STAT 626: Outlines of Lectures 19, 20 and 21
RW, Unit-Roots and ARCH Models (§5.3,§5.4)

1. **WN (Building Block of TS) $\Rightarrow$ RW, AR, ARIMA, SARIMA, ARCH/GARCH, ...**

2. **Random Walk vs AR(1):** $x_t = \phi x_{t-1} + w_t$,
$$H_0 : \phi = 1 \quad \text{vs} \quad H_1 : |\phi| < 1.$$

3. **Unit-Root Tests: DF, ADF, PP**.

4. **Why Unit-Root Test is Important in Economics and Finance?**

5. **Spurious Regression:** HW# 3, Due June 21 , 2:15 PM CST,
(a) Simulate two random walks $y_t, x_t, t = 1, \ldots, 100$, with initial values $x_0 = y_0 = 0$, using two independent $N(0, 1)$ white noises.

   (i) Plot y_t vs x_t in the (x, y)-plane. Describe the pattern of dependence (if any) you notice in the scatterplot.
(ii) Consider the linear regression model $y_t = \beta_0 + \beta_1 x_t + w_t$, and the null hypothesis $H_0 : \beta_1 = 0$ vs $H_1 : \beta_1 \neq 0$ at $\alpha = 0.05$. Do you expect H_0 would be rejected? Why?
(iii) Perform the test by fitting the regression line to the data using the least squares method and state your conclusion.

   (b) Repeat the above experiment 1000 times and count the number of H_0 rejected in (iii). Does it support your expectation in (ii)? If not, find an explanation for this phenomenon based on possible violations of assumptions of inference in regression models. (Hint: Recall the three assumptions of inference for linear regression models: (a) Independence, (b) Homogeniety of variances, (c) Normality, and check whether they hold for the data here.)

Random Walk; Engine of Financial Engineering

I. A Random Walk Down Wall Street, by Burton G. Malkiel

II. A Non-Random Walk Down Wall Street, by Andrew W. Lo & A. Craig MacKinlay

Books By Peter Bernstein:

III. Capital Ideas: The Improbable Origins of Modern Wall Street, (Free Press), 1991.

IV. Against the Gods: The Remarkable Story of Risk, (John Wiley & Son), 1996, Story of (Random Walk) Brownian Motion and how it Entered the World of Finance.

L. Bachelier Dissertation (1900).

Random Walk Hypothesis,

Efficient Market Hypothesis:

The weak form: All information about market prices is already reflected in the current stock price.

The strong form: All publicly available information about a company is already reflected in its stock price.

Lecture 20: ARCH Models

6. **Taking Care of Time-Varying Variances:** σ_t^2

7. **Time Series Decomposition:** $x_t = \mu_t + \sigma_t \varepsilon_t, \quad \mathbf{Var}(\sigma_t \varepsilon_t) = \sigma_t^2.$

8. **How to Model Time-Varying Variances?**
   Recall that Squared Residuals y_t^2 are Reasonable "Estimates" of σ_t^2:

$$y_t^2 \approx \sigma_t^2.$$

9. **Often y_t^2's appear more correlated than y_t's (Granger, 1970's).**

10. **AutoRegressive Conditionally Heteroscedastic (ARCH) Models:**(Engle, 1982)

$$y_t = \sigma_t \varepsilon_t,$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 y_{t-1}^2.$$

    AR Models for Squared Residuals y_t^2.

    This point of view is helpful in using the ACF and PACF of the series y_t^2 to identify the orders of the ARCH(p) models.

11. **Generalized ARCH (GARCH) Models**

    ARMA Models for Squared Residuals y_t^2.

Lecture 21: Regression with TS Errors

12. **Regression with Autocorrelated Errors (§5.6):**
    Taking Care of Correlations in the Residuals.

Coming Attractions

13. **Multivariate ARMAX Models (§5.8)**

~~and jointly estimating the parameters d and $\boldsymbol{\theta}$ using the Newton–Raphson method. If we are interested in a nonparametric estimator, using the conventional smoothed spectral estimator for the periodogram, adjusted for~~ the long memory component, say $g_k^d\ I(\omega_k)$ might be a possible approach.

5.3 Unit Root Testing

As discussed in the previous section, the use of the first difference $\nabla x_t = (1-B)x_t$ can be too severe a modification in the sense that the nonstationary model might represent an overdifferencing of the original process. For example, consider a causal AR(1) process (we assume throughout this section that the noise is Gaussian),

$$x_t = \phi x_{t-1} + w_t. \tag{5.26}$$

Applying $(1-B)$ to both sides shows that differencing, $\nabla x_t = \phi \nabla x_{t-1} + \nabla w_t$, or

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

where $y_t = \nabla x_t$, introduces extraneous correlation and invertibility problems. That is, while x_t is a causal AR(1) process, working with the differenced process y_t will be problematic because it is a non-invertible ARMA$(1,1)$.

A unit root test provides a way to test whether (5.26) is a random walk (the null case) as opposed to a causal process (the alternative). That is, it provides a procedure for testing

$$H_0\colon \phi = 1 \quad \text{versus} \quad H_1\colon |\phi| < 1.$$

An obvious test statistic would be to consider $(\widehat{\phi} - 1)$, appropriately normalized, in the hope to develop an asymptotically normal test statistic, where $\widehat{\phi}$ is one of the optimal estimators discussed in Chapter 3, §3.6. Unfortunately, the theory of §3.6 will not work in the null case because the process is nonstationary. Moreover, as seen in Example 3.35, estimation near the boundary of stationarity produces highly skewed sample distributions (see Figure 3.10) and this is a good indication that the problem will be atypical.

To examine the behavior of $(\widehat{\phi} - 1)$ under the null hypothesis that $\phi = 1$, or more precisely that the model is a random walk, $x_t = \sum_{j=1}^t w_j$, or $x_t = x_{t-1} + w_t$ with $x_0 = 0$, consider the least squares estimator of ϕ. Noting that $\mu_x = 0$, the least squares estimator can be written as

$$\widehat{\phi} = \frac{\frac{1}{n}\sum_{t=1}^n x_t x_{t-1}}{\frac{1}{n}\sum_{t=1}^n x_{t-1}^2} = 1 + \frac{\frac{1}{n}\sum_{t=1}^n w_t x_{t-1}}{\frac{1}{n}\sum_{t=1}^n x_{t-1}^2}, \tag{5.27}$$

where we have written $x_t = x_{t-1} + w_t$ in the numerator; recall that $x_0 = 0$ and in the least squares setting, we are regressing x_t on x_{t-1} for $t = 1, \dots, n$. Hence, under H_0, we have that

To obtain its approximate dist. one has to rely on LLN, CLT, etc.

What happens to this average when n goes to infinity?

$$\widehat{\phi} - 1 = \frac{\frac{1}{n\sigma_w^2}\sum_{t=1}^n w_t x_{t-1}}{\frac{1}{n\sigma_w^2}\sum_{t=1}^n x_{t-1}^2}. \tag{5.28}$$

Consider the numerator of (5.28). Note first that by squaring both sides of $x_t = x_{t-1} + w_t$, we obtain $x_t^2 = x_{t-1}^2 + 2x_{t-1}w_t + w_t^2$ so that

$$x_{t-1}w_t = \frac{1}{2}(x_t^2 - x_{t-1}^2 - w_t^2),$$

and summing,

$$\frac{1}{n\sigma_w^2}\sum_{t=1}^n x_{t-1}w_t = \frac{1}{2}\Big(\frac{x_n^2}{n\sigma_w^2} - \frac{\sum_{t=1}^n w_t^2}{n\sigma_w^2}\Big).$$

Because $x_n = \sum_1^n w_t$, we have that $x_n \sim N(0, n\sigma_w^2)$, so that $\frac{1}{n\sigma_w^2}x_n^2 \sim \chi_1^2$, the chi-squared distribution with one degree of freedom. Moreover, because w_t is white Gaussian noise, $\frac{1}{n}\sum_1^n w_t^2 \to_p \sigma_w^2$, or $\frac{1}{n\sigma_w^2}\sum_1^n w_t^2 \to_p 1$. Consequently, $(n \to \infty)$

$$\frac{1}{n\sigma_w^2}\sum_{t=1}^n x_{t-1}w_t \xrightarrow{d} \frac{1}{2}(\chi_1^2 - 1). \tag{5.29}$$

Next we focus on the denominator of (5.28). First, we introduce standard Brownian motion.

Definition 5.1 *A continuous time process* $\{W(t);\, t \ge 0\}$ *is called* **standard Brownian motion** *if it satisfies the following conditions:*

(i) $W(0) = 0$;
(ii) $\{W(t_2) - W(t_1), W(t_3) - W(t_2), \dots, W(t_n) - W(t_{n-1})\}$ *are independent for any collection of points,* $0 \le t_1 < t_2 \cdots < t_n$, *and integer* $n > 2$;
(iii) $W(t + \Delta t) - W(t) \sim N(0, \Delta t)$ *for* $\Delta t > 0$.

The result for the denominator uses the functional central limit theorem, which can be found in Billlingsley (1999, §2.8). In particular, if $\xi_1, \dots, \xi_n$ is a sequence of iid random variables with mean 0 and variance 1, then, for $0 \le t \le 1$, the continuous time process

$$S_n(t) = \frac{1}{\sqrt{n}}\sum_{j=1}^{[\![nt]\!]} \xi_j \;\xrightarrow{d}\; W(t), \tag{5.30}$$

as $n \to \infty$, where $[\![\ ]\!]$ is the greatest integer function and $W(t)$ is standard Brownian motion on $[0, 1]$. Note the under the null hypothesis, $x_s = w_1 + \cdots + w_s \sim \mathrm{N}(0, s\sigma_w^2)$, and based on (5.30), we have $\frac{x_s}{\sigma_w\sqrt{n}} \to_d W(s)$. From this fact, we can show that $(n \to \infty)$

$$\sum_{t=1}^n \Big(\frac{x_{t-1}}{\sigma_w\sqrt{n}}\Big)^2 \frac{1}{n} \xrightarrow{d} \int_0^1 W^2(t)\, dt\,. \tag{5.31}$$

This looks different.

The denominator in (5.28) is off from the left side of (5.31) by a factor of n^{-1}, and we adjust accordingly to finally obtain ($n \to \infty$),

$$n(\widehat{\phi} - 1) = \frac{\frac{1}{n\sigma_w^2} \sum_{t=1}^{n} w_t x_{t-1}}{\frac{1}{n^2\sigma_w^2} \sum_{t=1}^{n} x_{t-1}^2} \xrightarrow{d} \frac{\frac{1}{2}\left(\chi_1^2 - 1\right)}{\int_0^1 W^2(t)\, dt}. \tag{5.32}$$

The test statistic $n(\widehat{\phi} - 1)$ is known as the unit root or Dickey-Fuller (DF) statistic (see Fuller, 1996), although the actual DF test statistic is normalized a little differently. Because the distribution of the test statistic does not have a closed form, quantiles of the distribution must be computed by numerical approximation or by simulation. The R package `tseries` provides this test along with more general tests that we mention briefly.

Toward a more general model, we note that the DF test was established by noting that if $x_t = \phi x_{t-1} + w_t$, then $\nabla x_t = (\phi - 1)x_{t-1} + w_t = \gamma x_{t-1} + w_t$, and one could test $H_0 : \gamma = 0$ by regressing ∇x_t on x_{t-1}. They formed a Wald statistic and derived its limiting distribution [the previous derivation based on Brownian motion is due to Phillips (1987)]. The test was extended to accommodate AR(p) models, $x_t = \sum_{j=1}^{p} \phi_j x_{t-j} + w_t$, as follows. Subtract x_{t-1} from the model to obtain

$$\nabla x_t = \gamma x_{t-1} + \sum_{j=1}^{p-1} \psi_j \nabla x_{t-j} + w_t, \tag{5.33}$$

where $\gamma = \sum_{j=1}^{p} \phi_j - 1$ and $\psi_j = -\sum_{j=i}^{p} \phi_i$ for $j = 2, \ldots, p$. For a quick check of (5.33) when $p = 2$, note that $x_t = (\phi_1 + \phi_2)x_{t-1} - \phi_2(x_{t-1} - x_{t-2}) + w_t$; now subtract x_{t-1} from both sides. To test the hypothesis that the process has a unit root at 1 (i.e., the AR polynoimial $\phi(z) = 0$ when $z = 1$), we can test $H_0 : \gamma = 0$ by estimating γ in the regression of ∇x_t on $x_{t-1}, \nabla x_{t-1}, \ldots, \nabla x_{t-p+1}$, and forming a Wald test based on $t_\gamma = \widehat{\gamma}/\text{se}(\widehat{\gamma})$. This test leads to the so-called augmented Dickey-Fuller test (ADF). While the calculations for obtaining the asymptotic null distribution change, the basic ideas and machinery remain the same as in the simple case. The choice of p is crucial, and we will discuss some suggestions in the example. For ARMA(p, q) models, the ADF test can be used by assuming p is large enough to capture the essential correlation structure; another alternative is the Phillips-Perron (PP) test, which differs from the ADF tests mainly in how they deal with serial correlation and heteroskedasticity in the errors.

One can extend the model to include a constant, or even non-stochastic trend. For example, consider the model

$$x_t = \beta_0 + \beta_1 t + \phi x_{t-1} + w_t.$$

If we assume $\beta_1 = 0$, then under the null hypothesis, $\phi = 1$, the process is a random walk with drift β_0. Under the alternate hypothesis, the process is a causal AR(1) with mean $\mu_x = \beta_0(1 - \phi)$. If we cannot assume $\beta_1 = 0$, then the interest here is testing the null that $(\beta_1, \phi) = (0, 1)$, simultaneously, versus the alternative that $\beta_1 \neq 0$ and $|\phi| < 1$. In this case, the null hypothesis is that the process is a random walk with drift, versus the alternative hypothesis that the process is stationary around a global trend (consider the global temperature series examined in Example 2.1).

See Example 2.6

Example 5.3 Testing Unit Roots in the Glacial Varve Series

In this example we use the R package `tseries` to test the null hypothesis that the log of the glacial varve series has a unit root, versus the alternate hypothesis that the process is stationary. We test the null hypothesis using the available DF, ADF and PP tests; note that in each case, the general regression equation incorporates a constant and a linear trend. In the ADF test, the default number of AR components included in the model, say k, is $[\![(n-1)^{\frac{1}{3}}]\!]$, which corresponds to the suggested upper bound on the rate at which the number of lags, k, should be made to grow with the sample size for the general ARMA(p, q) setup. For the PP test, the default value of k is $[\![.04n^{\frac{1}{4}}]\!]$.

```
library(tseries)
adf.test(log(varve), k=0)                  # DF test
```

```
Dickey-Fuller = -12.8572, Lag order = 0, p-value < 0.01
alternative hypothesis: stationary
```

```
adf.test(log(varve))                       # ADF test
```

```
Dickey-Fuller = -3.5166, Lag order = 8, p-value = 0.04071
alternative hypothesis: stationary
```

```
pp.test(log(varve))                        # PP test
```

```
Dickey-Fuller Z(alpha) = -304.5376,
Truncation lag parameter = 6, p-value < 0.01
alternative hypothesis: stationary
```

In each test, we reject the null hypothesis that the logged varve series has a unit root. The conclusion of these tests supports the conclusion of the previous section that the logged varve series is long memory rather than integrated.

5.4 GARCH Models

Recent problems in finance have motivated the study of the volatility, or variability, of a time series. Although ARMA models assume a constant variance, models such as the autoregressive conditionally heteroscedastic or ARCH model, first introduced by Engle (1982), were developed to model changes in volatility. These models were later extended to generalized ARCH, or GARCH models by Bollerslev (1986).

In §3.8, we discussed the return or growth rate of a series. For example, if x_t is the value of a stock at time t, then the return or relative gain, y_t, of the stock at time t is