Workshop 2: the PEN

STAT 464/864 - Fall 2024
Discrete Time Series Analysis
Skye P. Griffith, Queen's University

Setup

Quarto renders from a blank slate: it runs code chunks *in order,* and based on an *empty environment.* That means you'll have to load any packages you'll be using, even if they're already loaded in your R session. You'll also have to load any data you plan to work with. Do all this at the beginning of the document, so the rest of your chunks are ready to run.

Packages

`itsmr` (from the textbook) If you need to install it, run the code

```
install.packages("itsmr")
```

in the console. The CONSOLE. Not the SCRIPT.

```
# Load ITSMR
```

Data

Info	Description
Dataset	Happy Australian Red Wine Sales (unit = kilolitres)
Times Sampled	(Monthly) Jan, 1980 – Oct, 1991 (142 total obs.)
Source	ITSM Time Series Package

It's included with the ITSMR package, so we don't need to load any external files.

Plotting

```
# --- Plotting
# make the world's most basic plot of the data
```

Okay… this doesn't really tell us what's going on. It doesn't demonstrate the data's most interesting patterns, and it doesn't give the viewer any context. Let's now create a scientifically meaningful plot of the data, with an appropriate x (time) axis.

```
# --- Time Axis: create a function [axis.wine()] to add this axis on the fly
# indicate 1 tick every January (total = 12)

# bottom edge of plot
# tick placement
# tick labels

# --- Base plot: create a function [plot.wine()] to speed-plot the data
# main title
# x-axis label
# y-axis label
# lines + points
# bullets (trust me)
# NO X-AXIS TICKS!
# add appropriate axis

# --- Test wine-plotting function
```

Analysis

The Plan

We want to decompose the data according to the classical model

$$X_t = m_t + s_t + Y_t \qquad (\star)$$

Think of it like this: X_t is a pen, and x_t are the lines drawn by the pen. We've run out of ink. So now we have to get a tube of the same coloured ink that will fit the pen's model.

1. **Eliminate** m_t :
   This is the shell – the *general shape* of the pen. Remove it.
   What your left with is $\hat{r}_m = X_t - \hat{m}_t$.
2. **Extract** s_t :
   Remove the spring + clicky components that are responsible for the pen's *repetitive pattern* of being open-closed-open-closed.
   Now you have $\hat{r} = X_t - \hat{m}_t - \hat{s}_t$ (the ink tube).
3. **Examine** Y_t :
   Look at this tube of *residual* ink. How long/wide is it? What colour is the ink?

Later in this course, we'll learn where to get more ink and how to reconstruct the pen.

Eliminate $\mathrm{m_t}$: The Body of the pen

Polynomial Regression

We suspect there may be a linear trend. Or, it's possible that a shallow quadratic may fit the data. I'm going to go with linear, that is, $p = 1$.

```
# --- Estimation
# --- Plotting
# Subtitle indicating model
```

Now we need to compute and plot the residuals: $(x_t - \hat{m}_t)$.
This is like the ink tube with the spring still attached.

```
# --- Residuals
# --- Plot
# (draw a dashed line at y = 0)
```

What we want to see is some kind of noisy but regular wave-like structure, since we plan to model seasonality, next. These residuals look ready to go. If there was a linear/curved trend remaining, we would want to consider a different model.

MA Smoothing

Let's repeat the same process, using a moving average smoother ($q = 7$) instead of polynomial regression. We'll plot the estimate and residuals side-by-side.

```
# --- Trend Estimation & Residuals
# --- Plotting
# Matrix of plots: 1 row, 2 columns
# Estimated m_t
# Residuals from m_t
```

Exponential Smoothing

We now repeat the whole thing using an exponential smoother with parameter $\alpha = 0.2$

```
# --- Trend Estimation & Residuals
# --- Plotting
# Matrix of plots: 2 rows, 1 column
# Estimated m_t
# Residuals from m_t
```

Extract s_t : The Spring

We apply this not to the original data, but to the residuals we got when we removed the trend. Just like how we can't remove the spring from a pen without opening up the pen. Let's use the residuals from our polynomial regression estimate.

Harmonic Regression

We suspect there is a seasonal component of period $d = 12$. Let's model this, and plot it over our polynomial regression residuals. Are there other periods that might be relevant?

```
# --- Seasonal Component & Residuals
```

The S1 Method

Let's do the same thing using the `season()` function from ITSMR. Again, we choose $d = 12$.

```
# --- Seasonal Component & Residuals
```

Plotting

```
# --- Plotting
# Estimated s_t: Harmonic Regression
# Estimated s_t: S1 Method
```

Examine $\mathbf{Y_t}$: The Residual Ink

```
# --- Plotting
# Residuals from s_t: harmonic regression
# Residuals from s_t: S1 method
```

Autocorrelation

We will learn what this is on Wednesday. Basically, it describes the correlation between any two points in the series, as a function of the time-distance between those two points. Noise, by definition, has no such correlation across time.

We want our residuals $\hat{Y}_t$ to be *noisy.* We want our pen's ink to be *inky.*

```
# --- Plotting ACFs
# Residuals from s_t: harmonic regression
# Residuals from s_t: S1 method
```

Values exceeding the dashed blue lines indicate significant autocorrelation.

These ACFs suggest that the residuals left by the season fit are less correlated across time than those we obtained via harmonic regression. Thus the season fit is preferred.

Putting the pen back together

```
# Put the de-noised series back together, plot it on top of the original data
```