Linear Time Series

Recall the definition of a *linear time series* \{X_t\}: one that follows the relation,

\[
X_t = \mu + \sum_{i=0}^{\infty} \psi_i A_{t-i},
\]

where \{A_t\} is white noise.

(Recall “white noise” in Tsay means iid – but he’s not clear on the mean. He says (usually) “we assume the mean is 0”. Most people define white noise to be a process with 0 mean, finite and constant variance, and 0 autocovariance at all lags. Under that more common meaning of white noise, we must specify that \{A_t\} is iid with 0 mean, finite and constant variance. Just to bring these differing definitions full circle, if \{A_t\} is white noise under the most common sense, then expression (1) requires one more condition: \{A_t\} must be a martingale relative to \{X_t\}; that is, \(E(A_t|X_s, -\infty < s \leq t) = 0\) for \(t = 0, 1, \ldots\)

A linear time series is stationary if \(\sum_{i=0}^{\infty} \psi_i^2 < \infty\).
Linear Time Series

We sometimes use “time series” and “model of a time series” (or just “model”) synonymously.

ARMA models that are causal and invertible (“stationary”) are linear.

Integrated and seasonal linear ARMA models are linear.
Linear Spaces

A linear space is a collection of objects that is closed with respect to a binary operation (called vector addition) and multiplication by a real number (called scalar multiplication) with the property that scalar multiplication distributes over both real addition and vector addition.

An important class of linear spaces is a Hilbert space (has an inner product and is Cauchy with respect to the norm induced by the inner product).

(The norm induced by an inner product is an $L_2$-type norm.)

The most important linear space, which is also a Hilbert space, is $\mathbb{R}^d$.

This is the vector space of $d$-dimensional vectors, and in the following we’ll only consider it although what we say applies more generally to Hilbert spaces.
Projections in Linear Spaces

One of the most important properties of linear spaces is expressed by the “projection theorem”, which is an existence theorem.

For $\mathbb{R}^d$, this theorem is

Let $\mathcal{W} \subseteq \mathbb{R}^d$ be a linear space, and let $y \in \mathbb{R}^d$. Then there is a unique $\hat{y} \in \mathcal{W}$ and a unique $z \in \mathbb{R}^d \ni z \perp \mathcal{W}$ such that

$$y = \hat{y} + z$$

and

$$\|y - \hat{y}\|_2 < \|y - w\|_2 \quad \forall w \neq \hat{y} \in \mathcal{W}.$$ 

A corollary is

Let $A$ be an idempotent $d \times d$ matrix and let $\mathcal{W}$ be the vector space spanned by the columns of $A$. Then for any $y \in \mathbb{R}^d$, $\hat{y}$ in the theorem above is given by

$$\hat{y} = Ay.$$ 

Think of the application of this theorem in linear regression.
Properties of Projections


The projection of a vector onto a subspace is the “closest” vector in that subspace to the given vector.

The projection theorem is sort of a Pythagorean theorem.

If $A$ is an idempotent matrix in $\mathbb{R}^{d \times d}$ and $y$ is a vector in $\mathbb{R}^d$, then the projection $\hat{y} = Ay$ has the property that

$$\|y\| = \|\hat{y}\| + \|y - \hat{y}\|.$$

(Here and in the following “$\| \cdot \|”$ means “$\| \cdot \|_2$.”)

If $\hat{y} = Ay$, $\hat{y}$ is the projection of $y$ onto the columns of $A$.

Think of this fact in the context of linear regression.
Properties of Projections

For a given idempotent matrix, think of two spaces, the space spanned by the columns of the matrix, and the orthogonal complement to this space.

Notation:
\( \mathcal{V}(A) \) is the space spanned by the columns of \( A \).
\( \mathcal{N}(A) \) is the orthogonal complement to the space spanned by the columns of \( A \).

If the columns of \( A \) are in \( \mathbb{R}^d \), then

\[
\mathcal{V}(A) \oplus \mathcal{N}(A) = \mathbb{R}^d.
\]

and if \( A \) is a \( d \times d \) matrix, then

\[
\mathcal{N}(A) = \mathcal{V}(I - A).
\]
Projections of Random Variables


We extend the concept of projection to random variables.

The simplest instance is when we are given a single probability space.

The projection of the random variable $Y$ onto a subspace of that probability space is the random variable $X_p$ in the subspace such that

$$ E(\|Y - X_p\|) \leq E(\|Y - X\|), $$

where $X$ is any random variable in the subspace.

So $X_p$ is the “closest” random variable to $Y$ in that subspace.

We could also say $X_p$ is “best mean square predictor” of $Y$ in that subspace.

(We also sometimes drop “mean square”, but that’s what we mean, unless we specify some other criterion.)
Projections of Random Variables

Here’s a big deal:

\textit{Let }$Y$\textit{ be a }$d$\textit{-variate random variable such that }$E(\|Y\|) < \infty$\textit{ and let }$G$\textit{ be the set of all [measurable] functions from }$\mathbb{R}^k$\textit{ into }$\mathbb{R}^d$. \textit{Let }$X$\textit{ be a }$k$\textit{-variate random variable such that }$E(\|E(Y|X)\|) < \infty$. \textit{Let }$g_0(X) = E(Y|X)$. \textit{Then for any }$g$\textit{ in }$G$,

\[ E(\|Y - g_0(X)\|) \leq E(\|Y - g(X)\|). \]
Projections of random variables have similar properties to those of projections of vectors.

The angle between vectors is sorta like the correlation between random variables (that is, the cosine of the angle), and perpendicular vectors are sorta like random variables with 0 correlations.

Here’s a big deal:

Let $\mathcal{X}$ be a linear space of random variables with finite second moments. and let $X_p$ be a projection of the random variable $Y$ onto $\mathcal{X}$. Then,

$$E(X_p) = E(Y),$$

$$\text{Cov}(Y - X_p, X) = 0 \quad \forall X \in \mathcal{X},$$

and

$$\text{Cov}(Y, X) = \text{Cov}(X_p, X) \quad \forall X \in \mathcal{X}.$$
Projections and Predictions in Time Series

In a time series, we think of probability spaces as evolving; that is, getting “bigger”.

The idea is that “information” is increasing as successive random variables are encountered.

The probability spaces are represented by the $\sigma$-fields,
$$ \cdots \subseteq \mathcal{F}_{t-1} \subseteq \mathcal{F}_t \subseteq \mathcal{F}_{t+1} \subseteq \cdots. $$

Given a time series $\{X_t\}$, what is the best predictor of $X_{t+h}$ at time $t$?

It is
$$ \hat{X}_{t+h} = \mathbb{E}(X_{t+h} \mid \mathcal{F}_t). $$

This is a projection of course.

We can denote the probability space of the time series at time $t$ as $\mathcal{P}(\mathcal{F}_t)$. 
Predictors in Time Series

best mean square predictor

\( \hat{X}_{t+h} \) minimizes \( \mathbb{E}\left( (X_{t+h} - \hat{X}_{t+h})^2 \bigg| X_s, -\infty < s \leq t \right) \).

The solution is \( \hat{X}_{t+h} = \mathbb{E}(X_{t+h} \mid X_s, -\infty < s \leq t) \)

best linear predictor

\( \hat{X}_{t+h} = \sum_{i \leq t} b_i X_i \) such that

\[
\mathbb{E}\left( (X_{t+h} - \hat{X}_{t+h})^2 \bigg| X_s, -\infty < s \leq t \right) \leq \mathbb{E}\left( (X_{t+h} - \sum_{i \leq t} a_i X_i)^2 \bigg| X_s, -\infty < s \leq t \right)
\]

for any \( a_i \in \mathbb{R} \).
An Important Property of Linear Time Series

A linear time series \( \{X_t\} \), that is, one that follows the relation,

\[
X_t = \mu + \sum_{i=0}^{\infty} \psi_i A_{t-i},
\]

where \( \{A_t\} \) is iid with mean 0,

has the important property:

• Given \( \{X_s : -\infty < s \leq t\} \), the best mean square predictor of \( X_{t+h} \) is the same as the best linear predictor.
Best Linear Predictors

When is best linear best mean square?

- When we have a linear relationship and iid.
- When we have normality and 0 correlations.

What other cases?

- When we have 0 correlations and some kind of martingale property; that is, some kind of constant conditional expectation (as in the alternate definition of a linear time series).

Also, recall the Gauss-Markov theorem for $Y = X\beta + \epsilon$: least squares is best linear if $\mathbb{E}(\epsilon) = 0$ and $\mathbb{V}(\epsilon) = \sigma^2 I$. 
The Wold Decomposition

The Wold decomposition is a linear representation of any non-deterministic stationary time series.

If \( \{X_t\} \) is a nondeterministic stationary time series (i.e., it has constant mean, constant covariance at each lag \( h \), and has a stochastic component), then

\[
X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j} + V_t,
\]

where

1. \( \psi_0 = 1 \) and \( \sum_{j=0}^{\infty} \psi_j^2 < \infty \),
2. \( \{Z_t\} \) is white noise,
3. \( Z_t \in \mathcal{P}(\mathcal{F}_t) \),
4. \( \text{Cov}(Z_s, V_t) = 0 \) for all \( s \) and \( t \),
5. \( V_t \in \mathcal{P}(\mathcal{F}_{-\infty}) \),
6. \( V_t \) is deterministic.
Linear and Nonlinear Time Series

So does the Wold decomposition

\[ X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j} + V_t \]

say that all stochastic time series are linear?

No.

This gets to the subtle differences in independence and 0 correlations.
Nonlinear Time Series

Time series that follow ARMA models and most variations of these models are linear.

Statistical models often have the form

\[
\text{response} = \text{systematic component} + \text{random component}.
\]

In a linear time series \(\{X_t\}\), the systematic component is a linear function of previous values, \(\sum \phi_i x_{t-i}\) or else is a linear function of a well-behaved random process, \(\sum \theta_i w_{t-i}\).

Whenever the recursion is nonlinear, unusual behavior can result.
Consider the simple recursion

\[ x_t = 4x_{t-1}(1 - x_{t-1}) \quad \text{for} \ 0 < x_{t-1} < 1. \]

```r
chaos1<-function(x0,n) {
  for (i in 1:n) {
    x1<-4*x0*(1-x0)
    x0<-x1
  }
  print(x1)
}
```
Nonlinear Time Series

What types of time series are nonlinear?

There are four general types that are useful:

- models of squared quantities such as variances; these are often coupled with other models to allow stochastic volatility; ARIMA+GARCH, for example

- bilinear models

\[
X_t = c + \sum_{i=1}^{p} \phi_i X_{t-i} - \sum_{j=1}^{q} \theta_j A_{t-j} + \sum_{i=1}^{m} \sum_{j=1}^{s} \beta_{ij} X_{t-i} A_{t-j} + A_t
\]

- random coefficients models

\[
X_t = \sum_{i=1}^{p} (\phi_i + U_t^{(i)}) X_{t-i} + A_t
\]

- threshold models – Tsay describes a number of these in Chapters 3 and 4.
Fitting Time Series Models in R

In the standard \texttt{stats} package, the function \texttt{arima} fits general ARIMA models and the function \texttt{arima.sim} simulate data from such models.

There are also many formerly standalone software packages that have been incorporated into R, as separate packages.

One of these is \texttt{itsmr} from Colorado State. It has several functions, including the interesting \texttt{autofit}.

Another one that we have mentioned several times is \texttt{fGarch} from \texttt{Rmetrics}. It has a function \texttt{garchFit} that will fit ARIMA\(+\)GARCH models including most of the variations that Tsay mentions in Chapters 3 and 4.

It also has the function \texttt{garchSim} to simulate data from such models.