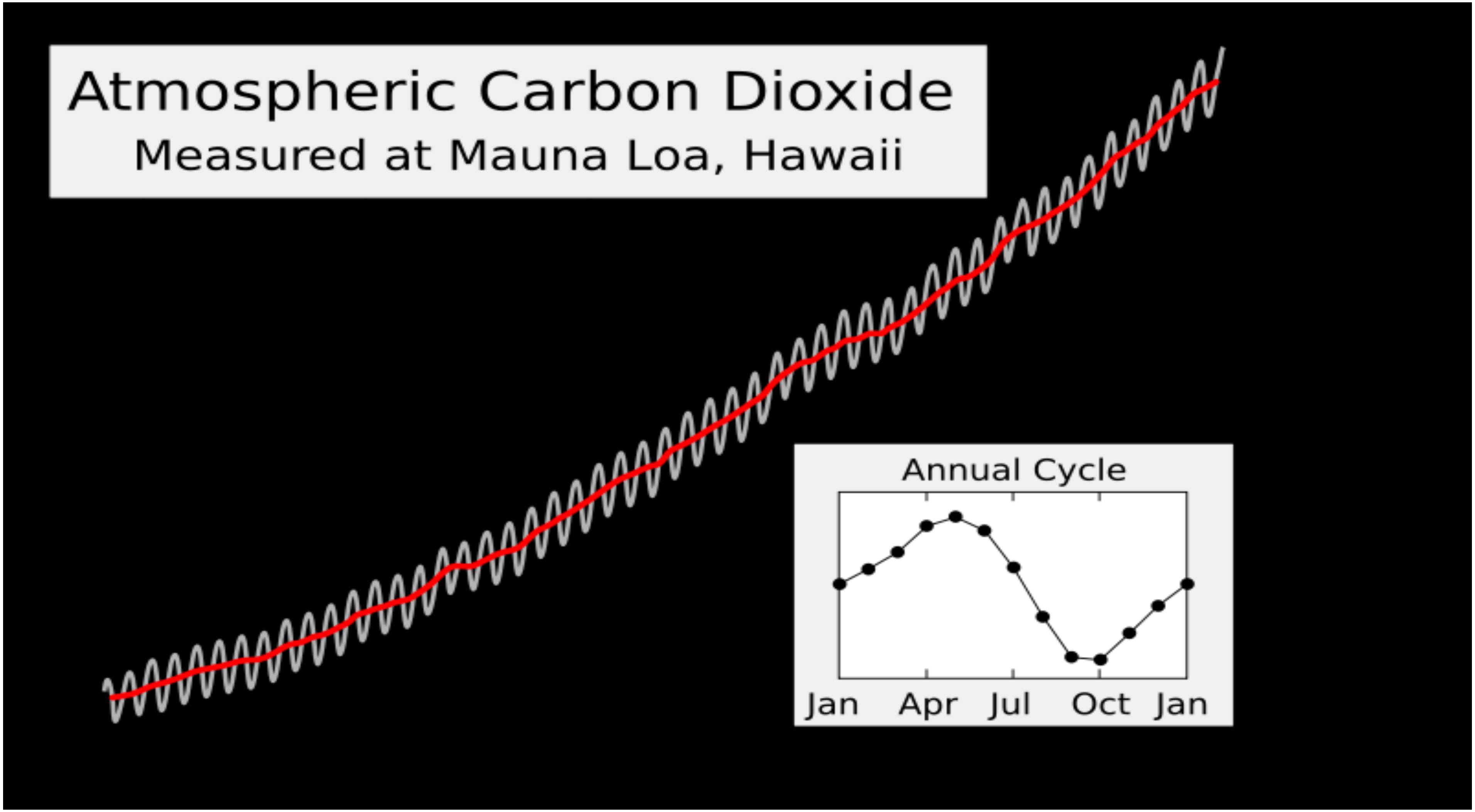
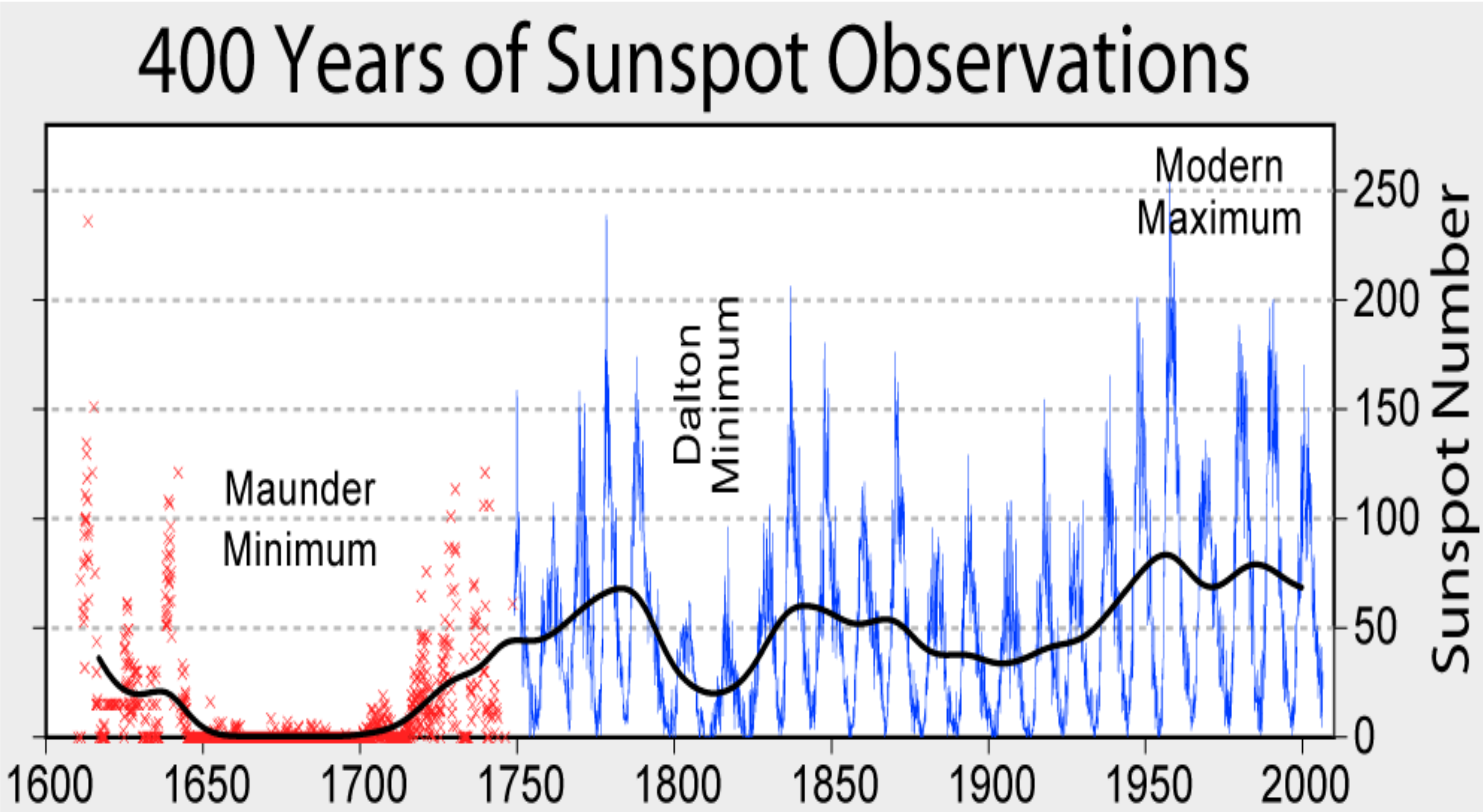
Time Series

- Look at the data!
- Common Models
- Multivariate Data
- Cycles/Seasonality
- Filters

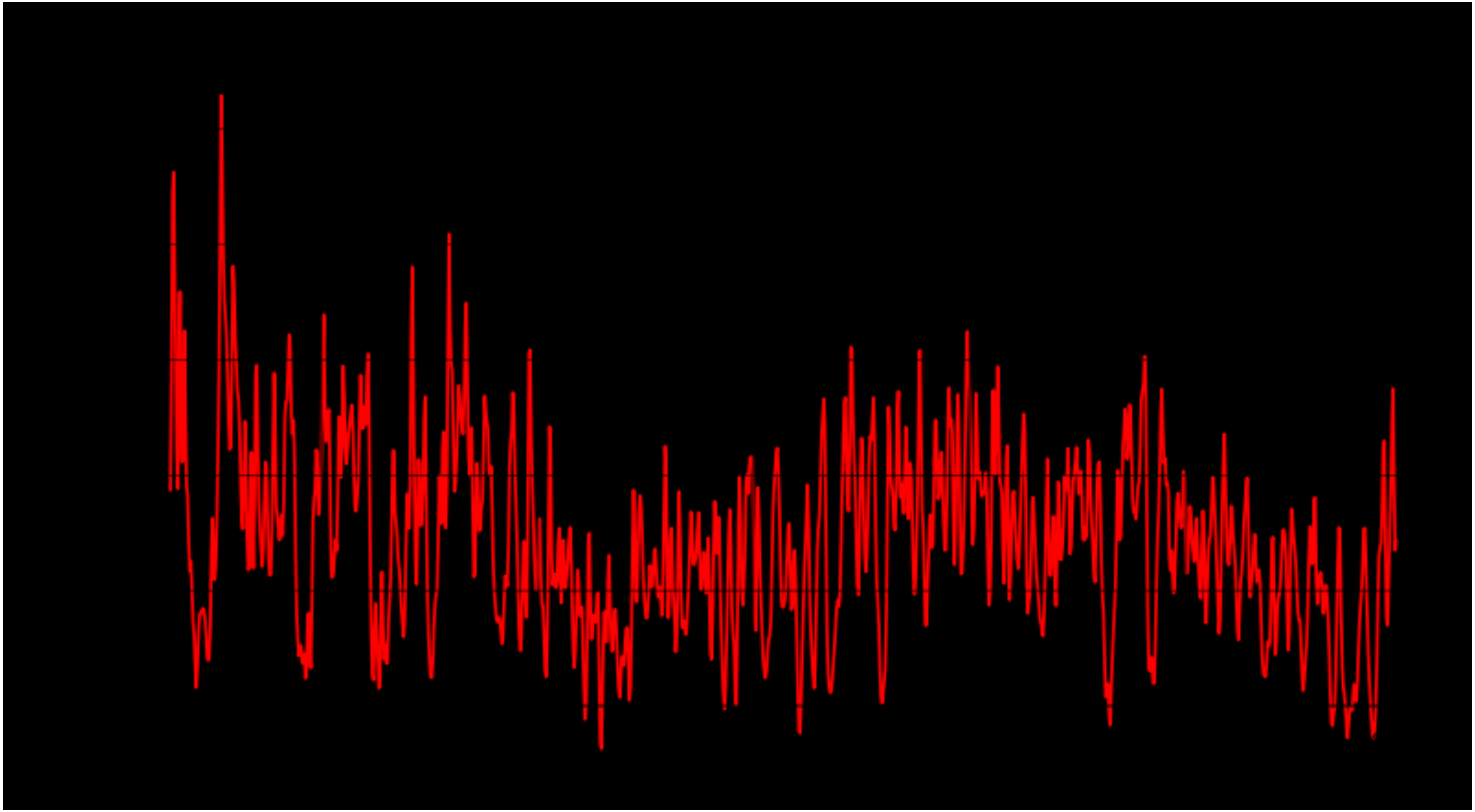
Atmospheric CO_2



Years: 1958 to now; vertical scale 300 to 400ish


400 Years of Sunspot Observations
Modern
Maximum
Dalton
Minimum
Maunder
Minimum
Sunspot Number
250
200
150
100
50
0
1600
1650
1700
1750
1800
1850
1900
1950
2000

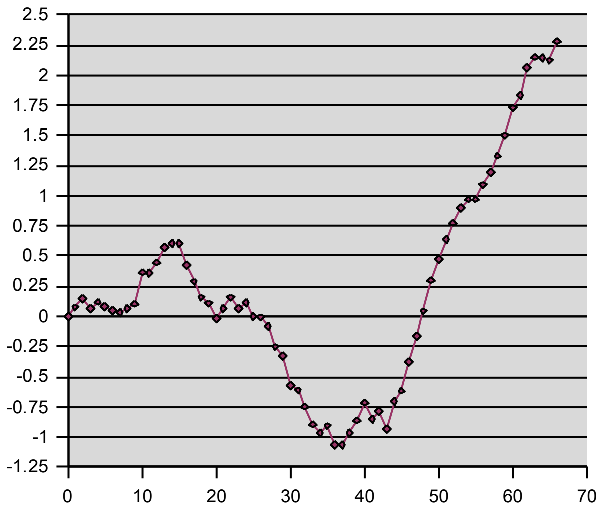
Ancient sunspot data

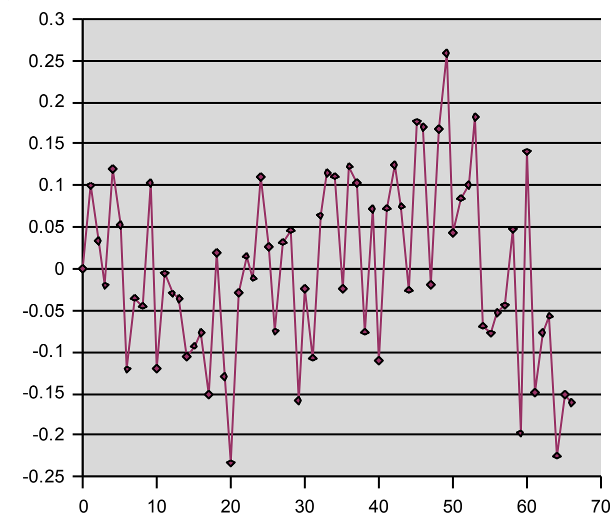


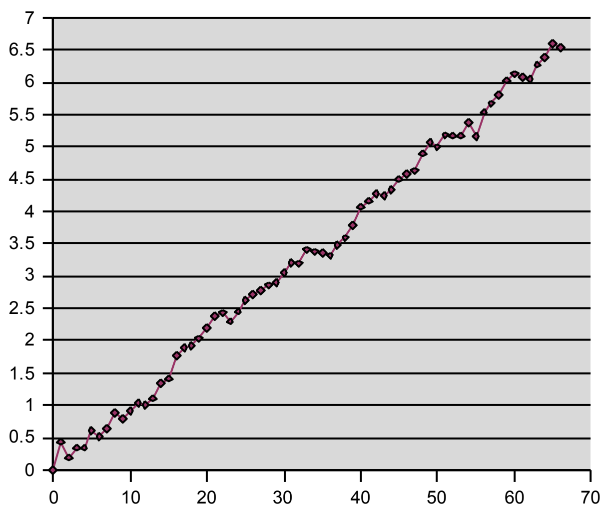
Our Basic Procedure

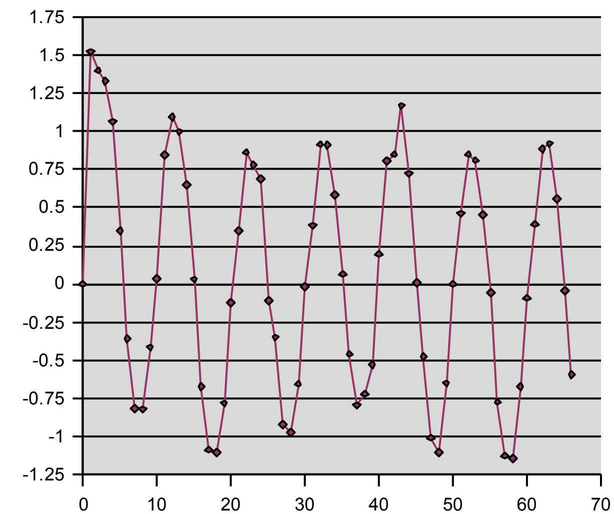
1. Look at the data
2. Quantify any pattern you see
3. Remove the pattern
4. Look at the residuals
5. Repeat at step 2 until no patterns left

One of these things is not like the others









Stationarity

- The upper-right-corner plot is Stationary.
- Mean doesn't change in time
  - no Trend
  - no Seasons (known frequency)
  - no Cycles (unknown frequency)
- Variance doesn't change in time
- Correlations don't change in time
  - Up to here, weakly stationary
- Joint Distributions don't change in time
  - That makes it strongly stationary

Basic Notation

- Time is “t”, not “n”
  - even though it's discrete
- State (value) is Y, not X
  - to avoid confusion with x-axis, which is time.
- Value at time t is Y_t, not Y(t)
  - because time is discrete

Detrending: deterministic trend

Trend stationary Process (TSP)

- Fit a plain linear regression, then subtract it out:
  - Fit $Y_t = m*t + b$,
  - New data is $Z_t = Y_t - m*t - b$
  - Or use quadratic fit, exponential fit, etc.

Detrending: stochastic trend

Difference Stationary Process (DSP)

- Differencing
  - For linear trend, new data is $Z_t = Y_t - Y_{t-1}$
  - To remove quadratic trend, do it again:
    - $W_t = Z_t - Z_{t-1} = Y_t - 2Y_{t-1} + Y_{t-2}$
  - Like taking derivatives

  - What's the equivalent if you think the trend is exponential, not linear?

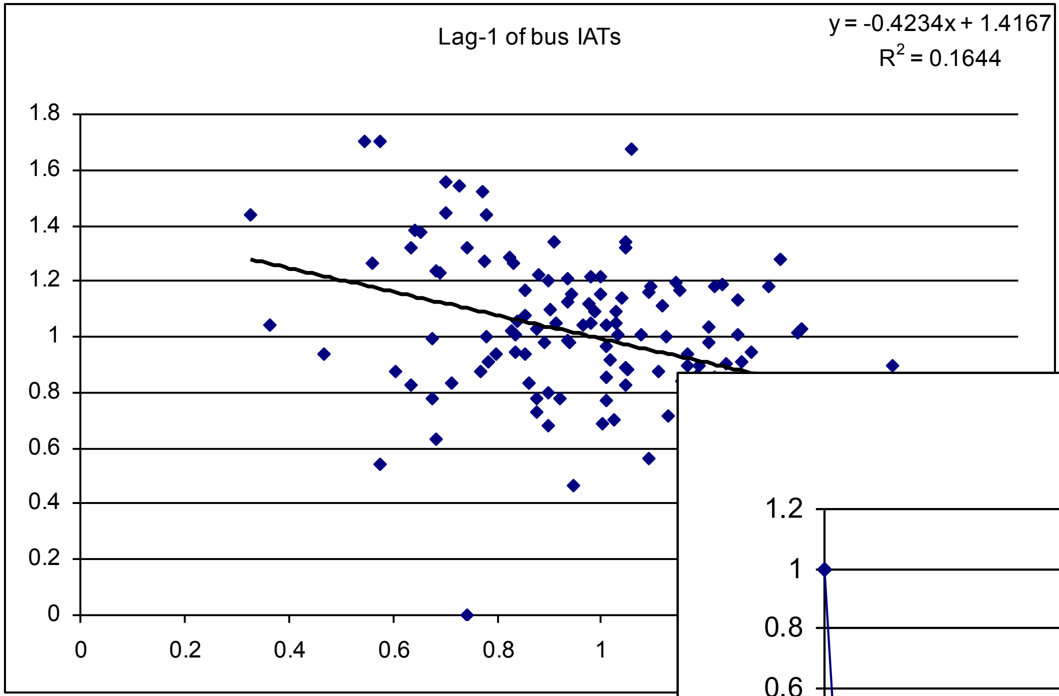
Hard to decide: regression or differencing?

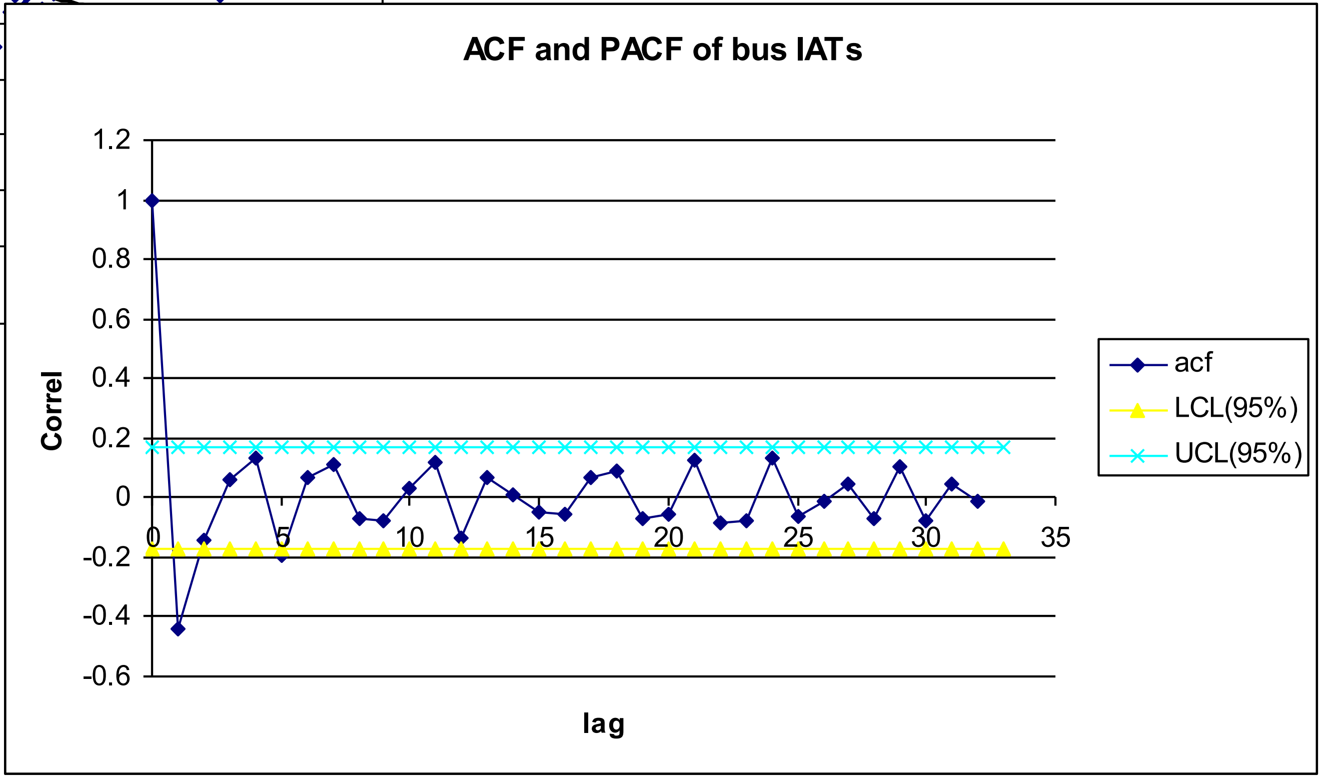
Problems with wrong choice of model		Assumed Model	
		TSP	DSP
Correct Model	TSP	√	Error becomes MA
	DSP	Error becomes Heteroscedastic	√

Autocorrelation Function

- How correlated is the series with itself at various lag values?
- E.g. If you plot Y_{t+1} versus Y_t and find the correlation, that's the correl. at lag 1
- ACF lets you calculate all these correls. without plotting at each lag value.
- ACF is a basic building block of time series analysis.

Fake data on bus IATs





Properties of ACF

- At lag 0, ACF=1
- Symmetric around lag 0
- Approx. confidence-interval bars around ACF=0
  - To help you decide when ACF drops to near-0
- Less reliable at higher lags
- Often assume ACF dies off fast enough so its absolute sum is finite.
  - If not, called “long-term memory”; e.g.
    - River flow data over many decades
    - Traffic on computer networks

ACF at lag h

$$C_h = \frac{1}{N-h}\sum_{t=1}^{N-h}(Y_t - \bar{Y})(Y_{t+h} - \bar{Y})$$

$$C_h = \frac{1}{N}\sum_{t=1}^{N-h}(Y_t - \bar{Y})(Y_{t+h} - \bar{Y})$$

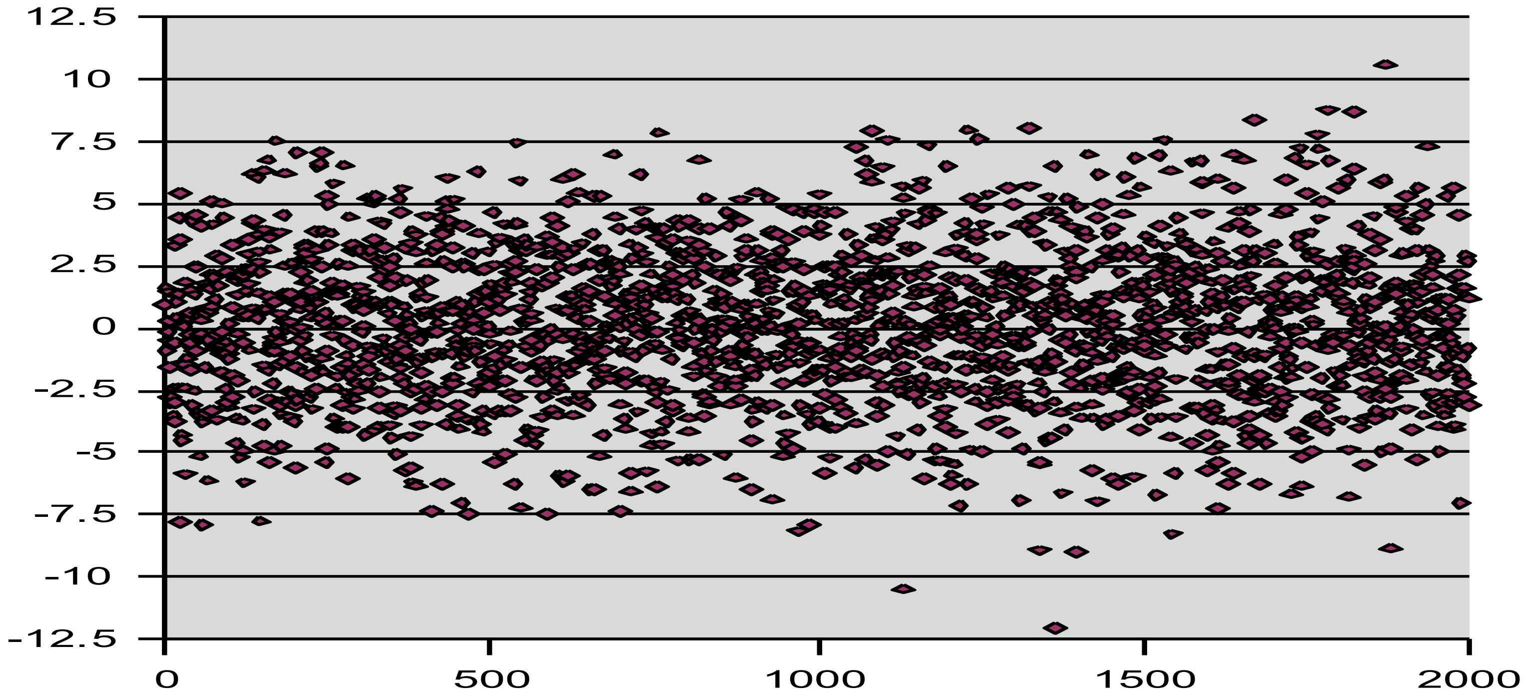
- Y-bar is mean of whole data set
  - Not just mean of N-h data points
- Left side: old way, can produce correl>1
- Right side: new way

Common Models

- White Noise
- AR
- MA
- ARMA
- ARIMA
- SARIMA
- ARMAX
- Kalman Filter
- Exponential Smoothing, trend, seasons

White Noise

- Sequence of I.I.D. Variables ε_t
- mean=zero, Finite std.dev., often unknown
- Often, but not always, Gaussian



AR: AutoRegressive

- Order 1: $Y_t = a*Y_{t-1} + \varepsilon_t$

  E.g. New = (90% of old) + random fluctuation
- Order 2: $Y_t = a_1*Y_{t-1} + a_2*Y_{t-2} + \varepsilon_t$
- Order p denoted AR(p)
- p=1,2 common; >2 rare
- AR(p) like p'th order ODE
- AR(1) not stationary if |a|>=1
- $E[Y_t] = 0$, can generalize

Things to do with AR

- Find appropriate order
- Estimate coefficients
  - via Yule-Walker eqn.
- Estimate std.dev. of white noise
- If estimated $|a|>0.98$:Unit Root Test.

Dickey – Fuller Unit Root Test

- Model : $Y_t = a^*Y_{t-1} + \varepsilon_t$
- Test for $H_0 : a = 1$ vs. $H_1 : a < 1$.
- If H_0 accepted, series non-stationary

Extensions:

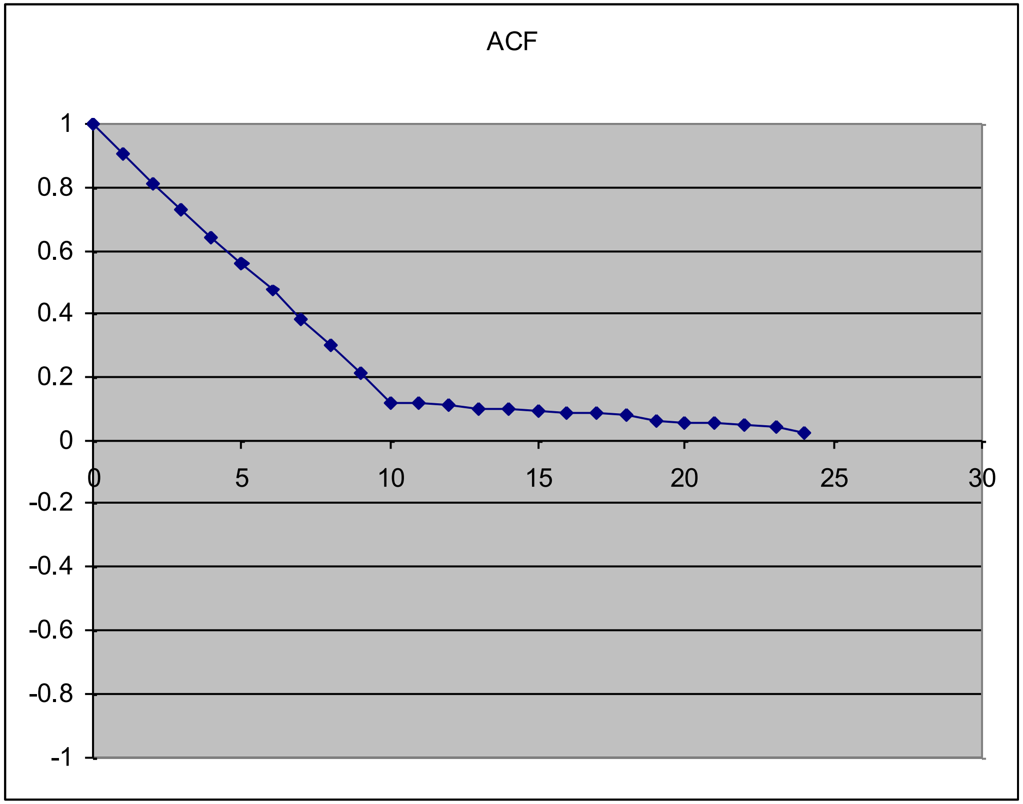
- ADF (additional AR terms),
- PP (switch H_0 and H_1),
- KPSS (allowing for heteroscedasticity)

MA: Moving Average

- Order 1:
  - $Y_t = b_0\varepsilon_t + b_1\varepsilon_{t-1}$
- Order q: MA(q)
- Important in theory of filters
- Stationary regardless of b values
- $E[Y_t] = 0$, can generalize

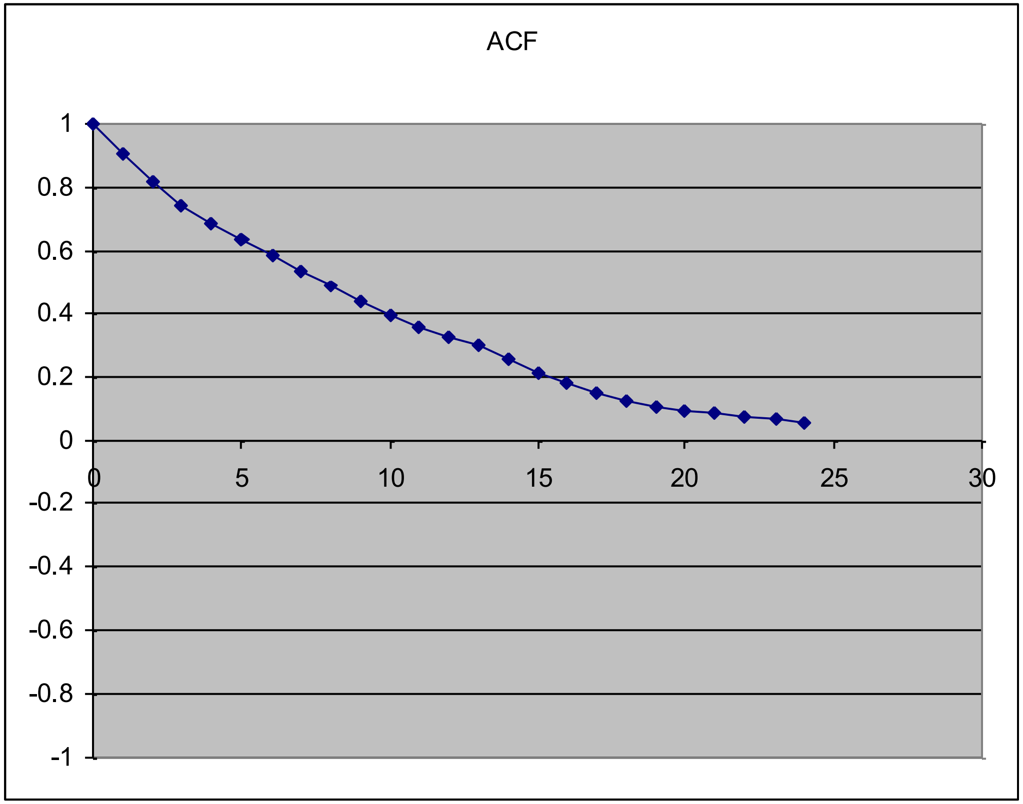
ACF of an MA process

- Drops to zero after lag=q
- That's a good way to determine what q should be!



ACF of an AR process?

- Never completely dies off, not useful for finding order p.
- AR(1) has exponential decay in ACF
- Instead, use Partial ACF = PACF, which dies after lag=p
- PACF of MA never dies.



ARMA

- ARMA(p,q) combines AR and MA
- Often p,q <= 1 or 2

ARIMA

- AR-Integrated-MA
- ARIMA(p,d,q)
- d=order of differencing before applying

  ARMA(p,q)
- For nonstationary data w/stochastic trend

SARIMA, ARMAX

- Seasonal ARIMA(p,d,q) – and – $(P,D,Q)_S$
- Often S=
  - 12 (monthly) or
  - 4 (quarterly) or
  - 52 (weekly)
- Or, S=7 for daily data inside a week

- ARMAX=ARMA with outside explanatory variables (halfway to multivariate time series)

State Space Model, Kalman Filter

- Underlying process that we don't see
- We get noisy observations of it
- Like a Hidden Markov Model (HMM), but state is continuous rather than discrete.
- AR/MA, etc. can be written in this form too.
- State evolution (vector): $\underline{S}_t = \mathbf{F} * \underline{S}_{t-1} + \underline{\eta}_t$
- Observations (scalar): $Y_t = \mathbf{H} * \underline{S}_t + \varepsilon_t$

ARCH, GARCH(p,q)

- (Generalized) AutoRegressive Conditional Heteroskedasticity
- Variance changes randomly in time according to ARMA process.
- Used for many financial models

Volatility

- Volatility – conditional variance of the process
  - Don’t observe this quantity directly (only one observation at each time point)
- Common features
  - Serially uncorrelated but a depended process
  - Stationary
  - Clusters of low and high volatility
  - Tends to evolve over time with jumps being rare
  - Asymmetric as a function of market increases or market decreases

The basic models

- Consider a process r(t) where

$$r(t) = \mu(t) + a(t)$$

$$\mu(t) = E(r(t) \mid F(t-1))$$

Conditional mean evolves as an ARMA process

$$\mu(t) = \phi_0 + \sum_{j=1}^{p} \phi_j r(t-j) + \sum_{k=1}^{q} \theta_k a(t-k)$$

$$\sigma^2(t) = Var(r(t) \mid F(t-1))$$

How does the conditional variance evolve?

Modeling the volatility

- Evolution of the conditional variance follows to basic sets of models
  - The evolution is set by a fixed equation (ARCH, GARCH,…)
  - The evolution is driven by a stochastic equation (stochastic volatility models).
- Notation:
  - a(t)=shock or mean-corrected return;
  - $\sigma(t)$ is the positive square root of the volatility

ARCH model

- We have the general format as before
- The equation defining the evolution of the volatility (conditional variance) is an AR(m) process.

$$a(t) = \sigma(t)\varepsilon(t)$$

Why would this model yield "volatility clustering"?

$$\sigma^2(t) = \alpha_0 + \alpha_1 a^2(t-1) + \cdots + \alpha_m a^2(t-m)$$

Basic properties ARCH(1)

Unconditional mean is 0.

$$
\begin{aligned}
E[a(t)] &= E[E(a(t) \mid F(t-1))] \\
&= E[E(\sigma(t)\varepsilon(t) \mid F(t-1))] \\
&= E[\sigma(t)E(\varepsilon(t)))] \\
&= 0
\end{aligned}
$$

Basic properties, ARCH(1)

Unconditional variance

$$Var[a(t)] = Var[E(a(t) \mid F(t-1))] + E[Var(a(t) \mid F(t-1)]$$

$$= 0 + E[\sigma^2(t)]$$

$$= E[\alpha_0 + \alpha_1 a^2(t-1)]$$

$$= \alpha_0 + \alpha_1 E[a^2(t-1)]$$

$$= \alpha_0 + \alpha_1 Var[a(t-1)]$$

$$= \alpha_0 + \alpha_1 Var[a(t)]$$

$$Var[a(t)] = \alpha_0 / (1 - \alpha_1)$$

What constraint does this put on α1?

Basic properties of ARCH

- $0 \leq \alpha_1 < 1$
- Higher order moments lead to additional constraints on the parameters
  - Finite positive (always the case) fourth moments requires

    $0 \leq \alpha_1^2 < 1/3$
- Moment conditions get more difficult as the order increases – see Enders
- Note that in general the kurtosis for a(t) is greater than 3 even if the ARCH model is built from normal random variates.
- Thus the tails are heavier and you expect more “outliers” than “normal”.

ARCH Estimation, Model Fitting and Forecasting

- MLE for normal and t-dist ε's is discussed in Enders
- The full likelihood is very difficult and thus the conditional likelihood is most generally used.
- The conditional likelihood ignores the component of the likelihood that involves unobserved values (in other words, obs 1 through m)
- MLE for joint estimation of parameters and degree of the t-distribution is given.
- Model selection
  - Fit ARMA model to mean structure
  - Review PACF to identify order of ARCH
  - Check the standardized residuals – should be WN
- Forecasting – identical to AR forecasting but we forecast volatility first and then forecast the process.

GARCH model

- Generalize the ARCH model by including an MA component in the model for the volatility or the conditional variance.

$$a(t) = \sigma(t)\varepsilon(t)$$

$$\sigma^2(t) = \alpha_0 + \sum_{j=1}^{m} \alpha_j a^2(t-j) + \sum_{k=1}^{s} \beta_k \sigma^2(t-k)$$

Proceed as before – using all you learned from ARMA models.

Exponential Smoothing = EWMA

- More a method than a model.
- Very common in practice
- Forecasting w/o much modeling of the process.
- A_t = forecast of series at time t
- Pick some parameter α between 0 and 1
- $A_t = \alpha Y_t + (1-\alpha)A_{t-1}$
  - or $A_t = A_{t-1} + \alpha *$(error in period t)
- Why call it “Exponential”?
  - Weight on Y_t at lag k is $(1-\alpha)^k$

How to determine the parameter

- Train the model: try various values of α
- Pick the one that gives the lowest sum of absolute forecast errors
- The larger α is, the more weight given to recent observations
- Common values are 0.10, 0.30, 0.50
- If best α is over 0.50, there's probably some trend or seasonality present

Holt-Winters

- Exponential smoothing: no trend or seasonality
  - Excel/Analysis Toolpak can do it if you tell it α
- Holt's method: accounts for trend.
  - Also known as double-exponential smoothing
- Holt-Winters: accounts for trend & seasons
  - Also known as triple-exponential smoothing

Multivariate

- Along with ACF, use Cross-Correlation
- Cross-Correl is not 1 at lag=0
- Cross-Correl is not symmetric around lag=0
- Leading Indicator: one series' behavior helps predict another after a little lag
  - Leading means “coming before”, not “better than others”
- Can also do cross-spectrum, aka coherence

Cycles/Seasonality

- Suppose a yearly cycle
- Sample quarterly: 3-med, 6-hi, 9-med, 12-low
- Sample every 6 months: 3-med, 9-med
  - Or 6-hi, 12-low
- To see a cycle, must sample at twice its freq.

The basic problem

- We have data, want to find
  - Cycle length (e.g. Business cycles), or
  - Strength of seasonal components
- Idea: use sine waves as explanatory variables
- If a sine wave at a certain frequency explains things well, then there's a lot of strength.
  - Could be our cycle's frequency
  - Or strength of known seasonal component
- Explains=correlates

Correlate with Sine Waves

- Ordinary covar: $$\sum_{t=0}^{T-1}(X_t - \overline{X})(Y_t - \overline{Y})$$

- At freq. Omega, $$\sum_{t=0}^{T-1}\sin(\omega t)Y_t$$

  (means are zero)

- Problem: what if that sine is out of phase with our cycle?

Solution

- Also correlate with a cosine
  - 90 degrees out of phase with sine
- Why not also with a 180-out-of-phase?
  - Because if that had a strong correl, our original sine would have a strong correl of opposite sign.
- Sines & Cosines, —combine using complex variables!

The Discrete Fourier Transform

$$d(\omega) = \sum_{t=0}^{T-1} e^{-i\omega t} Y_t$$

- Often a scaling factor like 1/T, 1/sqrt(T), 1/2pi, etc. out front.
- Some people use +i instead of -i
- Often look only at the frequencies $\omega_k = 2\pi k / T$
- k=0,...,T-1

$$d(\omega_k) = \sum_{t=0}^{T-1} e^{-2\pi i k / T} Y_t$$

- Define a matrix F whose j,k entry is

  exp(-i*j*k*2pi/T)

- Then $\vec{d} = \mathbf{F}\vec{Y}$
- Matrix multiplication takes T^2 operations
- This matrix has a special structure, can do it in about T log T operations
- That's the FFT=Fast Fourier Transform
- Easiest if T is a power of 2

So now we have complex values...

- Take magnitude & argument of each DFT result
- Plot squared magnitude vs. frequency
  - This is the “Periodogram”
- Large value = that frequency is very strong
- Often plotted on semilog-y scale, “decibels”

Interpretations

- Value at k=0 is mean of data series
  - Called “DC” component
- Area under periodogram is proportional to Var(data series)
  - Height at each point=how much of variance is explained by that frequency
- Plotting argument vs. frequency shows phase
- Often need to smooth with moving avg.

Long-memory time series

- Ordinary theory assumes that ACF dies off faster than 1/h
- But some time series don't satisfy that:
  - River flows
  - Packet amounts on data networks
- Connected to chaos & fractals

References

- Walter Enders (2003), Applied Econometric Time Series, Wiley
- Shumway R. H., Stoffer (2011), *Time Series Analysis and its Applications*, Springer.
- J. Campbell, A. Lo and C. Mackinlay (1997), The Econometrics of Financial Markets, Princeton University Press
- Hamilton, James (1994), Time Series Analysis, Princeton University Press.