_{n-1,i}\gamma_{n-i}\right]\nu_{n-1}^{-1} \qquad (50)$$

9

$$\begin{bmatrix} \phi_{n1} \\ \vdots \\ \phi_{n,n-1} \end{bmatrix} = \begin{bmatrix} \phi_{n-1,1} \\ \vdots \\ \phi_{n-1,n-1} \end{bmatrix} - \phi_{nn} \begin{bmatrix} \phi_{n-1,n-1} \\ \vdots \\ \phi_{n-1,1} \end{bmatrix} \tag{51}$$

and

$$\nu_n = \nu_{n-1} \left[1 - \phi_{nn}^2 \right] \tag{52}$$

Another, more general recursive algorithm for calculating the projects in (48) is the Innovations Algorithm:

The Innovations Algorithm (Proposition 5.2.2 in Brockwell and Davis): If $\{X_t\}$ is a mean-zero process with $\mathbb{E}X_i X_j = \kappa_{ij}$ and the matrix $[\kappa_{ij}]_{i,j=1}^{n}$ is non-singular for each $n \geq 1$ then the projections and mean squared errors in (48) and (49) can be calculated recursively as

$$\widehat{X}_{n+1} = \begin{cases} 0 & \text{if } n = 0 \\ \sum_{i=1}^{n} \theta_{ni} \left(X_{n+1-i} - \widehat{X}_{n+1-i} \right) & \text{if } n \geq 1 \end{cases} \tag{53}$$

$$\nu_0 = \kappa_{11} \tag{54}$$

$$\theta_{n,n-k} = \nu_k^{-1} \left(\kappa_{n+1,k+1} - \sum_{i=0}^{k-1} \theta_{k,k-i} \theta_{n,n-i} \nu_i \right) \quad k = 0, 1, ..., n-1 \tag{55}$$

$$\nu_n = \kappa_{n+1,n+1} - \sum_{i=0}^{n-1} \theta_{n,n-i}^2 \nu_i \tag{56}$$

6   Random Walk

Suppose $\{\epsilon_t\} \sim IID\left(0, \sigma^2\right)$ and $Y_t = Y_{t-1} + \epsilon_t$. For $r \in [0, 1]$, define

$$X_T(r) = \frac{1}{T} Y_{\lfloor rT \rfloor} \tag{57}$$

By the Functional Central Limit Theorem,

$$\sqrt{T} X_T(\cdot) \rightsquigarrow \sigma W(\cdot) \tag{58}$$

where W is a standard Brownian Motion. It follows from the Continuous Mapping Theorem that

$$T^{-3/2} \sum_{t=1}^{T} Y_{t-1} = \int_0^1 \sqrt{T} X_T(r) \, dr \rightsquigarrow \sigma \int_0^1 W(r) \, dr \tag{59}$$

10

Note that

$$
\begin{aligned}
T^{-3/2} \sum_{t=1}^{T} Y_{t-1} &= T^{-3/2} \sum_{t=1}^{T} (T-t) \, \epsilon_t \\
&= T^{-1/2} \sum_{t=1}^{T} \epsilon_t - T^{-3/2} \sum_{t=1}^{T} t\epsilon_t
\end{aligned}
\tag{60}
$$

Therefore

$$
\begin{aligned}
T^{-3/2} \sum_{t=1}^{T} t\epsilon_t &= T^{-1/2} \sum_{t=1}^{T} \epsilon_t - T^{-3/2} \sum_{t=1}^{T} Y_{t-1} \\
&\rightsquigarrow \sigma W(1) - \sigma \int_0^1 W(r) \, dr
\end{aligned}
\tag{61}
$$

By the Continuous Mapping Theorem we also have

$$
T^{-2} \sum_{t=1}^{T} Y_{t-1}^2 = \int_0^1 \left[\sqrt{T} X_T(r) \right]^2 dr \rightsquigarrow \sigma^2 \int_0^1 \left[W(r) \right]^2 dr
\tag{62}
$$

$$
T^{-5/2} \sum_{t=1}^{T} t Y_{t-1} = T^{-3/2} \sum_{t=1}^{T} (t/T) Y_{t-1} \rightsquigarrow \sigma \int_0^1 r W(r) \, dr
\tag{63}
$$

and

$$
T^{-3} \sum_{t=1}^{T} t Y_{t-1}^2 = T^{-2} \sum_{t=1}^{T} (t/T) Y_{t-1}^2 \rightsquigarrow \sigma^2 \int_0^1 r \left[W(r) \right]^2 dr
\tag{64}
$$

Since

$$
Y_T^2 = \sum_{t=1}^{T} \sum_{j=1}^{T} \epsilon_t \epsilon_j = 2 \sum_{t=1}^{T} Y_{t-1} \epsilon_t + \sum_{t=1}^{T} \epsilon_t^2
\tag{65}
$$

it follows that

$$
\frac{1}{T} \sum_{t=1}^{T} Y_{t-1} \epsilon_t = \frac{1}{2T} Y_T^2 - \frac{1}{2T} \sum_{t=1}^{T} \epsilon_t^2 \rightsquigarrow \frac{1}{2} \sigma^2 \left[W(1) \right]^2 - \frac{1}{2} \sigma^2
\tag{66}
$$

Then the OLS estimator of β in the regression $Y_t = \beta Y_{t-1} + \epsilon_t$ is

$$
\widehat{\beta} = \left[\sum_{t=1}^{T} Y_{t-1}^2 \right]^{-1} \sum_{t=1}^{T} Y_{t-1} Y_t = 1 + \left[\sum_{t=1}^{T} Y_{t-1}^2 \right]^{-1} \sum_{t=1}^{T} Y_{t-1} \epsilon_t
\tag{67}
$$

11

Hence

$$T\left(\widehat{\beta} - 1\right) = \left[T^{-2} \sum_{t=1}^{T} Y_{t-1}^2\right]^{-1} \frac{1}{T} \sum_{t=1}^{T} Y_{t-1} \epsilon_t$$

$$\rightsquigarrow \frac{\frac{1}{2}\sigma^2 \left[W\left(1\right)\right]^2 - \frac{1}{2}\sigma^2}{\sigma^2 \int_0^1 \left[W\left(r\right)\right]^2 dr} \tag{68}$$

$$\overset{L}{=} \frac{\frac{1}{2}\{\left[W\left(1\right)\right]^2 - 1\}}{\int_0^1 \left[W\left(r\right)\right]^2 dr}$$

7   Deterministic Time Trends

Suppose that

$$Y_t = \beta_1 + \beta_2 t + \epsilon_t \tag{69}$$

where $\{\epsilon_t\} \sim WN\left(0, \sigma^2\right)$. The OLS estimator for β is

$$\widehat{\beta} = \beta + \begin{bmatrix} T & \sum_{t=1}^{T} t \\ \sum_{t=1}^{T} t & \sum_{t=1}^{T} t^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{t=1}^{T} \epsilon_t \\ \sum_{t=1}^{T} t\epsilon_t \end{bmatrix} \tag{70}$$

It is simple to prove by induction that

$$\sum_{t=1}^{T} t = T\left(T + 1\right)/2 \qquad \sum_{t=1}^{T} t^2 = T\left(T + 1\right)\left(2T + 1\right)/6 \tag{71}$$

Hence

$$\begin{bmatrix} T & \sum_{t=1}^{T} t \\ \sum_{t=1}^{T} t & \sum_{t=1}^{T} t^2 \end{bmatrix}^{-1} = \frac{1}{T^2\left(T + 1\right)\left(2T + 1\right)/6 - T^2\left(T + 1\right)^2/4} \tag{72}$$

$$Var\left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left[\lambda_1 + \lambda_2\left(t/T\right)\right] \epsilon_t\right) = \frac{1}{T} \sum_{t=1}^{T} \sigma^2 \left[\lambda_1^2 + 2\lambda_1\lambda_2\left(t/T\right) + \lambda_2^2\left(t/T\right)^2\right]$$

$$\rightarrow \sigma^2 \left[\lambda_1^2 + \lambda_1\lambda_2 + \frac{1}{3}\lambda_2^2\right] \tag{73}$$

$$= \sigma^2 \lambda' \mathbf{Q} \lambda$$

where

$$\lambda = \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} \qquad \mathbf{Q} = \begin{bmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{3} \end{bmatrix} \tag{74}$$

It follows from the Central Limit Theorem for Martingale Difference Sequences and the Cramer Wold device that

$$\begin{bmatrix} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \epsilon_t \\ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{t}{T} \epsilon_t \end{bmatrix} \rightsquigarrow N\left(0, \sigma^2 \mathbf{Q}\right) \tag{75}$$

12

Then letting
$$R = \begin{bmatrix} \sqrt{T} & 0 \\ 0 & \sqrt{T^3} \end{bmatrix} \tag{76}$$

we have
$$\begin{bmatrix} \sqrt{T}\left(\widehat{\beta}_1 - \beta_1\right) \\ T^{3/2}\left(\widehat{\beta}_2 - \beta_2\right) \end{bmatrix} = R\left(\widehat{\beta} - \beta\right) = R \begin{bmatrix} T & \sum_{t=1}^{T} t \\ \sum_{t=1}^{T} t & \sum_{t=1}^{T} t^2 \end{bmatrix}^{-1} RR^{-1} \begin{bmatrix} \sum_{t=1}^{T} \epsilon_t \\ \sum_{t=1}^{T} t\epsilon_t \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{T}T & \frac{1}{T^2}\sum_{t=1}^{T} t \\ \frac{1}{T^2}\sum_{t=1}^{T} t & \frac{1}{T^3}\sum_{t=1}^{T} t^2 \end{bmatrix}^{-1} R^{-1} \begin{bmatrix} \sum_{t=1}^{T} \epsilon_t \\ \sum_{t=1}^{T} t\epsilon_t \end{bmatrix} \tag{77}$$

Since
$$\begin{bmatrix} \frac{1}{T}T & \frac{1}{T^2}\sum_{t=1}^{T} t \\ \frac{1}{T^2}\sum_{t=1}^{T} t & \frac{1}{T^3}\sum_{t=1}^{T} t^2 \end{bmatrix}^{-1} \xrightarrow{P} \mathbf{Q}^{-1} \tag{78}$$

it follows that
$$\begin{bmatrix} \sqrt{T}\left(\widehat{\beta}_1 - \beta_1\right) \\ T^{3/2}\left(\widehat{\beta}_2 - \beta_2\right) \end{bmatrix} \rightsquigarrow N\left(0, \sigma^2 \mathbf{Q}^{-1}\right) \tag{79}$$

8   Classical Time Series Decomposition

It is often the case that a time series will exhibit both a trend and cyclical component, typically called seasonality. The "classical decomposition" is the representation of a time series $\{Y_t\}$ as
$$Y_t = m_t + s_t + X_t \tag{80}$$

where m_t is a trend component, s_t is a seasonal or cyclical component with cycle d and X_t is covariance stationary. We review a few standard approaches to modeling (80).

8.1   Method 1

For the first method, we apply a low pass filter to the series to approximate the time trend while leaving the cyclical component intact. If d is odd, let $q = \frac{1}{2}(d-1)$, otherwise let $q = \frac{1}{2}d$. Then let

$$\widehat{m_t} = \begin{cases} \frac{1}{d}\left(\frac{1}{2}Y_{t-q} + \sum_{i=t-q+1}^{t+q-1} Y_i + \frac{1}{2}Y_{t+q}\right) & d \text{ even}, \ q < t \leq T - q \\ \frac{1}{d}\sum_{i=t-q}^{t+q} Y_i & d \text{ odd}, \ q < t \leq T - q \end{cases} \tag{81}$$

Setting $l_k = \lfloor \frac{1}{d}(q-k) \rfloor + 1$ and $u_k = \lfloor \frac{1}{d}(T-q-k) \rfloor$ for each $k \in \{1, ..., d\}$, define

$$w_k = \frac{1}{u_k - l_k + 1} \sum_{i=l_k}^{u_k} (Y_{k+id} - \widehat{m}_{k+id}) \tag{82}$$

13

We then estimate the seasonal component as

$$\widehat{s}_k = w_k - \frac{1}{d}\sum_{i=1}^{d} w_i \qquad k \in \{1, ..., d\} \tag{83}$$

and recursively as $\widehat{s}_k = \widehat{s}_{k-d}$ for $k > d$. The deseasonalized data is then defined as

$$d_t = Y_t - \widehat{s}_{t-d\lfloor t/d \rfloor} \tag{84}$$

We then re-estimate the trend component using this deseasonalized data using either a low pass filter as before or fitting a low order polynomial to produce the final estimate $\{\widehat{m}_t\}$. Finally, we produce an estimate of $\{X_t\}$ as

$$\widehat{X}_t = Y_t - \widehat{s}_t - \widehat{m}_t \tag{85}$$

8.2   Method 2

The second method is to apply a lag polynomial that effectively eliminates the seasonal and trend components. In particular, we look for a lag polynomial of the form $(1 - L)^k \left(1 - L^d\right)$ that results in a covariance stationary series when applied to $\{Y_t\}$. The term $\left(1 - L^d\right)$ removes the seasonal component. Ideally we find the smallest value of $k \in \mathbb{N}$ that results in a covariance stationary remainder.

9   Autoregressive Integrated Moving Averages ARIMA(p,d,q)

Given $\{\epsilon_t\} \sim WN\left(0, \sigma^2\right)$, an $ARIMA\left(p, d, q\right)$ is a time series satisfying

$$\phi\left(L\right)\left(1 - L\right)^d Y_t = \theta\left(L\right)\epsilon_t \tag{86}$$

where $\phi\left(L\right)$ and $\theta\left(L\right)$ are polynomials of degrees p and q, respectively, and all roots of these polynomials lie outside of the unit circle.