Statistics 5350/7110
Forecasting

Lecture 3
Stationary Time Series

Professor Stine

Admin Issues

- Questions?

- Assignments
  - Assignment #1 due next Thursday

- Quick review
  - White noise: simple data
  - Quantile plot, specifically normal quantile plot
  - Random walk and the Durbin-Watson test [much more to come]
  - Autoregression
  - Signal + noise heuristic

Today's Topics

Textbook § 2.1-2.3

- Stationarity
  - Assumption that underlies data analysis
- Autocovariance and autocorrelation
  - Notation
  - Effect of stationarity
- Estimation
  - Means
  - Autocovariance and autocorrelation

Stationarity

- Averaging a time series
  - "Time series" = { ... X_{t-1}, X_t, X_{t+1}, ... } is a collection of random variables, a discrete-time stochastic process
  - What does it mean to take the expectation of an element of a time series?

$$E(X_t) = ?$$

  - Recall discussion of the variance of the random walk: What is possible vs what we have seen.
  - How would you estimate this mean? We need to make an assumption.
- Stationarity (second-order stationarity)
  - Allows us to average over time
  - Assuming finite variance, a time series is stationary if the following hold

$$E(X_t) = \mu$$

$$E\left((X_t - \mu)(X_s - \mu)\right) = \text{Cov}(X_t, X_s) = \gamma_x(t, s) = \gamma_x(t - s)$$

  where the autocovariance function γ_x is symmetric, $\gamma_x(t-s) = \gamma_x(s-t)$.
- Challenge for data analysis
  - How from one realization do you recognize that a time series comes from a stationary process?

Examples

- Are these processes stationary?
  - White noise
  - Smoothing white noise with a linear filter (a.k.a., a moving average)
  - Autoregression
  - Random walk
- Second-order stationarity requires covariances invariant of a time shift
  - Mean is time invariant
  - Autocovariance function depends only on the time separation of the elements.
- Computing the autocovariance (Property 2.7)
  - Key property of covariances of linear combinations (weighted sums)

$$\text{Cov}\left(\sum_{j=1}^{m} a_j X_j,\ \sum_{k=1}^{r} b_j Y_k\right) = \sum_j \sum_k a_j\, b_k \text{Cov}(X_j, Y_k)$$

  - Simpler in matrix form. For random vectors X and Y and scalar vectors a and b

$$\text{Cov}(a^\top X, b^\top Y) = a^\top \left[\text{Cov}(X_j, Y_k)\right] b$$

Autocovariances

- White noise
  - Cov$(w_t, w_t) =$ Var$(w_t) = \sigma_w^2$ and Cov$(w_s, w_t) = 0$ for s ≠ t
- Moving average of white noise
  - Three term moving average $X_t = a_{-1}w_{t-1} + a_0 w_t + a_1 w_{t+1}$
  - Variance

$$\text{Cov}(X_t, X_t) = \text{Var}(X_t) = \sigma_w^2 (a_{-1}^2 + a_0^2 + a_1^2)$$

  - Overlap determines covariance

$$\text{Cov}(X_{t+1}, X_t) = \sigma_w^2 (a_{-1}a_0 + a_0 a_1)$$
$$\text{Cov}(X_{t+2}, X_t) = a_{-1}a_1\sigma_w^2$$
$$\text{Cov}(X_{t+3}, X_t) = 0$$

- Are these stationary processes?

Autocovariances

- First-order autoregression AR(1) process
  - What's the time origin of the process?

$$X_t = \phi X_{t-1} + w_t$$

- Moving average representation
  - Write the process as a moving average, assuming white noise extends infinitely far back

$$X_t = w_t + \phi w_{t-1} + \phi^2 w_{t-2} + \cdots = \sum_{s=0}^{\infty} \phi^s w_{t-s}$$

and where we assume that $|\phi| < 1$. Follows that

$$\text{Var}(X_t) = \sigma_w^2 \frac{1}{1 - \phi^2}$$

  - For the other autocovariances

$$\text{Cov}(X_{t+h}, X_t) = \sigma_w^2 \phi^h \sum_{j=0}^{\infty} \phi^{2j} = \phi^h \, \text{Var}(X_t)$$

- Stationary? Under what conditions?

Autocovariances

- First-order autoregression revisited
  - Suppose we know that the process is stationary, that $|\phi| < 1$

$$X_t = \phi X_{t-1} + w_t$$

- If known to be stationary
  - What does the mean have to be?
  - Finding autocovariances is simplified if we know a process is stationary
  - Take the variance of both sides, noting $\text{Cov}(w_s, X_t) = 0$ for $s > t$

$$\text{Var}(X_t) = \text{Var}(\phi X_{t-1} + w_t) = \phi^2 \text{Var}(X_{t-1}) + \sigma_w^2$$

If stationary, then $\text{Var}(X_t) = \text{Var}(X_{t-1})$ and arrive at the prior expression

$$\text{Var}(X_t) = \sigma_w^2 \frac{1}{1 - \phi^2}$$

  - For the rest of the autocovariances

$$\text{Cov}(X_{t+1}, X_t) = \text{Cov}(\phi\, X_t + w_{t+1}, X_t) = \phi\, \text{Var}(X_t)$$

$$\text{Cov}(X_{t+2}, X_t) = \text{Cov}(\phi^2\, X_t + w_{t+2} + \phi w_{t+1}, X_t) = \phi^2\, \text{Var}(X_t)$$

Wold Representation

- Every second-order stationary time series $\{X_t\}$ has a moving average representation of the form

$$X_t = \mu + \sum_{j=0}^{\infty} \psi^j w_{t-j}$$

Property 2.21

  - Weights are square summable

$$\sum \psi_j^2 < \infty.$$

  - By convention, $\psi_0 = 1$.
- Extensive implications
  - Every stationary process can be viewed as filtering some white noise process
  - Causal representation, where the current value depends on past white noise
  - The influence of the past drops off
  - Expression for the best squared-error predictor and its mean squared error

More Autocovariances

- Autocovariances exist for non-stationary processes
  - Depend on the time index
- Random walk
  - Define as

$$X_t = w_1 + w_2 + \cdots + w_t = \sum_{s=1}^{t} w_s$$

  - Coefficients of common terms as in moving average determine autocovariances
  - Variance is easy

$$\text{Cov}(X_t, X_t) = \text{Var}(X_t) = t\,\sigma_w^2$$

  - Covariances

$$\text{Cov}(X_s, X_t) = \text{Var}(X_t) = \min(s, t)\,\sigma_w^2 \qquad s, t = 1,2,\ldots$$

Autocorrelation

- Interpreting autocovariance
  - Suppose that $\{X_t\}$ is stationary.
  - If $\gamma_X(1) = \mathrm{Cov}(X_{t+1}, X_t) = 200000$, are these random variables strongly or weakly dependent?
  - Covariance depends on the scale of X_t.
- Autocorrelation standardizes the autocovariance
  - Same relationships between covariance and correlation, just with time lags
  - Autocorrelation standardizes autocorrelation to scale of -1 to +1
  - General case

$$\mathrm{Corr}(X_t, X_s) = \rho(t, s) = \frac{\mathrm{Cov}(X_t, X_s)}{\sqrt{\mathrm{Var}(X_t)\mathrm{Var}(X_s)}} = \frac{\gamma(t, s)}{\sqrt{\gamma(t, t), \gamma(s, s)}}$$

  - Stationary case

$$\mathrm{Corr}(X_t, X_s) = \rho(t - s) = \frac{\gamma(t - s)}{\gamma(0)}$$

- Roles in theory and practice
  - Easier to do math with covariances
  - Correlations make more sense in data analysis

Estimating the Mean

Textbook §2.3

- Estimating the mean of a stationary process
  - Sample mean is the usual choice, but is it the best estimator?

$$\overline{X} = \frac{1}{n}\sum_{t=1}^{n} X_t$$

- Properties of the average
  - Variance of the average depends on the autocovariance of the process

$$n^2 \text{Var}\left(\overline{X}\right) = \text{Var}\left(\sum_{t=1}^{n} X_t\right) = \sum_{s=1}^{n}\sum_{t=1}^{n} \gamma(s-t) = 1^{\top}\Gamma\, 1$$

  where Γ denotes the n x n covariance matrix $\Gamma = [\gamma_x(i-j)]$ and 1 is a vector of n 1's.

  - Summing up the elements of Γ gives

$$\text{Var}\left(\overline{X}\right) = \sum_{j=-n}^{n} \left(1 - \frac{|h|}{n}\right)\gamma_x(h)$$

Equation 2.20

  - Generalized least squares provides an alternative estimator which has smaller MSE, but you have to know the autocovariances $\gamma_x(h)$ in order to compute it!

Estimating the Mean

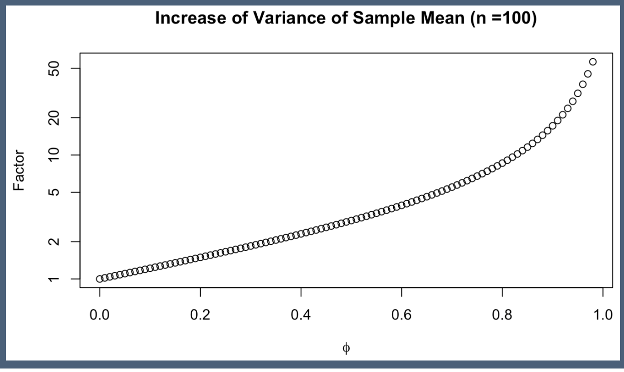
- Effects of covariances on variance of the mean
- White noise
  - Variance of the mean is the usual expression, namely

$$\text{Var}(\overline{X}) = \frac{\sigma_w^2}{n}$$

- Autoregression with coefficient ϕ
  - General expression is "messy"

$$\text{Var}(\overline{X}) = \frac{1}{n^2}\text{Var}\left(\sum_{t=1}^{n} X_t\right) = \frac{\sigma_w^2}{n(1-\phi^2)}\underbrace{\frac{2}{n}\left(\frac{n}{2} + (n-1)\phi + (n-2)\phi^2 + \cdots + \phi^{n-1}\right)}$$

  - Numerical example
    How much larger is the variance of the mean compared to the usual calculation that would use $\text{Var}(X_t)/n$?



$0 < \phi$

Estimating the Autocorrelations

- Estimating the autocovariances
  - Use the sample mean

$$\hat{\gamma}_x(h) = \frac{1}{n}\sum_{t=1}^{n-h}(X_{t+h}-\overline{X})(X_t-\overline{X}) \qquad h = 0,1,2,\ldots,n-1$$

  - Biased estimator since fewer terms as h increases (but leads to a p.s.d estimator $\widehat{\Gamma}$)
- Estimating the autocorrelation
  - Use the estimated autocovariances

$$\hat{\rho}_x(h) = \frac{\hat{\gamma}_x(h)}{\hat{\gamma}_x(0)}$$

Equation 2.21

- Properties of the estimated autocorrelation
  - As for the sample mean, properties of $\hat{\rho}_x(h)$ depend on the properties of the true process
  - The estimated $\hat{\rho}_x(h)$ are generally more autocorrelated than the process itself!

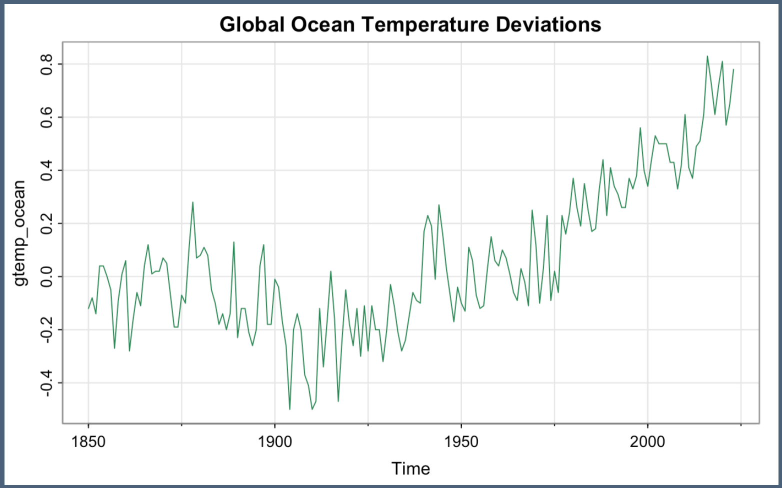
$$\text{Cov}(\hat{\rho}_x(r+h),\hat{\rho}_x(r)) \approx \frac{1}{n}\sum_{j=-\infty}^{\infty}\rho_x(j+h)\,\rho_x(j) \quad \text{and} \quad \text{Var}(\hat{\rho}_x(r)) \approx \frac{1}{n}\sum_{j=-\infty}^{\infty}\rho_x(j)^2$$

See Property 2.28

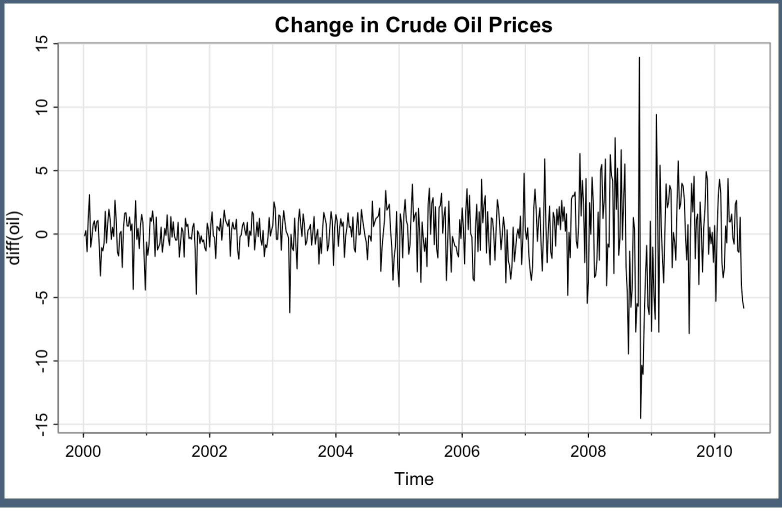
More on this in the next class...

Many Time Series Don't Look Stationary

- Ocean temperatures



- Change in crude prices



R Examples

- Two types of plots
  - Lag plots
  - Estimates of the autocorrelation function

- Cautions for correlation estimates
  - Correlation only measures linear dependence
  - Influence of nonlinearity and outliers

What's Next?

- Examples of autocorrelation estimates
- Multiple time series
  - Cross-correlation
  - Leading and lagging series
- Spurious correlations
- Dealing with non-stationary data
  - Removing a deterministic trend
  - Differencing
  - Logs