Statistics 5350/7110
Forecasting

Lecture 25
GARCH Models

Professor Robert Stine

Preliminaries

- Questions?

- Final exam
  - Similar to Midterm: short answer, multiple choice
  - Friday, December 13
  - JMHH F45
  - 3-5 pm

- Quick review
  - Martingales and random walks
  - Evolution of forecasts: center and standard error
  - Key role of the MA representation of an ARMA process

Today's Topics

Textbook §8.1

- Financial returns
  - Daily returns on S&P 500
  - Index isn't stationary in the mean
  - Returns aren't stationary in the variance

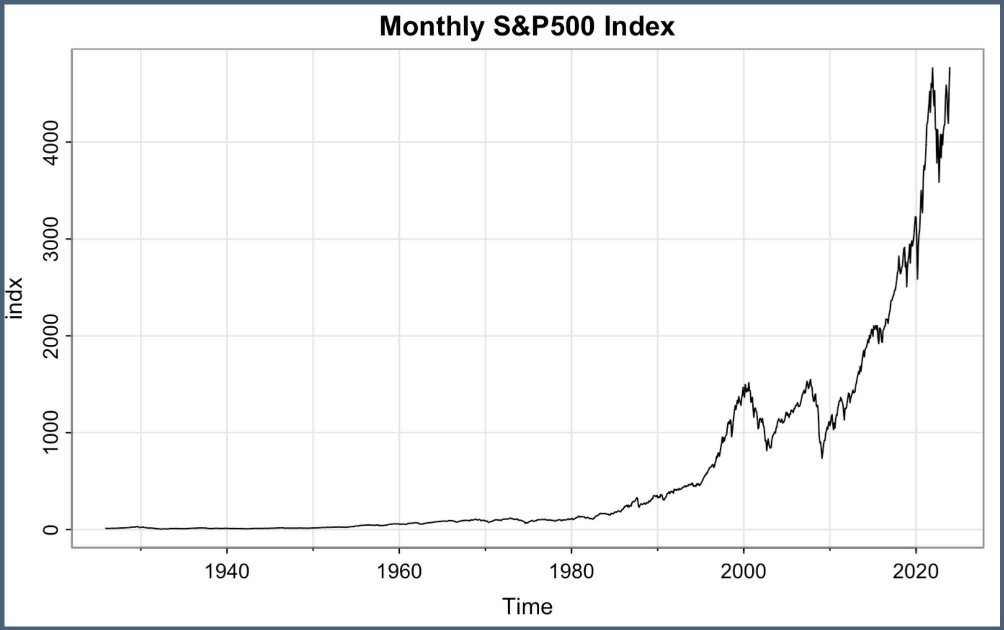
- ARCH models
  - Autoregressive conditional heteroscedastic models
  - One-step ahead variance
    - – ARMA models: Constant one-step ahead forecast variance
    - – ARCH models: Allow variance to change

- Estimation
  - Weighted least squares
  - Specialized software: Maximum likelihood estimates

Stock Market Returns

S&P 500

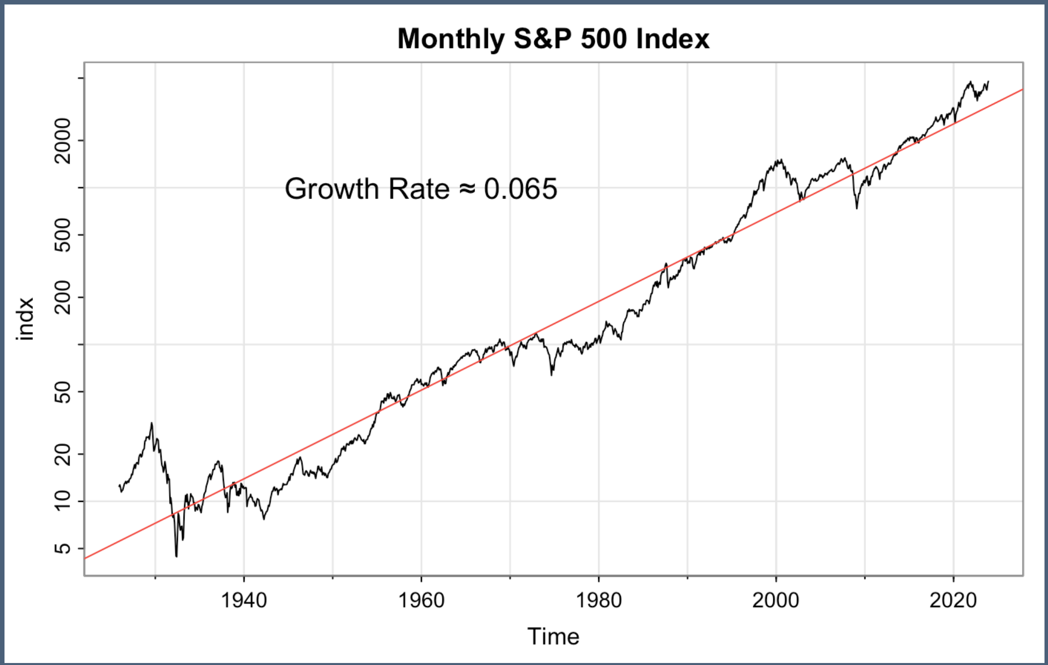
- Portfolio
  - Value-weighted portfolio of stocks from 500 companies with largest capitalized value
  - Alternative Dow Jones is not value weighted
- Index history
  - Monthly value, 1926-2023



Where's the Great Depression, WWII?

S&P 500

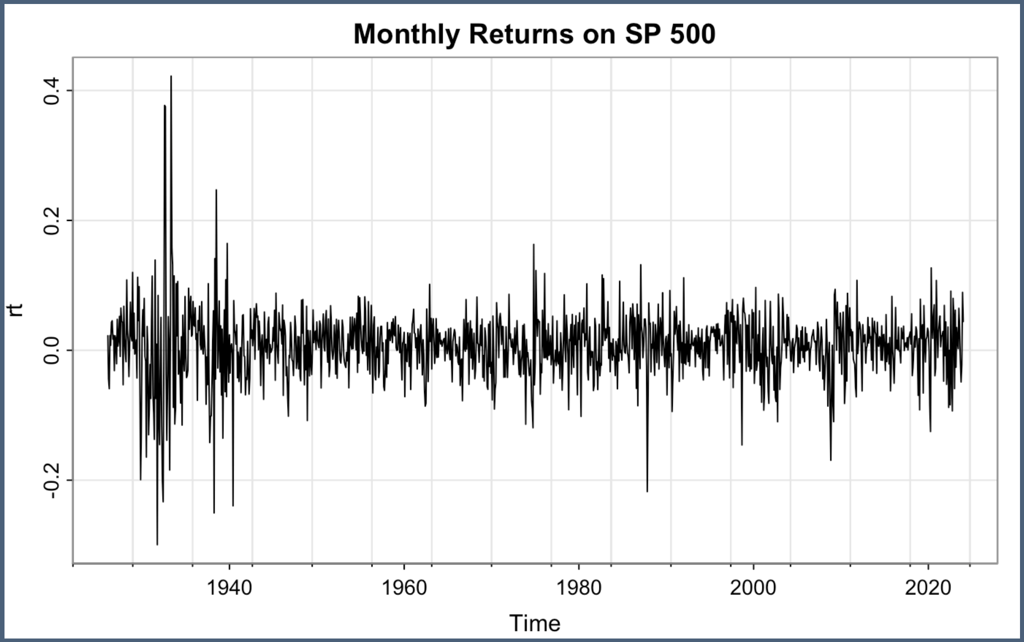
- Portfolio
  - Value-weighted portfolio of stocks from 500 companies with largest capitalized value
  - Alternative Dow Jones is not value weighted
- Index history
  - Monthly value, 1926-2023… Log scale



Where's the Great Depression, WWII?

S&P 500

- Portfolio
  - Value-weighted portfolio of stocks from 500 companies with largest capitalized value
  - Alternative Dow Jones is not value weighted
- Index history
  - Monthly value, 1926-2023… Returns show volatility

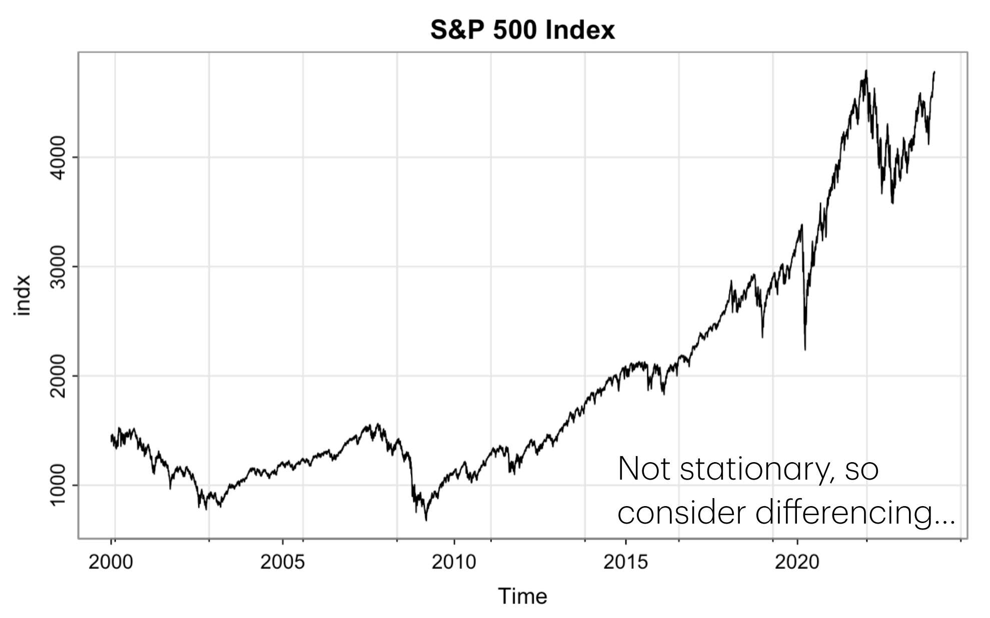


No problem finding the Great Depression in the returns.

Recent S&P 500

- Portfolio
  - Value-weighted portfolio of stocks from 500 companies with largest capitalized value
  - Alternative Dow Jones is not value weighted
- Index history
  - Daily, 2000 - 2023

No.	Symbol	Company Name	Market Cap
1	AAPL	Apple Inc.	3,521.61B
2	NVDA	NVIDIA Corporation	3,331.62B
3	MSFT	Microsoft Corporation	3,115.22B
4	AMZN	Amazon.com, Inc.	2,120.29B
5	GOOGL	Alphabet Inc.	2,051.84B
6	GOOG	Alphabet Inc.	2,026.84B
7	META	Meta Platforms, Inc.	1,425.40B
8	TSLA	Tesla, Inc.	1,090.49B
9	BRK.B	Berkshire Hathaway Inc.	1,027.36B
10	AVGO	Broadcom Inc.	768.45B



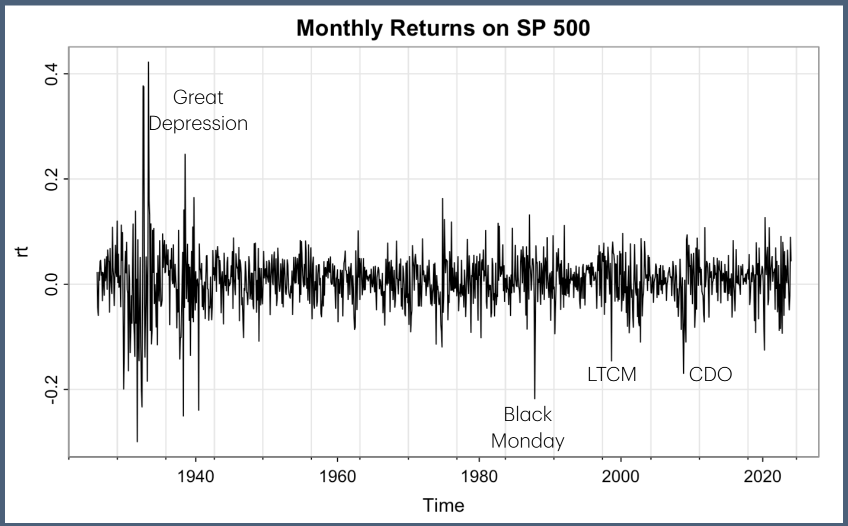
S&P 500 Daily Returns

- Returns
  - "Little r" returns

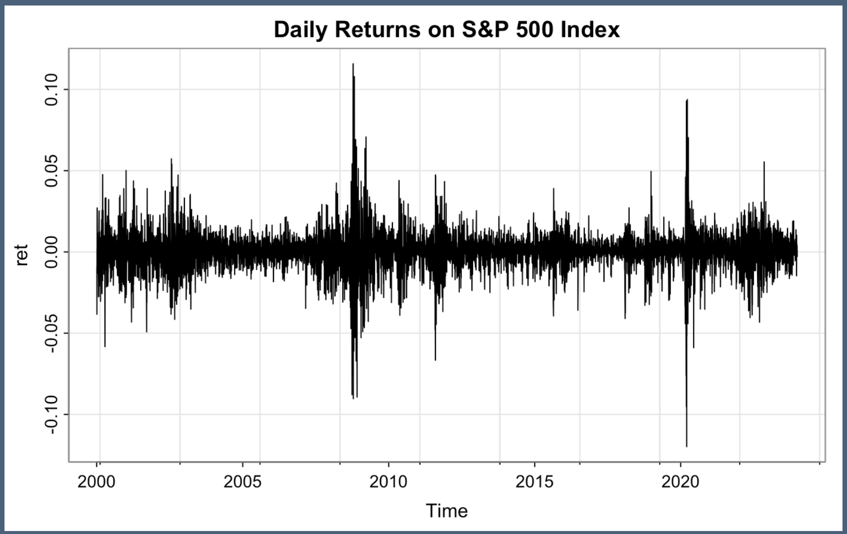
$$r_t = \frac{S_t - S_{t-1}}{S_{t-1}}$$

  - Returns on S&P 500 highly correlated with returns on the entire market
  - Appears stationary in mean, but has sustained periods of high volatility

We actually study the formal returns which are adjusted for dividends and related financial activity.



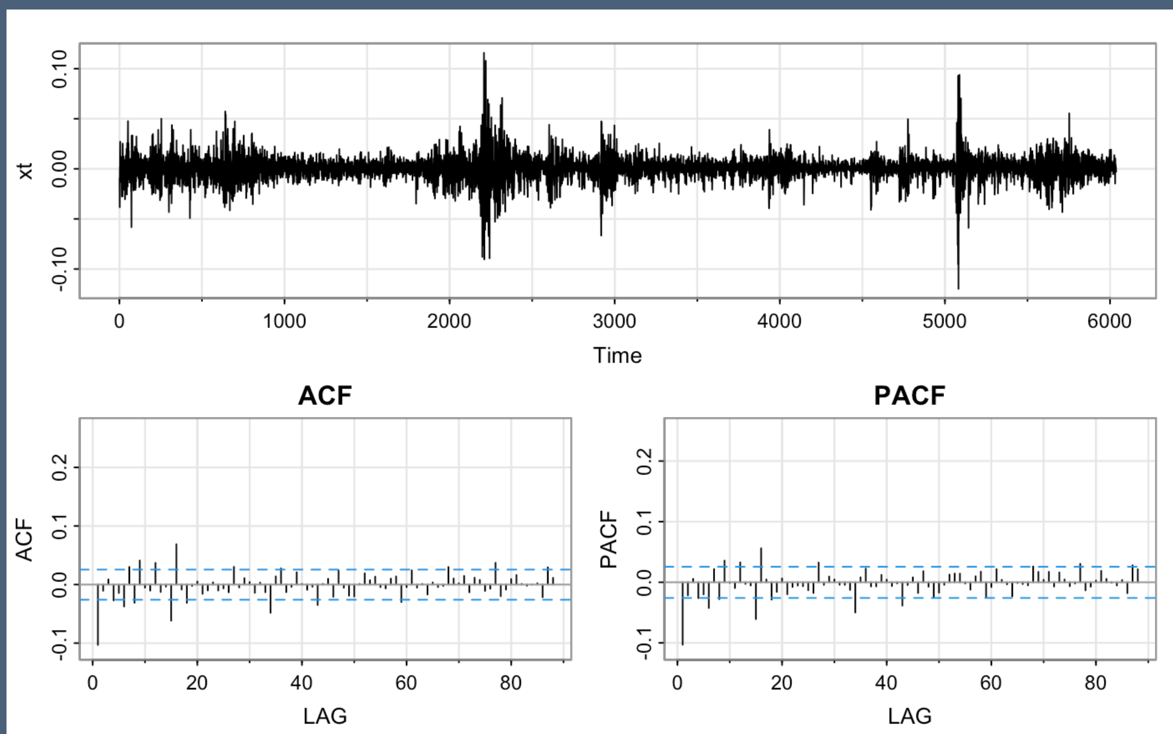
n=1177

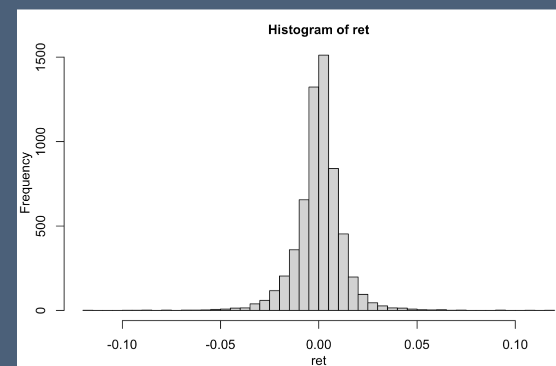


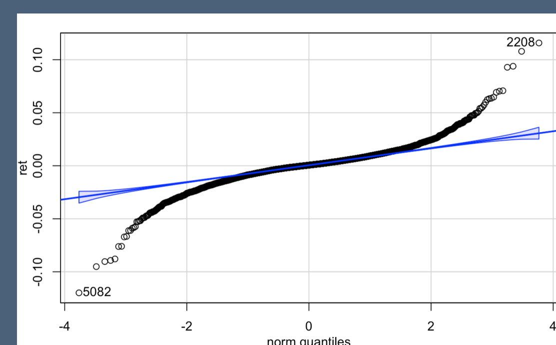
n=6037

S&P 500 Daily Returns

- Basic statistics
  - ACF and PACF appear to "cut off". Is that possible?
  - Distribution has heavy tails, more outliers than a normal distribution

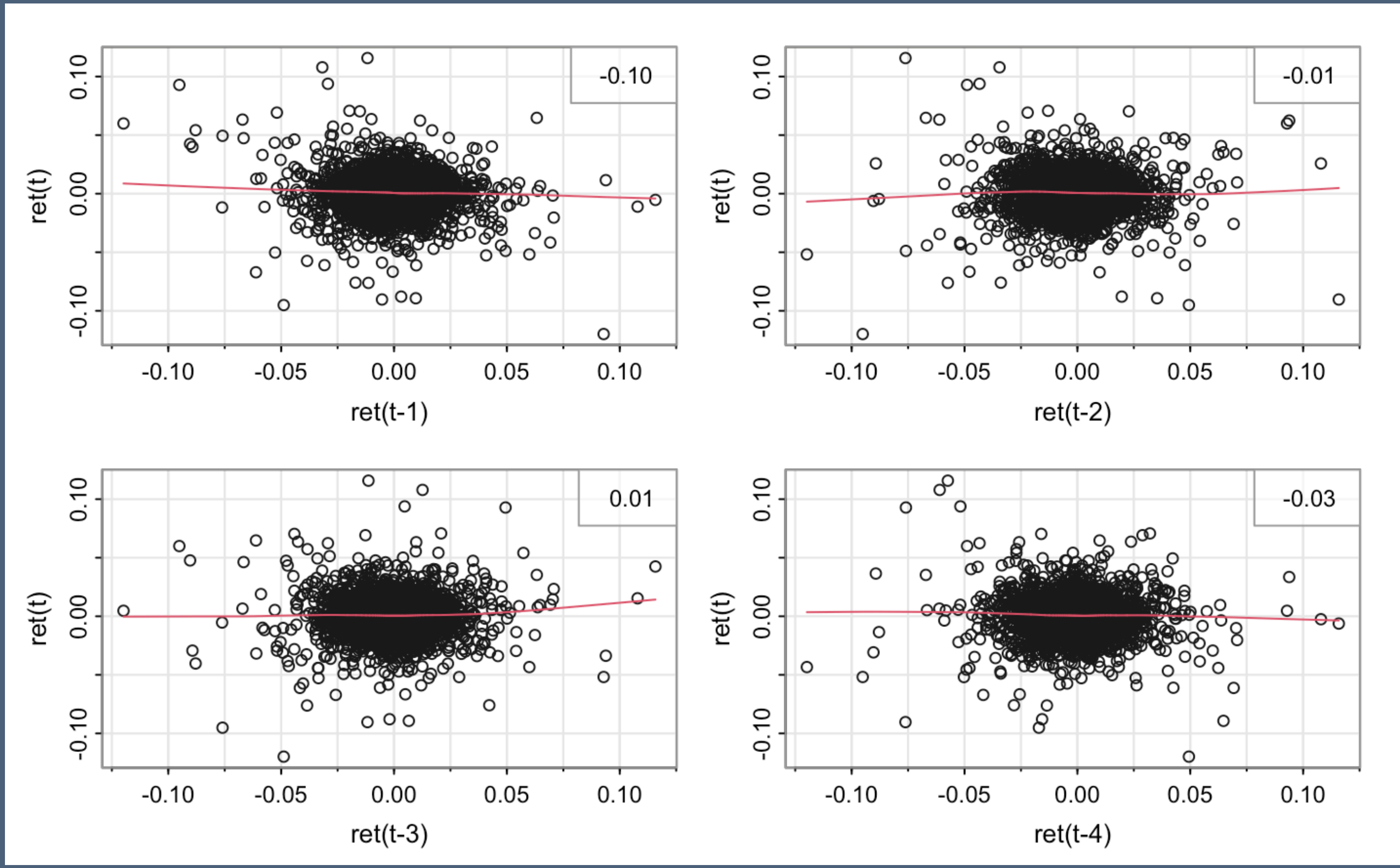






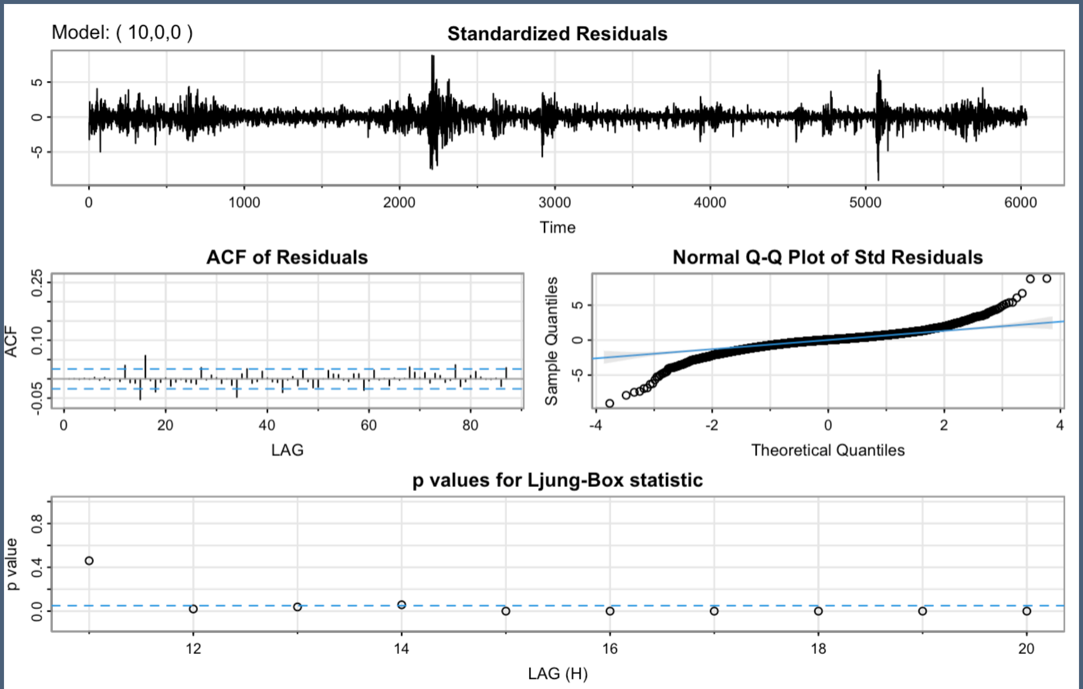
S&P 500 Daily Returns

- Lag plots
  - Influential outliers?
  - 0.10 is a small correlation



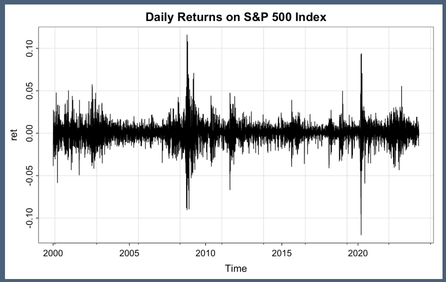
Autoregressive Model

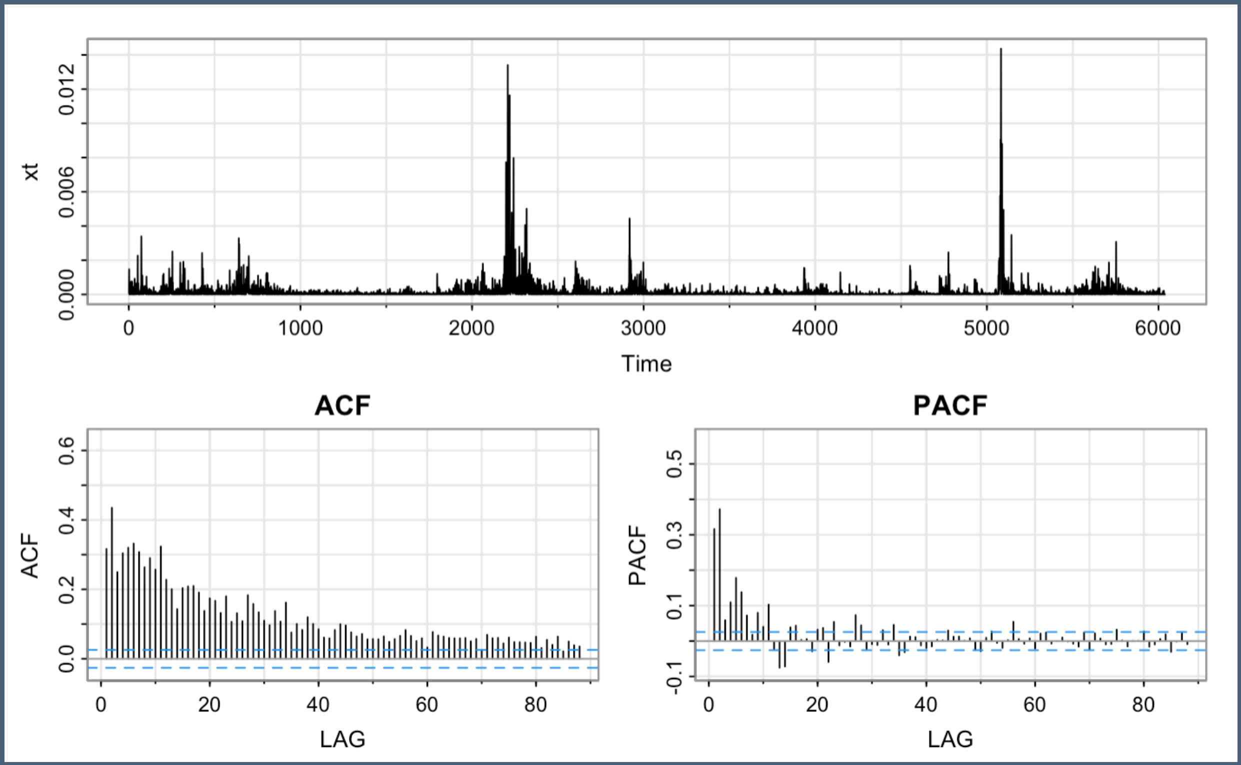
- Fit autoregression to returns
  - AR(10) model
  - Several significant coefficients
  - Explains only 1.7% of variation, and clearly not capturing volatile periods



Volatility

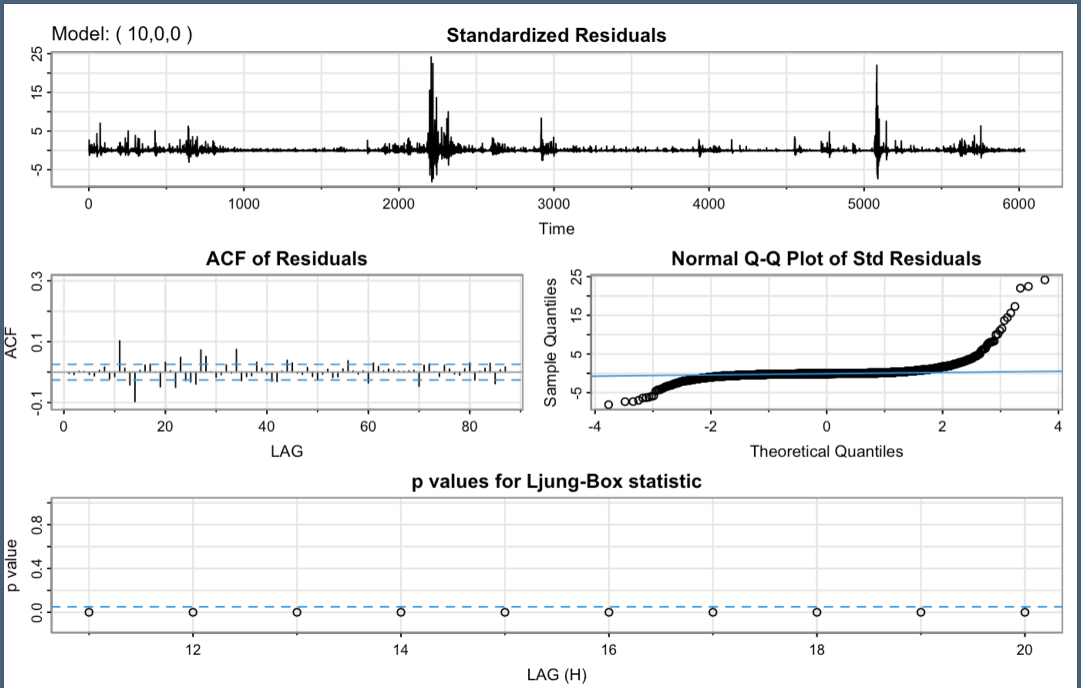
- Dependence in the squares of the returns
  - Unlike returns, substantial autocorrelation
  - Concentrated in ≈ first 10 lags





Autoregressive Model for Volatility

- Fit autoregression
  - AR(10) model for squared returns
  - Many very significant coefficients (p-value < 0.0001)
  - Explains 29% of variation among squared returns, but does not capture volatile periods



Need something better for modeling this type of time series

ARCH Model

ARCH Model

- Autoregressive, conditional heteroscedasticity
  - Two-equation model: one equation for observation and one for hidden state
  - Proposed by Engle (1982) in study of UK inflation
- ARCH(1)
  - Gaussian noise process (independent of other terms) $\epsilon_t \sim N(0,1)$
  - Observed time series $r_t = \sigma_t \epsilon_t$ Equation 8.2
  - "Hidden" variation process (not directly observed) $\sigma_t^2 = \alpha_0 + \alpha_1 r_{t-1}^2$ Equation 8.3
  - Combine and arrange equations

$$r_t^2 = \alpha_0 + \alpha_1 r_{t-1}^2 + \underbrace{\sigma_t^2(\epsilon_t^2 - 1)}_{v_t}$$ Equation 8.4

  where v_t is white noise.
- Estimation
  - Maximum likelihood is essentially weighted LS, noting that $r_t \mid r_{t-1} \sim N\left(0,\, \alpha_0 + \alpha_1 r_{t-1}^2\right)$

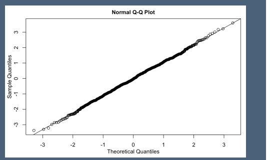
Fat tails?

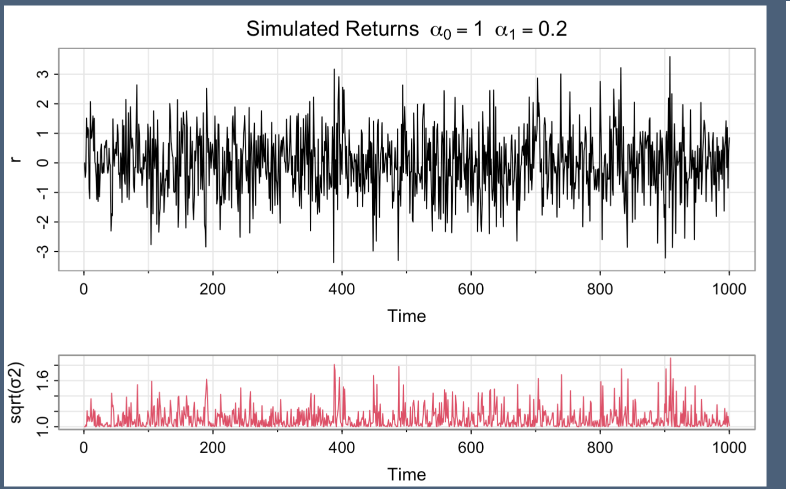
ARCH(1) Example

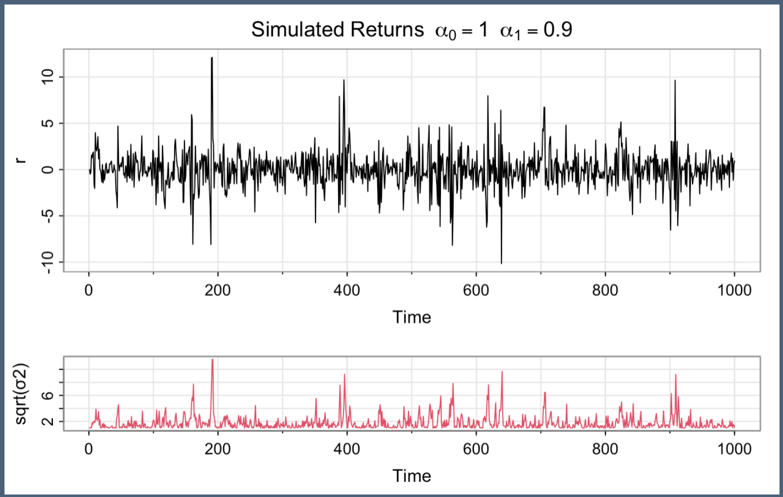
- Simulate process
  - Simulate Gaussian noise process (independent of other terms) $\epsilon_t \sim N(0,1)$
  - Recursively generate the observed time series

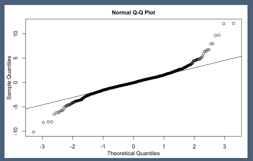
$$\sigma_t^2 = \alpha_0 + \alpha_1 r_{t-1}^2 \quad \text{followed by} \quad r_t = \sigma_t \epsilon_t$$

- Question
  - Resemble returns on the S&P 500?









ARCH(1) Estimates

- Estimation
  - Can we recover the coefficients used in the simulation?
- Weighted least squares
  - Supply weights to the `dylm` function
  - Regress r_t^2 on it's lags; use 1/fit as weights

`weights`	an optional vector of weights to be used in the fitting process. If specified, weighted least squares is used with weights `weights` (that is, minimizing `sum(w*e^2)`); otherwise ordinary least squares is used.

WLS iterations

```
         α0     α1
Fit 0 1.8798 0.5423
Fit 1 1.2496 0.6958
Fit 2 1.2001 0.7079
Fit 3 1.1962 0.7088
True  1.0000 0.9000
```

```
Coefficients:
            Estimate Std. Error
(Intercept)  1.19616    0.13497
L(sqr, 1)    0.70883    0.04312
```

WLS underestimates coefficient in the variance equation

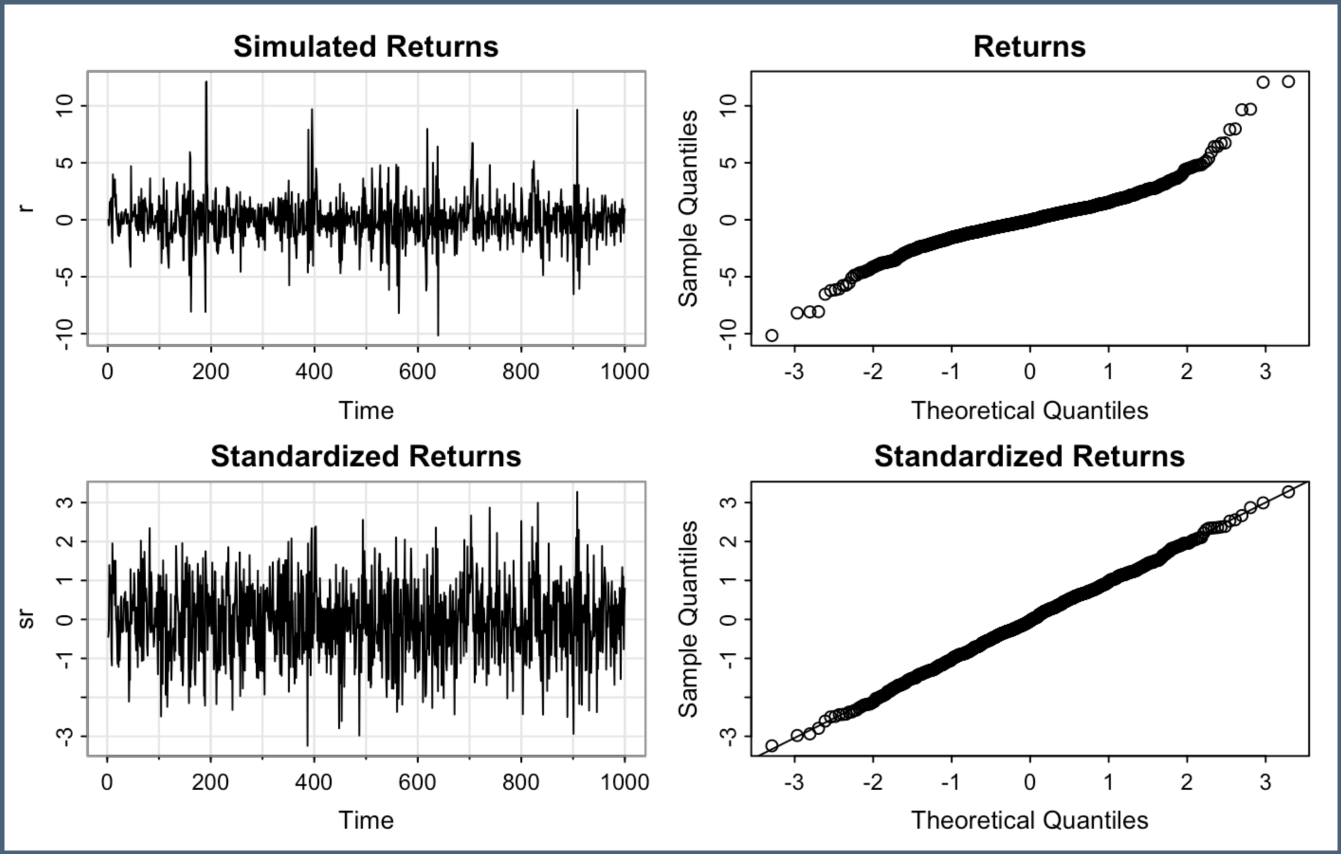
```
Error Analysis:
        Estimate  Std. Error
mu      -0.01808     0.03561
omega    1.05103     0.08680
alpha1   0.82944     0.07834
```

fGarch package does a better job for this example

fGarch includes many supplemental diagnostics as well.

Standardized Results

- Compare before/after



Using fit from WLS

ARCH(1) Estimates: S&P 500

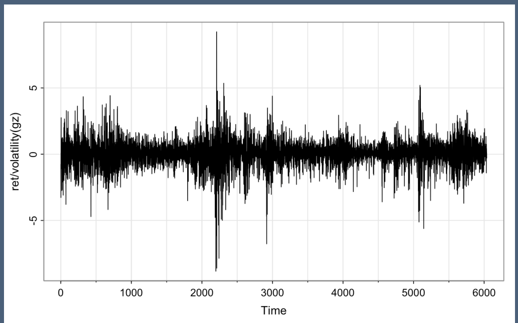
- Fit model to daily returns
  - Highly significant coefficients
  - Diagnostics indicate problems

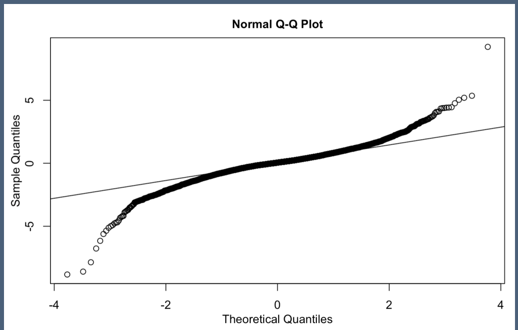
Error Analysis:

	Estimate	Std. Error	t value
mu	5.856e-04	1.326e-04	4.416
omega	9.644e-05	2.452e-06	39.331
alpha1	3.980e-01	2.817e-02	14.130

Standardised Residuals Tests:

			Statistic	p-Value
Jarque-Bera Test	R	Chi^2	11917.86564	0.000000000
Shapiro-Wilk Test	R	W	NA	NA
Ljung-Box Test	R	Q(10)	19.25891	0.037094782
Ljung-Box Test	R	Q(15)	35.37306	0.002175275
Ljung-Box Test	R	Q(20)	50.42186	0.000192673
Ljung-Box Test	R^2	Q(10)	1714.96031	0.000000000
Ljung-Box Test	R^2	Q(15)	2193.77782	0.000000000
Ljung-Box Test	R^2	Q(20)	2560.64958	0.000000000





- How to improve this model?
  - Better mean function?
  - Different volatility model?
  - Normality assumption?

ARCH(q) Model

Extensions of ARCH Model

- ARCH(q)
  - Expand expression for conditional variance to have more lags

$$\sigma_t^2 = \alpha_0 + \alpha_1 r_{t-1}^2 + \cdots + \alpha_q r_{t-q}^2$$

Equation 8.7

  - Book labels these ARCH(p). Will see explanation for using other symbol in Lecture 26

- Mean term
  - Allows returns to have a non-zero mean trend

$$r_t = \mu_t + \sigma_t \epsilon_t$$

Equation 8.8

where the mean might be, for example, a regression

$$\mu_t = \beta_0 + \beta_1 X_{t,1} + \beta_2 X_{t,2}$$

or perhaps an ARMA model

Constraints

- Assume ARCH(2)
  - Mean of return is 0 (assuming $\mu_t = 0$) $\quad E\, r_t = E(\sigma_t\, \epsilon_t) = 0$
  - Conditional variance

$$\text{Var}(r_t \,|\, r_{t-1}, r_{t-2}) = \sigma_t^2$$

Remember that
$\sigma_t^2 \in \mathcal{F}_{t-1}$

  - Marginal variance

$$\text{Var}(r_t) = E\, r_t^2 - (E\, r_t)^2 = E\, r_t^2$$

which is (iterated expectations)

$$E\, r_t^2 = E\left(E\,(r_t^2 \,|\, r_{t-1}, r_{t-2})\right) = E\left(\alpha_0 + \alpha_1 r_{t-1}^2 + \alpha_2 r_{t-2}^2\right) = \alpha_0 + \alpha_1 E(r_{t-1}^2) + \alpha_2 E(r_{t-2}^2)$$

  - Assuming variance stationary, then

$$\text{Var}(r_t) = \alpha_0/(1 - \alpha_1 - \alpha_2)$$

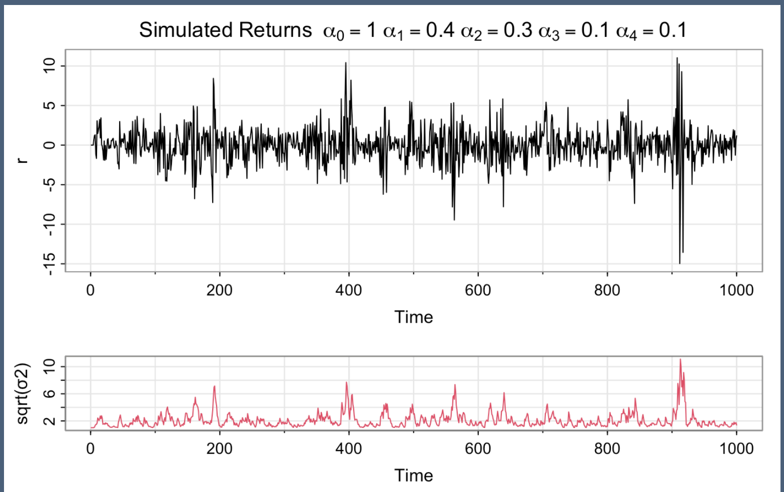
- Constraints
  - Guarantee variance is not negative $\quad 0 \leq \alpha_j$
  - Sum $\quad \alpha_1 + \alpha_2 < 1$

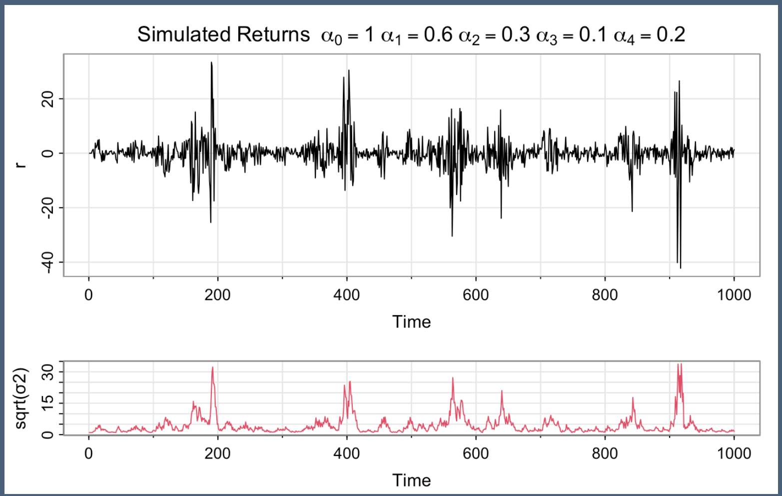
ARCH(4) Simulated Examples

- Simulated data
  - Simulate Gaussian noise process (independent of other terms) $\epsilon_t \sim N(0,1)$
  - Recursively generate the observed time series

$$\sigma_t^2 = \alpha_0 + \alpha_1 r_{t-1}^2 + \alpha_2 r_{t-2}^2 + \alpha_3 r_{t-3}^2 + \alpha_4 r_{t-4}^2 \quad \text{followed by} \quad r_t = \sigma_t \epsilon_t$$

- Question
  - Resemble returns on the S&P 500?



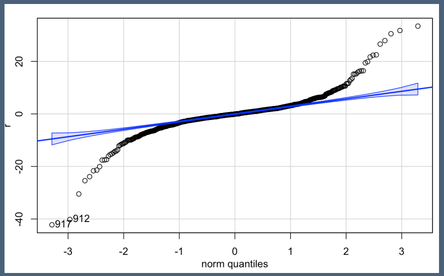


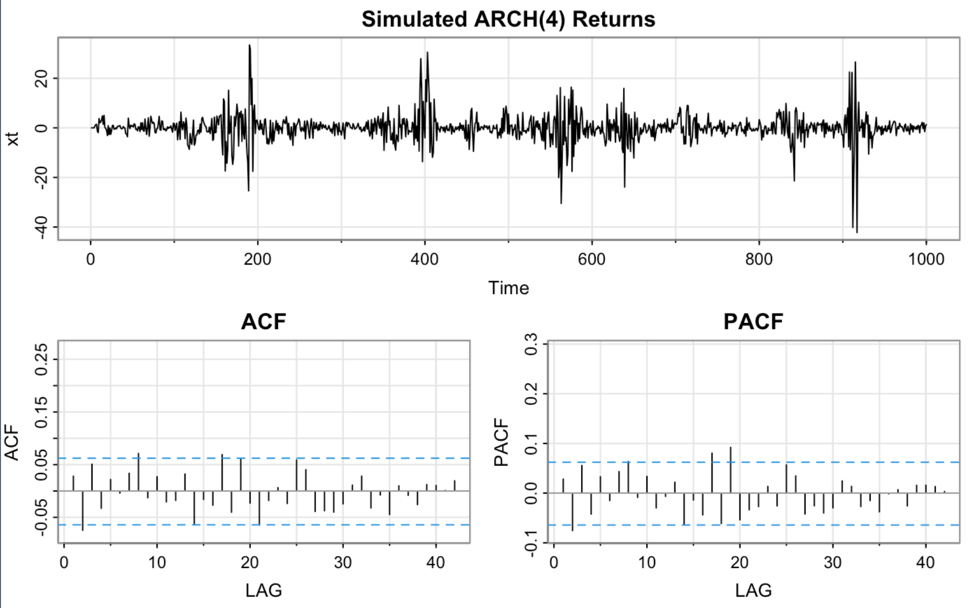
Unconstrained, but more like returns.

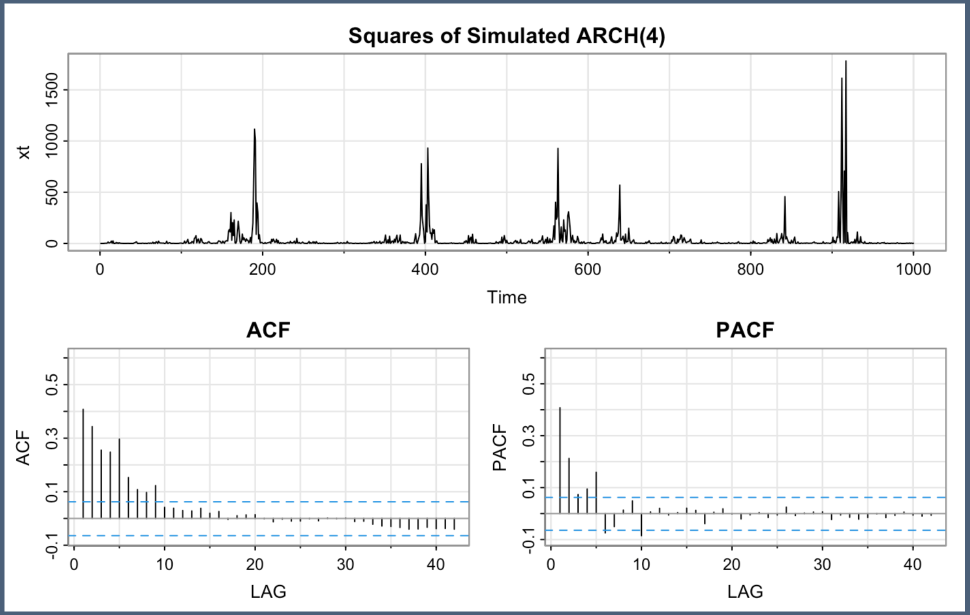
Appearance changes depending on the random seed.

ARCH(4) Example

- Autocorrelations of simulated returns
  - Last example of ARCH(4)
  - Returns are very fat tailed (variance mixture of normals)
  - Returns lack much dependence, whereas squares are autocorrelated







ARCH(4) Simulated Example

- Estimation
  - Can we recover the coefficients used in the simulation?
- Weighted least squares
  - Supply weights to the `dylm` function
  - Regress r_t^2 on it's lags; use 1/fit as weights

`weights`	an optional vector of weights to be used in the fitting process. If specified, weighted least squares is used with weights `weights` (that is, minimizing `sum(w*e^2)`); otherwise ordinary least squares is used.

	α_0	α_1	α_2	α_3	α_4
True	10.2749	0.3097	0.0704	0.0111	0.2385
Fit 0	3.5547	0.4675	0.1880	0.0183	0.1982
Fit 1	2.4927	0.4734	0.2114	0.0385	0.1871
Fit 2	2.2747	0.4759	0.2130	0.0429	0.1865
Fit 3	2.2298	0.4764	0.2132	0.0438	0.1865
Fit 4	1.0000	0.6000	0.3000	0.1000	0.2000

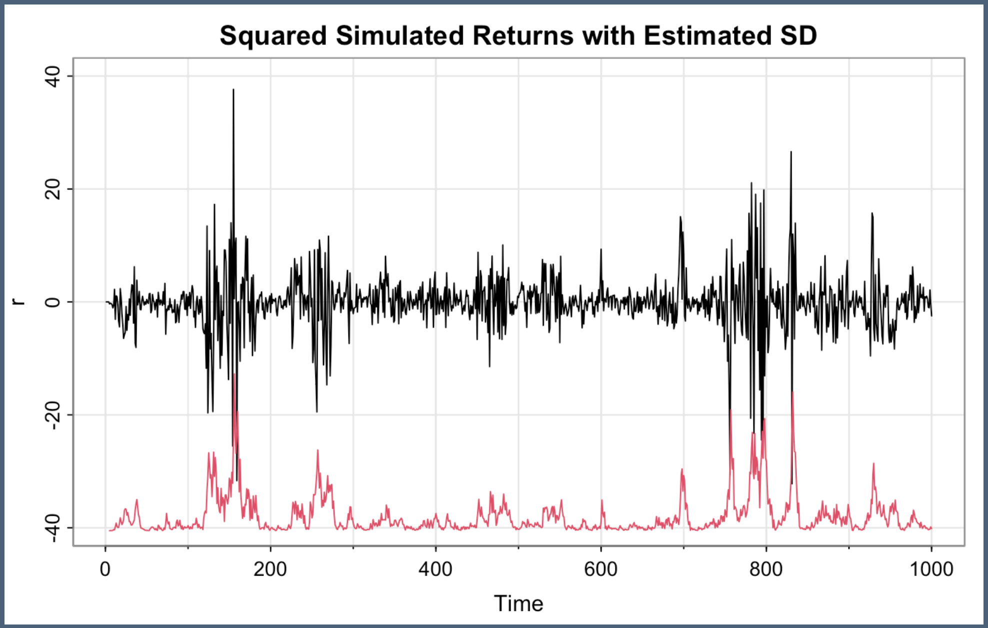
Coefficients:

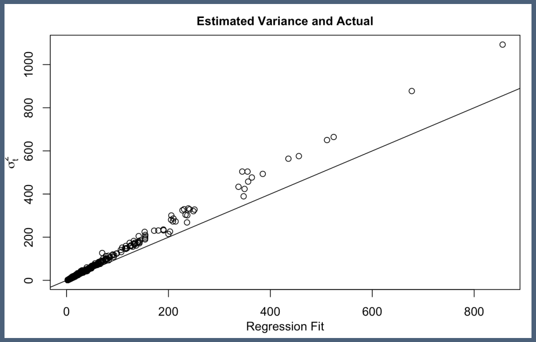
	Estimate	Std. Error	t value
(Intercept)	2.22983	0.63843	3.493
L(sqr, 1:4)1	0.47639	0.04290	11.104
L(sqr, 1:4)2	0.21319	0.03690	5.777
L(sqr, 1:4)3	0.04380	0.02971	1.474
L(sqr, 1:4)4	0.18648	0.03344	5.577

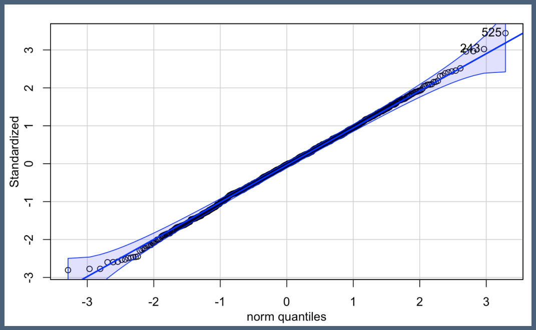
R-code: garchFit again does better!

ARCH(4) Example

- Fitted values from regression
  - Estimates of conditional variance σ_t^2
  - Much more visually similar to bursts of volatility in the SP500 returns
  - WLS underestimates large variances but standardized residuals look normal (qqplot)







ARCH(4) for S&P 500 Returns

- Fit ARCH(4) to S&P 500 returns
  - Clearly more dependence than 4 lags… significant at 10
  - Lots of AR terms suggests what?

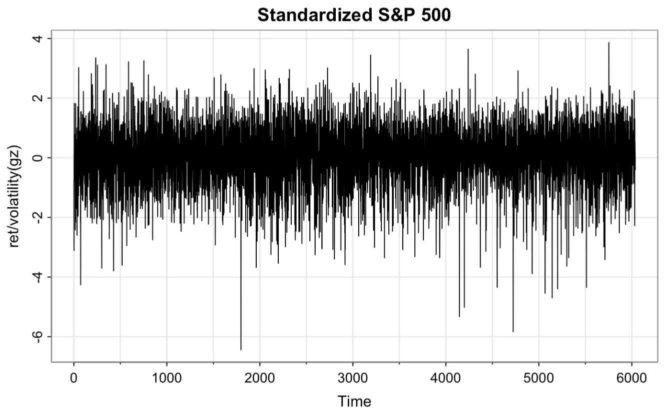
Error Analysis:

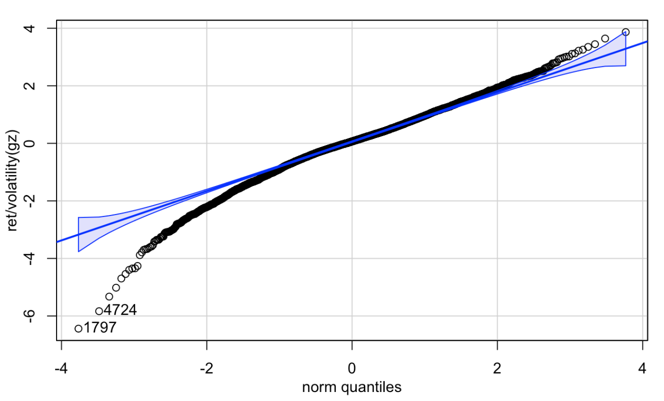
	Estimate	Std. Error	t value
mu	6.442e-04	1.017e-04	6.335
omega	2.002e-05	1.420e-06	14.098
alpha1	8.219e-02	1.411e-02	5.825
alpha2	1.561e-01	1.752e-02	8.910
alpha3	1.142e-01	1.677e-02	6.809
alpha4	1.246e-01	1.728e-02	7.212
alpha5	7.822e-02	1.488e-02	5.259
alpha6	6.570e-02	1.516e-02	4.333
alpha7	5.713e-02	1.394e-02	4.099
alpha8	8.948e-02	1.601e-02	5.588
alpha9	5.787e-02	1.479e-02	3.913
alpha10	5.326e-02	1.402e-02	3.798

Error Analysis:

	Estimate	Std. Error	t value
mu	6.442e-04	1.027e-04	6.274
omega	4.086e-06	5.798e-07	7.046
alpha1	8.626e-02	1.387e-02	6.220
alpha2	1.259e-01	1.895e-02	6.647
beta1	2.138e-01	1.808e-01	1.182
beta2	5.448e-01	1.618e-01	3.368

Much better, but something
else seems to be happening





What's next?

- GARCH process
  - Extension of the model
  - Analogous of AR –> ARMA
  - Other error distributions aside from normal
- Modeling examples
  - Stock returns
  - Other processes