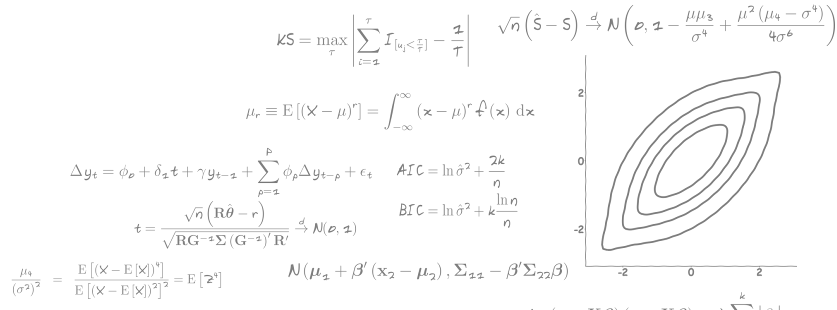
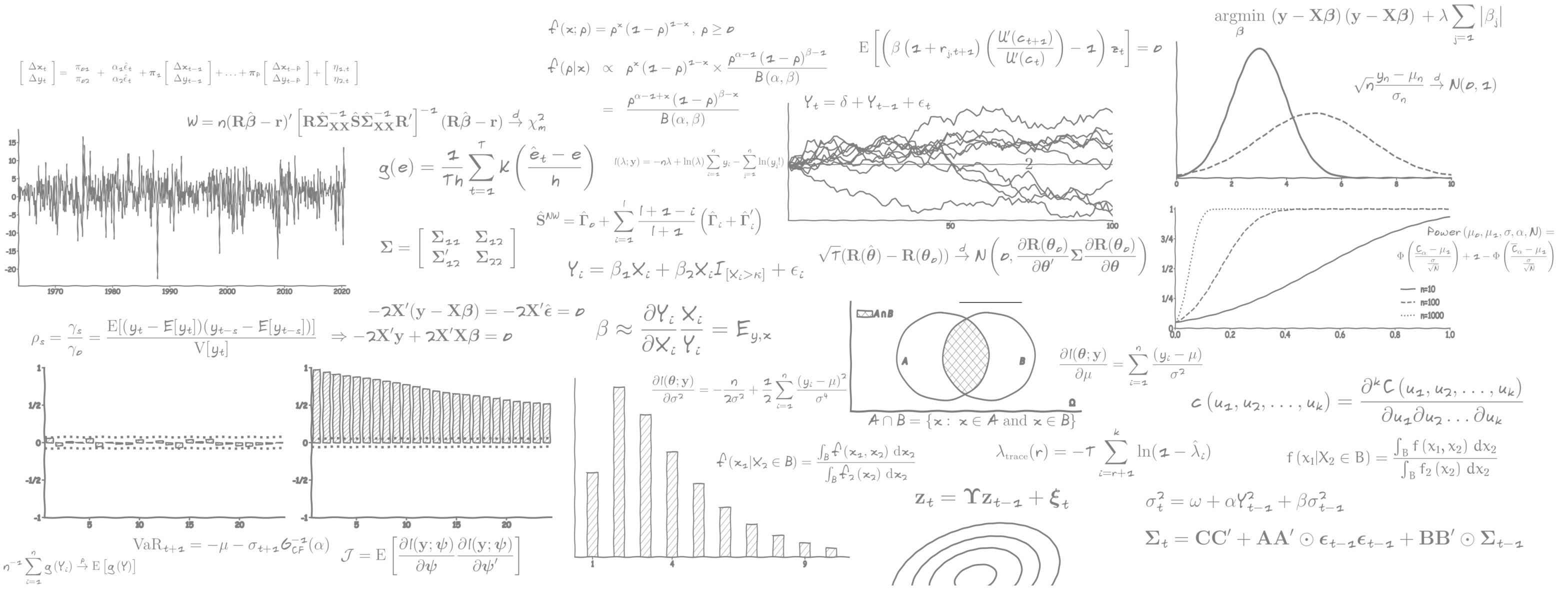
Univariate Volatility Modeling

Kevin Sheppard
https://kevinsheppard.com/teaching/mfe/
February 4, 2022





Volatility Overview

- What is volatility?
- Why does it change?
- What are ARCH, GARCH, TARCH, EGARCH, SWARCH, ZARCH, APARCH, STARCH, *etc.* models?
- What does time-varying volatility *look like*?
- What are the basic properties of ARCH and GARCH models?
- What is the news impact curve?
- How are the parameters of ARCH models estimated? What about inference?
- Twists on the standard model
- Forecasting conditional variance
- *Realized Variance*
- Implied Volatility

What is *volatility*?

- Volatility
  - Standard deviation
- Realized Volatility

$$\hat{\sigma} = \sqrt{T^{-1}\sum_{t=1}^{T}(r_t - \hat{\mu})^2}$$

  - Other meaning: variance computed from ultra-high frequency (UHF) data
- Conditional Volatility

$$\mathrm{E}_t[\sigma_{t+1}]$$

- Implied Volatility
- Annualized Volatility ($\sqrt{252}\times$daily, $\sqrt{12}\times$monthly)
  - Mean scales linearly with time ($252\times$daily, $12\times$monthly)
- Variance is squared volatility

Why does volatility change?

- Possible explanations:
  - News Announcements
  - Leverage
  - Volatility Feedback
  - Illiquidity
  - State Uncertainty
- None can explain all of the time-variation
- Most theoretical models have none

Review

Key Concepts

Leverage Effect, Liquidity, Volatility Feedback

Questions

- What factors are used to convert daily, weekly, and monthly volatility to annual?
- What factor would you use to convert daily volatility to annual if an asset traded 7 days a week?

Problems

1. If the annualized volatility of an asset is 48%, what is its daily, weekly, and monthly volatility?
2. If the daily return of an asset is .0476% and its daily volatility is 1.512%, what is the asset's annual Sharpe ratio?

A basic volatility model: the ARCH(1) model

$$
\begin{aligned}
r_t &= \epsilon_t \\
\sigma_t^2 &= \omega + \alpha_1 \epsilon_{t-1}^2 \\
\epsilon_t &= \sigma_t e_t \\
e_t &\overset{\text{i.i.d.}}{\sim} N(0,1)
\end{aligned}
$$

- **A**uto**r**egressive **C**onditional **H**eteroskedasticity
- Key model parameters
  - ω sets the long run level
  - α determines both the persistence and volatility of volatility (VoVo or VolVol)

Key Properties

- Conditional Mean: $\mathrm{E}_{t-1}[r_t] = \mathrm{E}_{t-1}[\epsilon_t] = 0$
- More on this later
  - Unconditional Mean: $\mathrm{E}[\epsilon_t] = 0$
    - Follows directly from the conditional mean and the LIE
- Conditional Variance: $\mathrm{E}_{t-1}[r_t^2] = \mathrm{E}_{t-1}[\epsilon_t^2] = \sigma_t^2$
- σ_t^2 and e_t^2 are independent
- $\mathrm{E}_{t-1}[e_t^2] = \mathrm{E}[e_t^2] = 1$
- $1 - \alpha_1 > 0$: Required for stationarity, also $\alpha_1 \geq 0$
  - $\omega > 0$ is also required for stationarity (technical, but obvious)

Unconditional Variance

- Unconditional Variance

$$\mathrm{E}[\epsilon_t^2] = \frac{\omega}{1 - \alpha_1}$$

- Unconditional relates the dynamic parameters to average variance

$$\mathrm{E}[\sigma_t^2] =$$

More properties of the ARCH(1)

- ARCH models are really Autoregressions in disguise
- Add $\epsilon_t^2 - \sigma_t^2$ to both sides

$$\begin{aligned}
\sigma_t^2 &= \omega + \alpha_1 \epsilon_{t-1}^2 \\
\sigma_t^2 + \epsilon_t^2 - \sigma_t^2 &= \omega + \alpha_1 \epsilon_{t-1}^2 + \epsilon_t^2 - \sigma_t^2 \\
\epsilon_t^2 &= \omega + \alpha_1 \epsilon_{t-1}^2 + \epsilon_t^2 - \sigma_t^2 \\
\epsilon_t^2 &= \omega + \alpha_1 \epsilon_{t-1}^2 + \nu_t
\end{aligned}$$

$$y_t = \phi_0 + \phi_1 y_{t-1} + \nu_t$$

  - AR(1) in ϵ_t^2
  - $\nu_t = \epsilon_t^2 - \sigma_t^2$ is a mean 0 white noise (WN) process
  - ν_t Captures variance *surprise* : $\epsilon_t^2 - \sigma_t^2 = \sigma_t^2(e_t^2 - 1)$

Autocovariance/Autocorrelations

- First Autocovariance

$$\mathrm{E}[(\epsilon_t^2 - \bar{\sigma}^2)(\epsilon_{t-1}^2 - \bar{\sigma}^2)] = \alpha_1 \mathrm{V}[\epsilon_t^2]$$

  - Same as in AR(1)
- j^th Autocovariance is

$$\alpha_1^j \mathrm{V}[\epsilon_t^2]$$

- j^th Autocorrelation is

$$\mathrm{Corr}(\epsilon_t^2, \epsilon_{t-j}^2) = \frac{\alpha_1^j \mathrm{V}[\epsilon_t^2]}{\mathrm{V}[\epsilon_t^2]} = \alpha_1^j$$

- Again, same as AR(1)
- ARCH(P) is AR(P)
  - Just apply results from AR models

Kurtosis

- Kurtosis effect is important
- Variance is not constant $\Rightarrow$ Volatility of Volatility > 0

$$\kappa = \frac{\mathrm{E}\left[\epsilon_t^4\right]}{\mathrm{E}\left[\epsilon_t^2\right]^2} = \qquad \geq 3$$

- Alternative: $\mathrm{E}[\sigma_t^4] = \mathrm{V}[\sigma_t^2] + \mathrm{E}[\sigma_t^2]^2$
  - Law of Iterated Expectations
- In ARCH(1):

$$\kappa = \frac{3(1-\alpha_1^2)}{(1-3\alpha_1^2)} > 3$$

- Finite if $\alpha_1 < \sqrt{\frac{1}{3}} \approx .577$

Describing Tail Risks

- "Fat-tailed" and "Thin-tailed"

Definition (Leptokurtosis)

A random variable x_t is said to be leptokurtotic if its kurtosis,

$$\kappa = \frac{\mathrm{E}[(x_t - \mathrm{E}[x_t])^4]}{\mathrm{E}[(x_t - \mathrm{E}[x_t])^2]^2}$$

is greater than that of a normal ($\kappa > 3$). Leptokurtotic variables are also known as "heavy tailed" or "fat tailed".

Definition (Platykurtosis)

A random variable x_t is said to be platykurtotic if its kurtosis,

$$\kappa = \frac{\mathrm{E}[(x_t - \mathrm{E}[x_t])^4]}{\mathrm{E}[(x_t - \mathrm{E}[x_t])^2]^2}$$

is less than that of a normal ($\kappa < 3$). Platykurtotic variables are also known as "thin tailed".

The ARCH(P) model

Definition (P[th] Order ARCH)

An Autoregressive Conditional Heteroskedasticity process or order P is given by

$$\begin{aligned}
r_t &= \mu_t + \epsilon_t \\
\mu_t &= \phi_0 + \phi_1 r_{t-1} + \ldots + \phi_s r_{t-S} \\
\sigma_t^2 &= \omega + \alpha_1 \epsilon_{t-1}^2 + \alpha_2 \epsilon_{t-2}^2 + \ldots + \alpha_P \epsilon_{t-P}^2 \\
\epsilon_t &= \sigma_t e_t \\
e_t &\overset{\text{i.i.d.}}{\sim} N(0,1).
\end{aligned}$$

- Mean μ_t can be an appropriate form - AR, MA, ARMA, ARMAX, etc.
  - $\mathrm{E}_t\left[r_t - \mu_t\right] = 0$
- e_t is the standardized residual, often assumed normal
- σ_t^2 is the conditional variance

Alternative expression of an ARCH(P)

- Model where both mean and variance are time varying
  - Natural extension of model definition for time varying mean model

$$\begin{aligned}
r_t | \mathcal{F}_{t-1} &\sim N(\mu_t, \sigma_t^2) \\
\mu_t &= \phi_0 + \phi_1 r_{t-1} + \ldots + \phi_s r_{t-S} \\
\sigma_t^2 &= \omega + \alpha_1 \epsilon_{t-1}^2 + \alpha_2 \epsilon_{t-2}^2 + \ldots + \alpha_P \epsilon_{t-P}^2 \\
\epsilon_t &= r_t - \mu_t
\end{aligned}$$

- "r_t given the information set at time $t-1$ is conditionally normal with mean μ_t and variance σ_t^2"

Review

Key Concepts

ARCH Model, Volatility Clustering, Conditional Variance, Unconditional Variance, Leptokurtosis

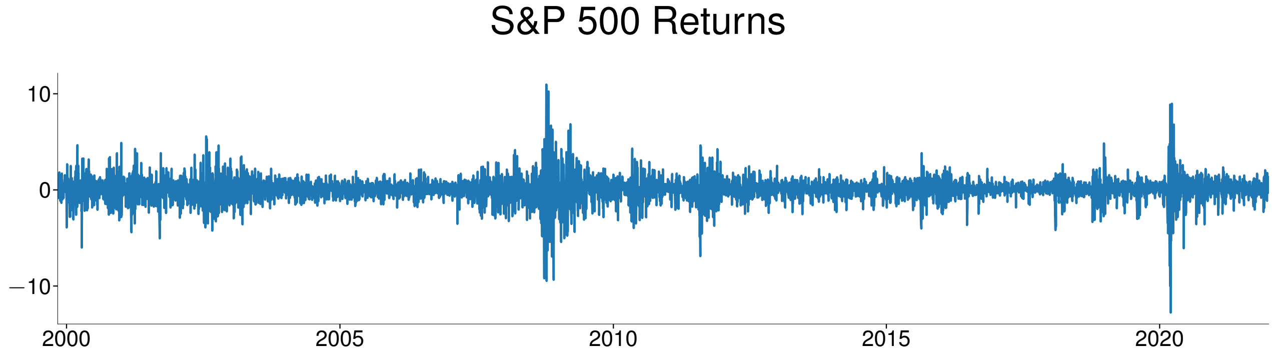
Questions

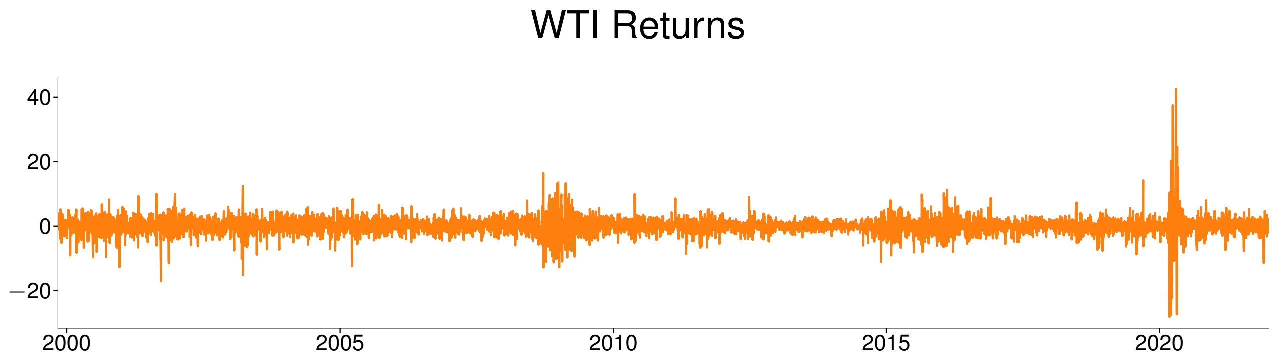
- Why does time-varying volatility always increase kurtosis?
- How is an ARCH(1) model like an AR(1)?

The data

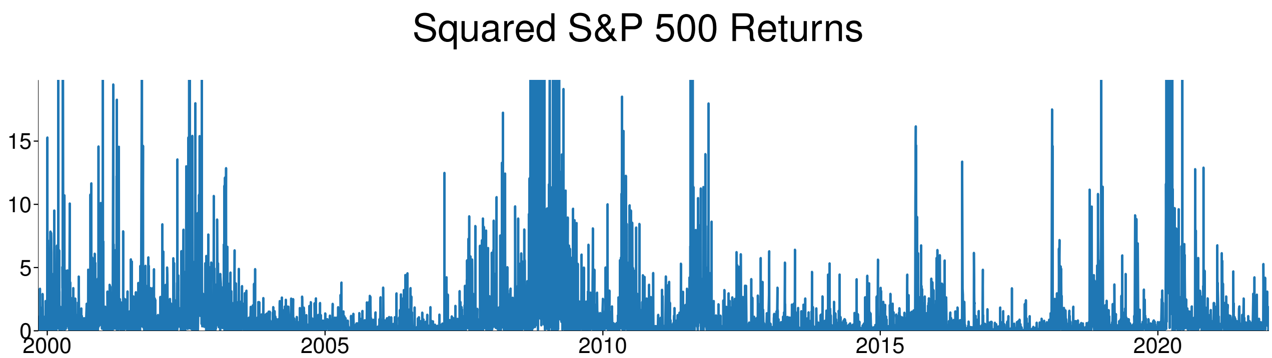
- S&P 500
  - Source: Yahoo! Finance
  - Daily January 1, 1999 - December 31, 2021
  - 5,575 observations
- WTI Spot Prices
  - Source: EIA
  - Daily January 1, 1999 - December 31, 2021
  - 5,726 observations
- All represented as $100\times$ log returns

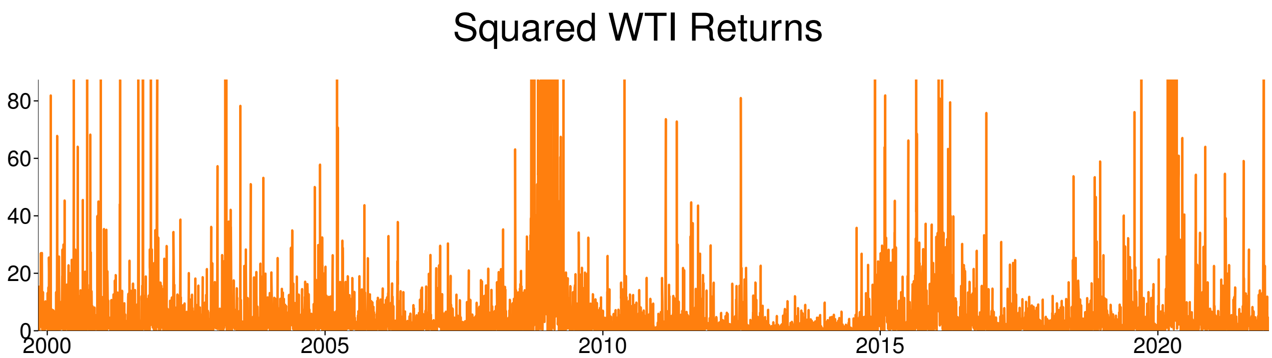
Graphical Evidence of ARCH



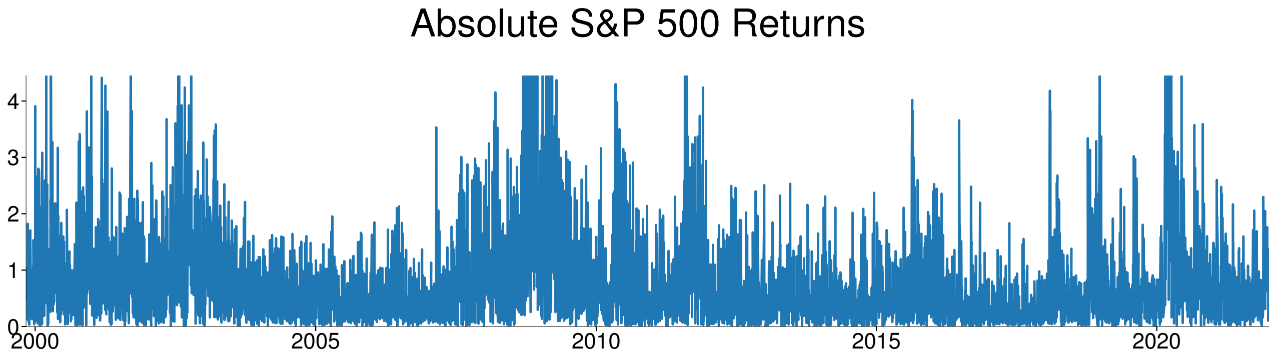


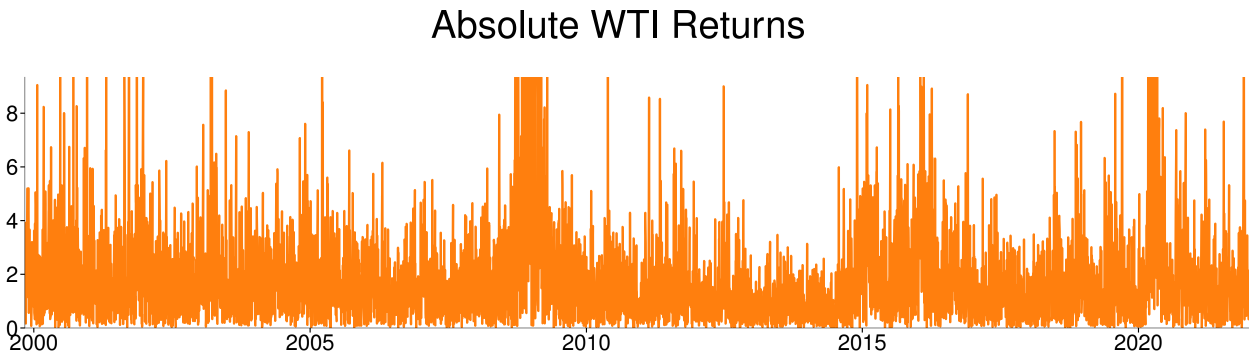
Graphical Evidence: Squared Data Plot





Graphical Evidence: Absolute Data Plot





A simple GARCH(1,1)

$$
\begin{aligned}
r_t &= \epsilon_t \\
\sigma_t^2 &= \omega + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \\
\epsilon_t &= \sigma_t e_t \\
e_t &\overset{\text{i.i.d.}}{\sim} N(0,1)
\end{aligned}
$$

- Adds lagged variance to the ARCH model
- ARCH(∞) in disguise

$$\sigma_t^2 =$$

Important Properties

$$\sigma_t^2 = \omega + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

- Unconditional Variance

$$\bar{\sigma}^2 = \mathrm{E}[\sigma_t^2] = \frac{\omega}{1 - \alpha_1 - \beta_1}$$

- Kurtosis

$$\kappa = \frac{3(1 + \alpha_1 + \beta_1)(1 - \alpha_1 - \beta_1)}{1 - 2\alpha_1\beta_1 - 3\alpha_1^2 - \beta_1^2} > 3$$

- Stationarity
  - $\alpha_1 + \beta_1 < 1$
  - $\omega > 0,\ \alpha_1 \geq 0,\ \beta_1 \geq 0$
  - ARMA in disguise

$$\begin{aligned}
\sigma_t^2 + \epsilon_t^2 - \sigma_t^2 &= \omega + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 + \epsilon_t^2 - \sigma_t^2 \\
\epsilon_t^2 &= \omega + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 + \epsilon_t^2 - \sigma_t^2 \\
\epsilon_t^2 &= \omega + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \epsilon_{t-1}^2 - \beta_1 \nu_{t-1} + \nu_t \\
\epsilon_t^2 &= \omega + (\alpha_1 + \beta_1)\epsilon_{t-1}^2 - \beta_1 \nu_{t-1} + \nu_t
\end{aligned}$$

The Complete GARCH model

Definition (GARCH(P,Q) process)

A Generalized Autoregressive Conditional Heteroskedasticity (GARCH) process of orders P and Q is defined as

$$\begin{aligned}
r_t &= \mu_t + \epsilon_t \\
\mu_t &= \phi_0 + \phi_1 r_{t-1} + \ldots + \phi_s r_{t-S} \\
\sigma_t^2 &= \omega + \sum_{p=1}^{P} \alpha_p \epsilon_{t-p}^2 + \sum_{q=1}^{Q} \beta_q \sigma_{t-q}^2 \\
\epsilon_t &= \sigma_t e_t,\ e_t \overset{\text{i.i.d.}}{\sim} N(0,1)
\end{aligned}$$

- Mean model can be altered to fit data – $AR(S)$ here
- Adds lagged variance to ARCH

Exponentially Weighted Moving Average Variance

A special case of a GARCH(1,1)

- Restricted model where $\mu_t = 0$ for all t, $\omega = 0$ and $\alpha = 1 - \beta$

$$\sigma_t^2 = (1-\lambda)\, r_{t-1}^2 + \lambda \sigma_{t-1}^2$$
$$\sigma_t^2 = (1-\lambda) \sum_{i=0}^{\infty} \lambda^i r_{t-i-1}^2$$

- Note that $\sum_{i=0}^{\infty} \lambda^i = {}^1\!/_{1-\lambda}$ so that $(1-\lambda)\sum_{i=0}^{\infty} \lambda^i = 1$
  - Leads to random-walk-like features

Review

Key Concepts

Generalized ARCH, EWMA Variance

Questions

- How does GARCH improve ARCH?
- How many lags are needed in an ARCH to match the fit of a GARCH?
- What restrictions are needed on a GARCH model produce an EWMA variance?

Glosten-Jagannathan-Runkle GARCH

- Extends GARCH(1,1) to include an asymmetric term

Definition (Glosten-Jagannathan-Runkle (GJR) GARCH process)

A GJR-GARCH(P,O,Q) process is defined as

$$
\begin{aligned}
r_t &= \mu_t + \epsilon_t \\
\mu_t &= \phi_0 + \phi_1 r_{t-1} + \ldots + \phi_s r_{t-S} \\
\sigma_t^2 &= \omega + \sum_{p=1}^{P} \alpha_p \epsilon_{t-p}^2 + \sum_{o=1}^{O} \gamma_o \epsilon_{t-o}^2 I_{[\epsilon_{t-o}<0]} + \sum_{q=1}^{Q} \beta_q \sigma_{t-q}^2 \\
\epsilon_t &= \sigma_t e_t \\
e_t &\overset{\text{i.i.d.}}{\sim} N(0,1)
\end{aligned}
$$

where $I_{[\epsilon_{t-o}<0]}$ is an indicator function that takes the value 1 if $\epsilon_{t-o} < 0$ and 0 otherwise.

GJR-GARCH(1,1,1) example

- GJR(1,1,1) model

$$
\begin{aligned}
\sigma_t^2 = \omega + &\alpha_1 \epsilon_{t-1}^2 + \gamma_1 \epsilon_{t-1}^2 I_{[\epsilon_{t-1}<0]} + \beta_1 \sigma_{t-1}^2 \\
&\alpha_1 + \gamma_1 \geq 0 \\
&\alpha_1 \geq 0 \\
&\beta_1 \geq 0 \\
&\omega > 0
\end{aligned}
$$

- $\gamma_1 \epsilon_{t-1}^2 I_{[\epsilon_{t-1}<0]}$: Variances are larger after negative shocks than after positive shocks
- "Leverage Effect"

Threshold ARCH

- Threshold ARCH is similar to GJR-GARCH
- Also known as ZARCH (Zakoain (1994)) or AVGARCH when symmetric

Definition (Threshold ARCH (TARCH) process)

A TARCH(P,O,Q) process is defined

$$
\begin{aligned}
r_t &= \mu_t + \epsilon_t \\
\mu_t &= \phi_0 + \phi_1 r_{t-1} + \ldots + \phi_s r_{t-S} \\
\sigma_t &= \omega + \sum_{p=1}^{P} \alpha_p |\epsilon_{t-p}| + \sum_{o=1}^{O} \gamma_o |\epsilon_{t-o}| I_{[\epsilon_{t-o}<0]} + \sum_{q=1}^{Q} \beta_q \sigma_{t-q} \\
\epsilon_t &= \sigma_t e_t \\
e_t &\overset{\text{i.i.d.}}{\sim} N(0,1)
\end{aligned}
$$

where $I_{[\epsilon_{t-o}<0]}$ is an indicator function that is 1 if $\epsilon_{t-o} < 0$ and 0 otherwise.

TARCH(1,1,1) example

- TARCH(1,1,1) model

$$
\begin{aligned}
\sigma_t = \omega + \alpha_1 |\epsilon_{t-1}| + \gamma_1 |\epsilon_{t-1}| I_{[\epsilon_{t-1}<0]} + \beta_1 \sigma_{t-1} \\
\alpha_1 + \gamma_1 \geq 0 \\
\omega > 0, \alpha_1 \geq 0, \beta_1 \geq 0
\end{aligned}
$$

- Note the different power: σ_t and $|\epsilon_{t-1}|$
  - Model for conditional standard deviation
- *Nonlinear* variance models complicate some things
  - Forecasting
  - Memory of volatility
  - *News impact curves*
- GARCH(P,Q) becomes TARCH(P,O,Q) or GJR-GARCH(P,O,Q)
- TARCH and GJR-GARCH are sometimes (*wrongly*) used interchangeably.

EGARCH

Definition (EGARCH(P,O,Q) process)

An Exponential Generalized Autoregressive Conditional Heteroskedasticity (EGARCH) process of order P, O and Q is defined

$$
\begin{aligned}
r_t &= \mu_t + \epsilon_t \\
\mu_t &= \phi_0 + \phi_1 r_{t-1} + \ldots + \phi_s r_{t-S} \\
\ln(\sigma_t^2) &= \omega + \sum_{p=1}^{P} \alpha_p \left(\left| \frac{\epsilon_{t-p}}{\sigma_{t-p}} \right| - \sqrt{\frac{2}{\pi}} \right) + \sum_{o=1}^{O} \gamma_o \frac{\epsilon_{t-o}}{\sigma_{t-o}} + \sum_{q=1}^{Q} \beta_q \ln(\sigma_{t-q}^2) \\
\epsilon_t &= \sigma_t e_t \\
e_t &\overset{\text{i.i.d.}}{\sim} N(0,1)
\end{aligned}
$$

In the original parametrization of Nelson (1991), P and O were required to be identical.

EGARCH(1,1,1)

- EGARCH(1,1,1)

$$
\begin{aligned}
r_t &= \mu + \epsilon_t \\
\ln(\sigma_t^2) &= \omega + \alpha_1 \left(\left| \frac{\epsilon_{t-1}}{\sigma_{t-1}} \right| - \sqrt{\frac{2}{\pi}} \right) + \gamma_1 \frac{\epsilon_{t-1}}{\sigma_{t-1}} + \beta_1 \ln(\sigma_{t-1}^2) \\
\epsilon_t &= \sigma_t e_t, \quad e_t \overset{\text{i.i.d.}}{\sim} N(0,1)
\end{aligned}
$$

- Modeling using $\ln$ removes any parameter restrictions ($|\beta_1| < 1$)
- AR(1) with *two* shocks

$$
\ln(\sigma_t^2) = \omega + \alpha_1 \left(|e_{t-1}| - \sqrt{\frac{2}{\pi}} \right) + \gamma_1 e_{t-1} + \beta_1 \ln(\sigma_{t-1}^2)
$$

  - Symmetric shock $\left(|e_{t-1}| - \sqrt{\frac{2}{\pi}} \right)$ and asymmetric shock e_{t-1}
    - Note, shocks are standardized residuals (unit variance)
- Often provides a better fit that GARCH(P,Q)

Asymmetric Power ARCH

- Nests ARCH, GARCH, TARCH, GJR-GARCH, EGARCH (almost) and other specifications
- Only present the APARCH(1,1,1):

$$\sigma_t^\delta = \omega + \alpha_1 \left(|\epsilon_{t-1}| + \gamma_1 \epsilon_{t-1}\right)^\delta + \beta_1 \sigma_{t-1}^\delta$$
$$\alpha_1 > 0, \quad -1 \leq \gamma_1 \leq 1, \quad \delta > 0, \quad \beta_1 \geq 0, \quad \omega > 0$$

- Parametrizes the "power" parameter
- Different values for δ affect the persistence.
  - Lower values $\Rightarrow$ higher persistence of shocks
    - ARCH: $\gamma = 0, \beta = 0, \delta = 2$
    - GARCH: $\gamma = 0, \delta = 2$
    - GJR-GARCH: $\delta = 2$
    - AVGARCH: $\gamma = 0, \delta = 1$
    - TARCH: $\delta = 1$
    - EGARCH: (almost) $\lim \delta \to 0$

Review

Key Concepts

Threshold ARCH, GJR-GARCH, Exponential GARCH, Asymmetric Power ARCH

Questions

- How goes a GJR-GARCH model improve a GARCH model?
- How do TARCH and EGARCH differ from GJR-GARCH?
- Why does the EGARCH model contains the term $-\sqrt{\frac{2}{\pi}}$?
- Do the asymmetric models allow asymmetries in both directions (i.e., more sensitive to positive than negative, or more sensitive to negative than to positive)?

Problems

1. Show that APARCH and GARCH are equivalent under the necessary parameter restrictions.

S&P Results

ARCH(5)

ω	α_1	α_2	α_3	α_4	α_5	Log Lik.
0.288 (0.000)	0.104 (0.000)	0.199 (0.000)	0.182 (0.000)	0.194 (0.000)	0.152 (0.000)	-6712

GARCH(1,1)

ω	α_1	β_1	Log Lik.
0.019 (0.000)	0.106 (0.000)	0.881 (0.000)	-6597

EGARCH(1,1,1)

ω	α_1	γ_1	β_1	Log Lik.
0.000 (0.983)	0.137 (0.000)	-0.153 (0.000)	0.974 (0.000)	-6484

Comparing different models

- Comparing models which are not nested can be difficult
- The *News Impact Curve* provides one method
- Defined:

$$n(e_t) = \sigma^2_{t+1}(e_t|\sigma^2_t = \bar{\sigma}^2)$$

$$NIC(e_t) = n(e_t) - n(0)$$

- Measures the effect of a shock *starting* at the unconditional variance
- Allows for asymmetric shapes

GARCH(1,1)

$$NIC(e_t) = \alpha_1\bar{\sigma}^2 e_t^2$$

GJR-GARCH(1,1,1)

$$NIC(e_t) = (\alpha_1 + \gamma_1 I_{[e_t<0]})\bar{\sigma}^2 e_t^2$$

TARCH(1,1,1)

$$NIC(e_t) = (\alpha_1 + \gamma_1 I_{[\epsilon_t<0]})^2\bar{\sigma}^2 e_t^2 + (2\omega + 2\beta_1\bar{\sigma})(\alpha_1 + \gamma_1 I_{[e_t<0]})|e_t|$$

Review

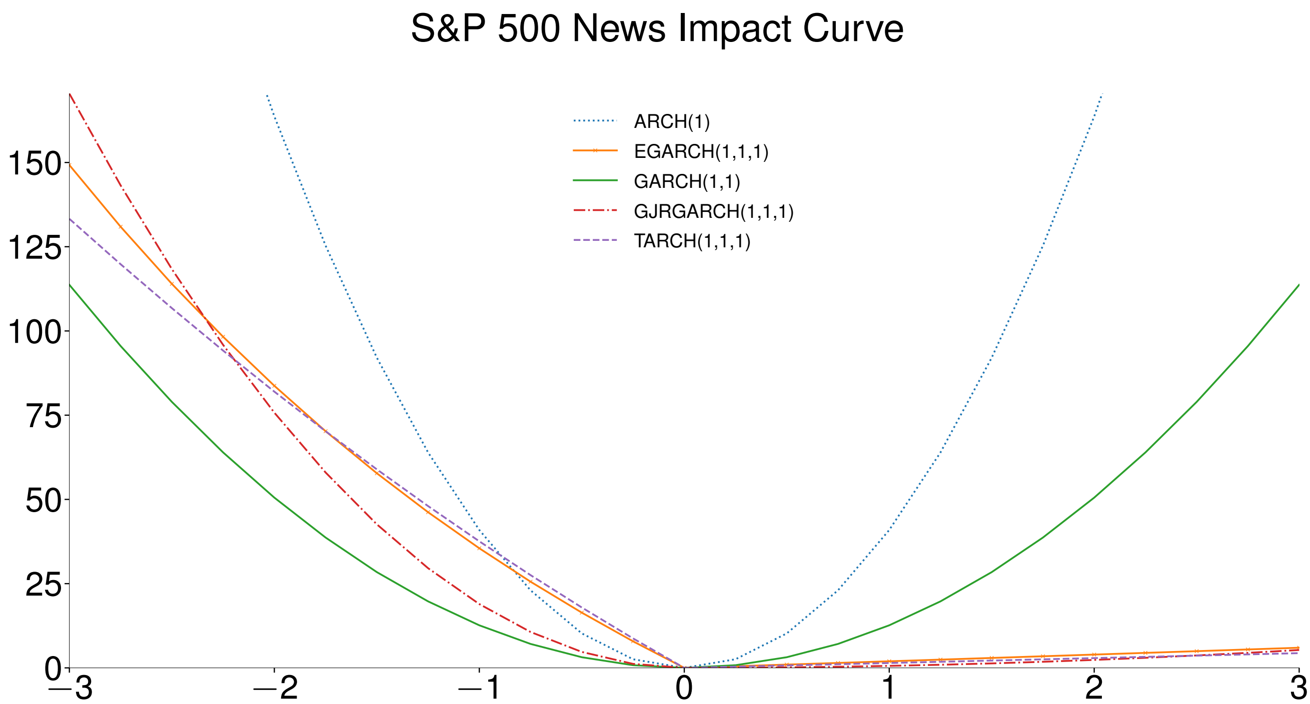
Key Concepts

News Impact Curve

Questions

- How are News Impact Curves used?
- Why is the unconditional variance/volatility used in NICs?

S&P 500 News Impact Curves



Estimation

$$\begin{aligned} r_t &= \mu_t + \epsilon_t \\ \sigma_t^2 &= \omega + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \\ \epsilon_t &= \sigma_t e_t \\ e_t &\overset{\text{i.i.d.}}{\sim} N(0,1) \end{aligned}$$

- So:

$$r_t | \mathcal{F}_{t-1} \sim N(\mu_t, \sigma_t^2)$$

- Need initial values for σ_0^2 and ϵ_0^2 to start recursion
  - Normal Maximum Likelihood is a natural choice

$$f(\mathbf{r}; \boldsymbol{\theta}) = \prod_{t=1}^{T} (2\pi\sigma_t^2)^{-\frac{1}{2}} \exp\left(-\frac{(r_t - \mu_t)^2}{2\sigma_t^2}\right)$$

$$l(\boldsymbol{\theta}; \mathbf{r}) = \sum_{t=1}^{T} -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma_t^2) - \frac{(r_t - \mu_t)^2}{2\sigma_t^2}.$$

Inference

- MLE are asymptotically normal

$$\sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N(0, \mathcal{I}^{-1}), \quad \mathcal{I} = -\mathrm{E}\left[\frac{\partial^2 l(\boldsymbol{\theta}_0; r_t)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}\right]$$

- If data are not conditionally normal, Quasi MLE (QMLE)

$$\sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N(0, \mathcal{I}^{-1}\mathcal{J}\mathcal{I}^{-1}), \quad \mathcal{J} = \mathrm{E}\left[\frac{\partial l(\boldsymbol{\theta}_0; r_t)}{\partial \boldsymbol{\theta}} \frac{\partial l(\boldsymbol{\theta}_0; r_t)}{\partial \boldsymbol{\theta}'}\right]$$

- Known as Bollerslev-Wooldridge Covariance estimator in GARCH models
  - Also known as a "*sandwich*" covariance estimator
  - Default `cov_type="robust"` in `arch` package code
  - White and Newey-West Covariance estimators are also sandwich estimators

Independence of the mean and variance

- Use LS to estimate mean parameters, then use estimated residuals in GARCH
- Efficient estimates one of two ways
- Joint estimation of mean and variance parameters using MLE
- GLS estimation
  - Estimate mean and variance in 2-steps as above
  - Re-estimate mean using GLS
  - Re-estimate variance using new set of residuals

The mean and the variance can be estimated consistently using 2-stages. Standard errors are also correct as long as a robust VCV estimator is used.

Review

Key Concepts

Quasi MLE, 2-step estimation, Bollerslev-Wooldridge Covariance

Questions

- How are parameters of ARCH model estimated?
- When are Bollerslev-Wooldridge standard errors needed?
- Under what condition is 2-step estimation consistent?
- What is needed when making inference about the mean parameters when using 2-step estimation?

Alternative Distributional Assumptions

- Equity returns are *not* conditionally normal
- Can replace the normal likelihood with a more realistic one
- Common choices:
- Standardized Student's t
  - Nests the normal as $\nu \to \infty$
- Generalized error distribution
  - Nests the normal when $\nu = 2$
- Hansen's Skew-T
  - Captures both skewness and heavy tails
  - Use *hyperparameters* to control shape (ν and λ)
- All can have heavy tails
- Only Skew-T is skewed
- Dozens more in academic research
- But for what gain?

Review

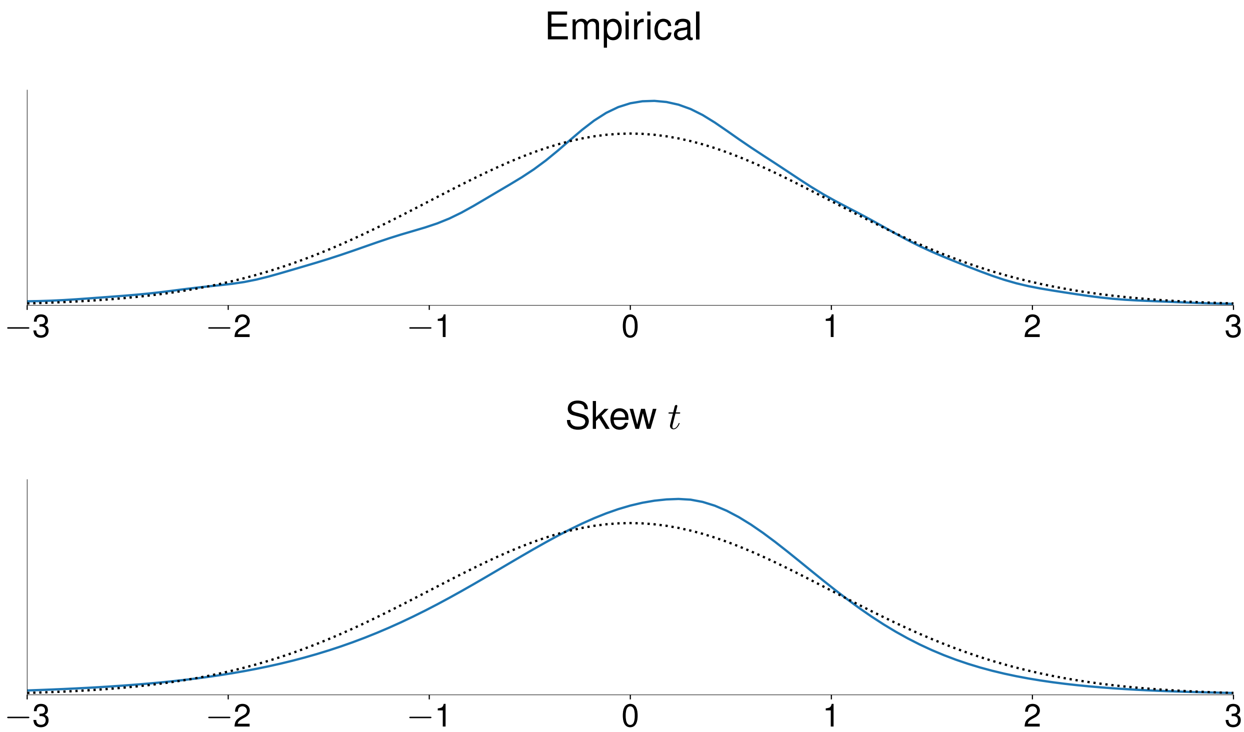
Key Concepts

GARCH-in-mean, Standardized Student's t, Generalized Error Distribution, Skew t Distribution

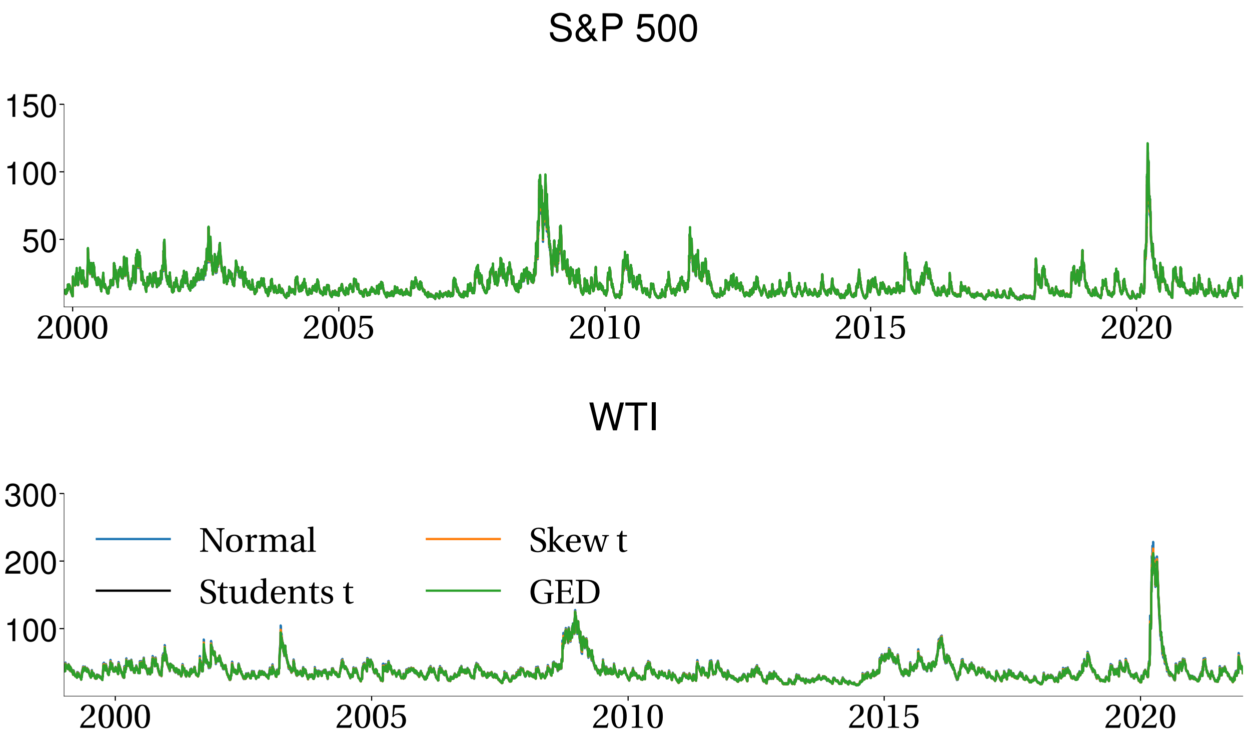
Questions

- Why can GARCH models be estimated in 2 steps while GARCH-in-mean cannot?
- What features are missing from the normal distribution when modeling financial return data?

S&P 500 Density



Effect of dist. choice on estimated volatility



Model Building

- ARCH and GARCH models are essentially ARMA models
  - Box-Jenkins Methodology
    - Parsimony principle

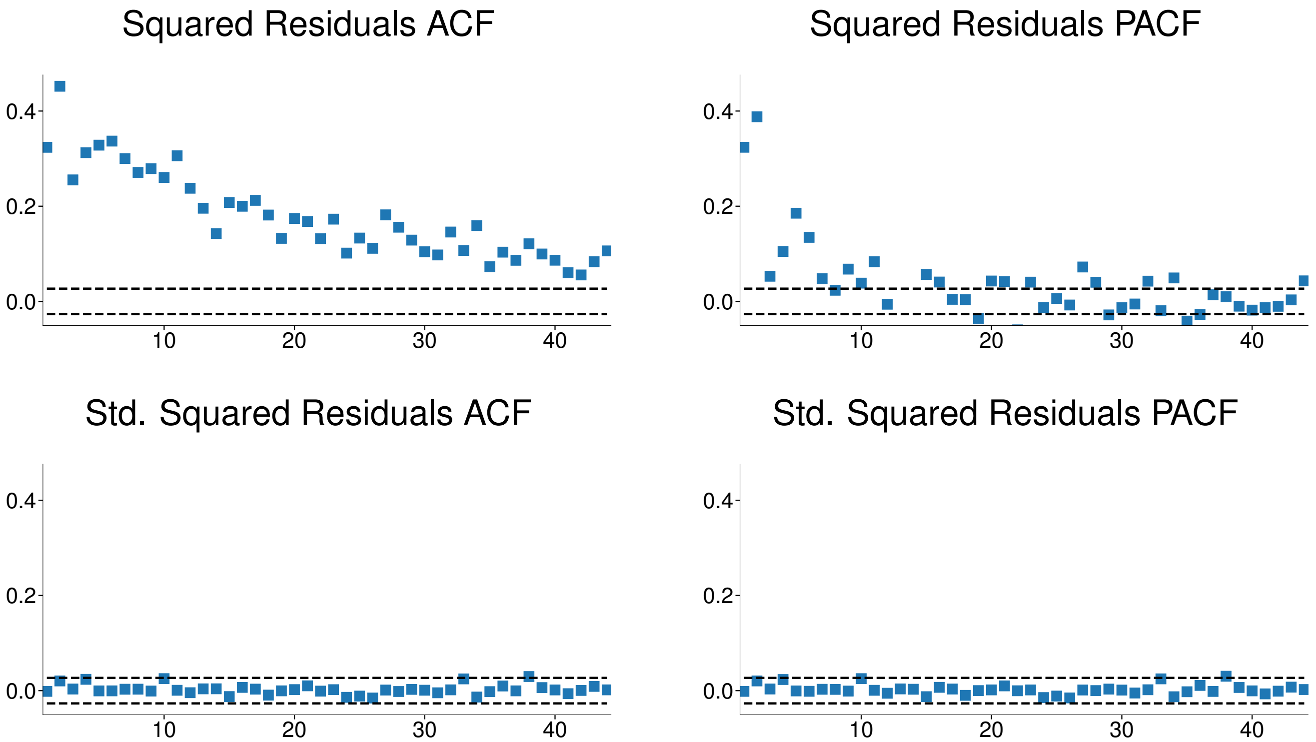
Steps:

1. Inspect the ACF and PACF of ϵ_t^2

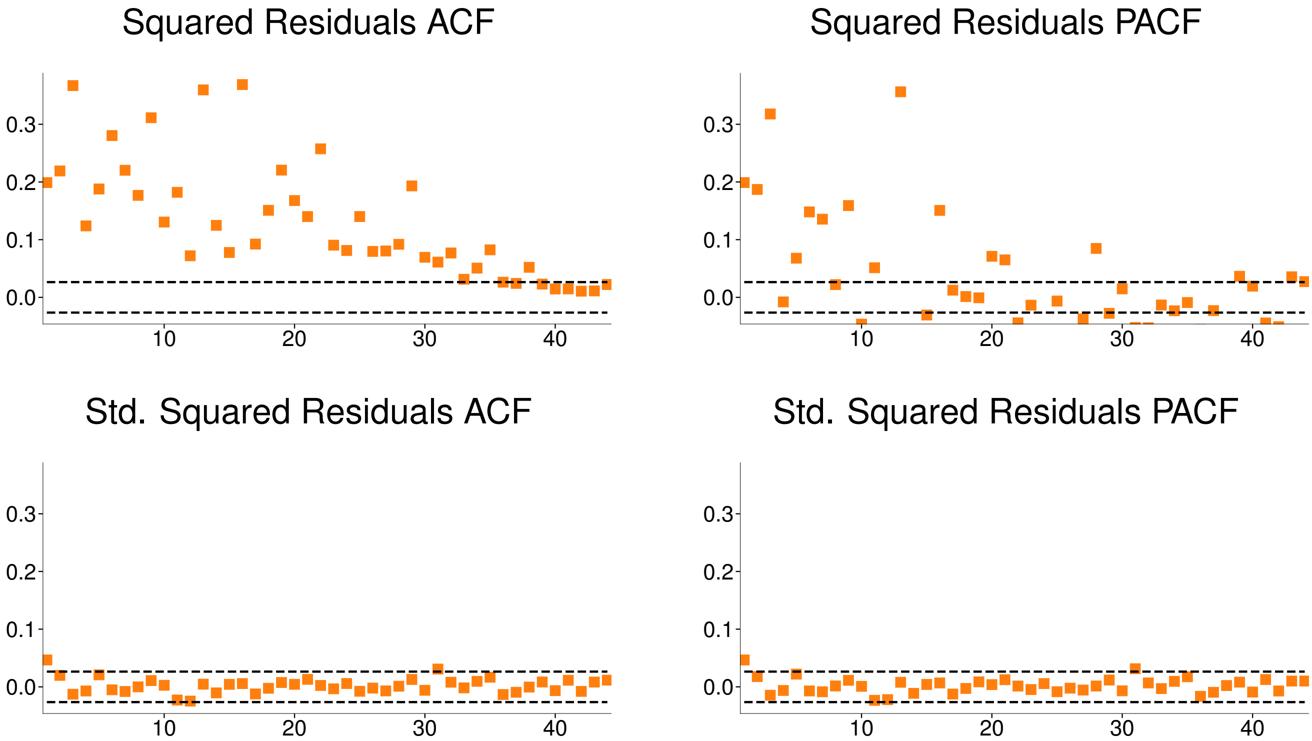
$$\epsilon_t^2 = \omega + (\alpha + \beta)\epsilon_{t-1}^2 - \beta\nu_{t-1} + \nu_t$$

   - ACF indicates α (or ARCH of any kind)
   - PACF indicates β
2. Build initial model based on these observation
3. Iterate between model and ACF/PACF of $\hat{e}_t^2 = \frac{\epsilon_t^2}{\hat{\sigma}_t^2}$

S&P 500 ϵ_t^2 ACF/PACF



WTI ϵ_t^2 ACF/PACF



How I built a model for the S&P 500

	α_1	α_2	γ_1	γ_2	β_1	β_2	Log Lik.
GARCH(1,1)	0.106 (0.000)				0.881 (0.000)		−6597.4
GARCH(1,2)	0.106 (0.000)				0.881 (0.000)	0.000 (0.999)	−6597.4
GARCH(2,1)	0.073 (0.002)	0.049 (0.105)			0.861 (0.000)		−6594.1
GJR-GARCH(1,1,1)	0.000 (0.999)		0.184 (0.000)		0.889 (0.000)		−6491.0
GJR-GARCH(1,2,1)	0.000 (0.999)		0.165 (0.000)	0.024 (0.603)	0.885 (0.000)		−6490.7
TARCH(1,1,1)⋆	0.000 (0.999)		0.173 (0.000)		0.907 (0.000)		−6469.4
TARCH(1,2,1)	0.000 (0.999)		0.169 (0.000)	0.005 (0.888)	0.907 (0.000)		−6469.4
TARCH(2,1,1)	0.000 (0.999)	0.003 (0.938)	0.172 (0.000)		0.906 (0.000)		−6469.3
EGARCH(1,0,1)	0.217 (0.000)				0.978 (0.000)		−6619.9
EGARCH(1,1,1)	0.137 (0.000)		−0.153 (0.000)		0.974 (0.000)		−6484.3
EGARCH(1,2,1)	0.129 (0.000)		−0.212 (0.000)	0.067 (0.055)	0.976 (0.000)		−6479.5
EGARCH(2,1,1)	0.029 (0.535)	0.121 (0.014)	−0.161 (0.000)		0.970 (0.000)		−6476.8

Testing for (G)ARCH

- ARCH is autocorrelation in ϵ_t^2
- All ARCH processes have this, whether GARCH or EGARCH or other
  - ARCH-LM test
  - Directly test for autocorrelation:

$$\epsilon_t^2 = \phi_0 + \phi_1 \epsilon_{t-1}^2 + \ldots + \phi_P \epsilon_{t-P}^2 + \eta_t$$

  - $H_0 : \phi_1 = \phi_2 = \ldots = \phi_P = 0$
  - $T \times R^2 \xrightarrow{d} \chi_P^2$
  - Standard LM test from a regression.
  - More powerful test: Fit an ARCH(P) model
  - The forbidden hypothesis

$$\sigma_t^2 = \omega + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

$$H_0 : \alpha_1 = 0, \; H_1 : \alpha > 0$$

Review

Key Concepts

ARCH-LM test

Questions

- How is model building of ARCH models similar to model building of ARMA models?
- What does an ARCH-LM test detect?

Forecasting: ARCH(1)

- Simple ARCH model

$$\epsilon_t \sim N(0, \sigma_t^2)$$
$$\sigma_t^2 = \omega + \alpha_1 \epsilon_{t-1}^2$$

  - **1-step ahead forecast is known today**
  - All ARCH-family models have this property

$$\begin{aligned} \epsilon_t &\sim N(0, \sigma_t^2) \\ \sigma_t^2 &= \omega + \alpha_1 \epsilon_{t-1}^2 \\ \mathrm{E}_t[\sigma_{t+1}^2] &= \mathrm{E}_t[\omega + \alpha_1 \epsilon_t^2] \\ &= \omega + \alpha_1 \epsilon_t^2 \end{aligned}$$

  - Note: $\mathrm{E}_t[\epsilon_{t+1}^2] = \mathrm{E}_t[e_{t+1}^2 \sigma_{t+1}^2] = \sigma_{t+1}^2 \mathrm{E}_t[e_{t+1}^2] = \sigma_{t+1}^2$
  - Further: $\mathrm{E}_t[\epsilon_{t+h}^2] = \mathrm{E}_t[\mathrm{E}_{t+h-1}[e_{t+h}^2 \sigma_{t+h}^2]] = \mathrm{E}_t[\mathrm{E}_{t+h-1}[e_{t+h}^2]\sigma_{t+h}^2] = \mathrm{E}_t[\sigma_{t+h}^2]$

Forecasting: ARCH(1)

- 2-step ahead

$$\mathrm{E}_t[\sigma_{t+2}^2] =$$

- h-step ahead forecast

$$\mathrm{E}_t[\sigma_{t+h}^2] = \sum_{i=0}^{h-1} \alpha_1^i \omega + \alpha_1^h \epsilon_t^2$$

  - Just the AR(1) forecasting formula
    - Why?

Forecasting: GARCH(1,1)

- 1-step ahead

$$\begin{aligned}\mathrm{E}_t[\sigma^2_{t+1}] &= \mathrm{E}_t[\omega + \alpha_1\epsilon_t^2 + \beta_1\sigma_t^2] \\ &= \omega + \alpha_1\epsilon_t^2 + \beta_1\sigma_t^2\end{aligned}$$

- 2-step ahead

$$\mathrm{E}_t[\sigma^2_{t+2}] =$$

Forecasting: GARCH(1,1)

- h-step ahead

$$\mathrm{E}_t[\sigma^2_{t+h}] = \sum_{i=0}^{h-1}(\alpha_1+\beta_1)^i\omega + (\alpha_1+\beta_1)^{h-1}(\alpha_1\epsilon_t^2 + \beta_1\sigma_t^2)$$

- Also essentially an AR(1), technically ARMA(1,1)

Forecasting: TARCH(1,0,0)

- This one is a mess
  - *Nonlinearities* cause problems
    - All ARCH-family models are nonlinear, but some are linearity in ϵ_t^2
    - Others are not

$$\sigma_t = \omega + \alpha_1 |\epsilon_{t-1}|$$

  - **Forecast for $t+1$ is known at time t**
    - **Always, always, always,** . . .

$$\begin{aligned}
\mathrm{E}_t[\sigma_{t+1}^2] &= \mathrm{E}_t[(\omega + \alpha_1|\epsilon_t|)^2] \\
&= \mathrm{E}_t[\omega^2 + 2\omega\alpha_1|\epsilon_t| + \alpha_1^2\epsilon_t^2] \\
&= \omega^2 + 2\omega\alpha_1\mathrm{E}_t[|\epsilon_t|] + \alpha_1^2\mathrm{E}_t[\epsilon_t^2] \\
&= \omega^2 + 2\omega\alpha_1|\epsilon_t| + \alpha_1^2\epsilon_t^2
\end{aligned}$$

TARCH(1,0,0) continued...

- Multi-step is less straightforward

$$\begin{aligned}
\mathrm{E}_t[\sigma_{t+2}^2] &= \mathrm{E}_t[(\omega + \alpha_1|\epsilon_{t+1}|)^2] \\
&= \mathrm{E}_t[\omega^2 + 2\omega\alpha_1|\epsilon_{t+1}| + \alpha_1^2\epsilon_{t+1}^2] \\
&= \omega^2 + 2\omega\alpha_1\mathrm{E}_t[|\epsilon_{t+1}|] + \alpha_1^2\mathrm{E}_t[\epsilon_{t+1}^2] \\
&= \omega^2 + 2\omega\alpha_1\mathrm{E}_t[|e_{t+1}|\sigma_{t+1}] + \alpha_1^2\mathrm{E}_t[e_t^2\sigma_{t+1}^2] \\
&= \omega^2 + 2\omega\alpha_1\mathrm{E}_t[|e_{t+1}|]\mathrm{E}_t[\sigma_{t+1}] + \alpha_1^2\mathrm{E}_t[e_t^2]\mathrm{E}_t[\sigma_{t+1}^2] \\
&= \omega^2 + 2\omega\alpha_1\mathrm{E}_t[|e_{t+1}|](\omega + \alpha_1|\epsilon_t|) + \alpha_1^2 \cdot 1 \cdot (\omega^2 + 2\omega\alpha_1|\epsilon_t| + \alpha_1^2\epsilon_t^2)
\end{aligned}$$

If $e_{t+1} \sim N(0,1)$, $E[|e_{t+1}|] = \sqrt{\frac{2}{\pi}}$

$$\mathrm{E}_t[\sigma_{t+2}^2] = \omega^2 + 2\omega\alpha_1\sqrt{\frac{2}{\pi}}(\omega + \alpha_1|\epsilon_t|) + \alpha_1^2(\omega^2 + 2\omega\alpha_1|\epsilon_t| + \alpha_1^2\epsilon_t^2)$$

Simulation-based Forecasting

- Multi-step forecasting using simulation is simple
- Two options
  - Parametric: $e_t \overset{\text{i.i.d.}}{\sim} F\left(0, 1, \hat{\theta}\right)$
  - Bootstrap: Sample i.i.d. from $\{\hat{e}_i\}_{i=1}^{t}$ where $\hat{e}_i = \hat{\epsilon}_i / \hat{\sigma}_i = (r_i - \hat{\mu}_i)/\hat{\sigma}_i$

Algorithm (Simulation-based Forecast)

For $b = 1, \ldots, B$ do:

1. *Sample $h-1$ i.i.d. values from either the parametric or bootstrap distribution*
2. *Simulate the model for h periods and store $\hat{\sigma}^2_{t+h|t,b}$*

Construct the forecast as $\hat{\sigma}^2_{t+h|t} = B^{-1} \sum_{b=1}^{B} \hat{\sigma}^2_{t+h|t,j}$

Notes

- If model parametrizes $g\left(\sigma_t^2\right)$ than at each period $h > 1$ the simulated value is $\epsilon_{t+h,j} = \sqrt{g^{-1}\left(g\left(\sigma^2_{t+h|t,j}\right)\right)}\eta_{h,j}$ where $\eta_{h,j}$ are the i.i.d.samples
- $\sigma^2_{t+1|t}$ is always known at time t and so simulation is never needed for 1-step forecasting

Review

Key Concepts

Linearity in ϵ_t^2, Iterated Expectations

Questions

- What property do all ARCH models share in terms of their forecasts?
- What happens to the long-run forecast from an ARCH model?
- Why do models that are linear in ϵ_t^2 simple to use in forecasting?
- Why are models like TARCH difficult to forecast over multiple steps?

Problems

1. If $Y_t = \phi Y_{t-1} + \epsilon_t$ where $|\phi| < 1$ and $\sigma_t^2 = \omega + \alpha \epsilon_{t-1}^2$, what are the 1 and 2-step forecasts of $\mathrm{V}_t\left[\epsilon_{t+h}\right]$?
2. What are the 1 and 2-step forecasts of $\mathrm{V}_t\left[Y_{t+h}\right]$?

Assessing forecasts: Augmented MZ

- Start from $\mathrm{E}_t[r^2_{t+h}] \approx \sigma^2_{t+h|t}$
  - Standard Augmented MZ regression:

$$\epsilon^2_{t+h} - \hat{\sigma}^2_{t+h|t} = \gamma_0 + \gamma_1 \hat{\sigma}^2_{t+h|t} + \gamma_2 z_{1t} + \ldots + \gamma_{K+1} z_{Kt} + \eta_t$$

  - η_t is heteroskedastic in proportion to σ^2_t: Use GLS.
  - An improved GMZ regression (GMZ-GLS)

$$\frac{\epsilon^2_{t+h} - \hat{\sigma}^2_{t+h|t}}{\hat{\sigma}^2_{t+h|t}} = \gamma_0 \frac{1}{\hat{\sigma}^2_{t+h|t}} + \gamma_1 1 + \gamma_2 \frac{z_{1t}}{\hat{\sigma}^2_{t+h|t}} + \ldots + \gamma_{K+1} \frac{z_{Kt}}{\hat{\sigma}^2_{t+h|t}} + \nu_t$$

  - Better to use *Realized Variance* to evaluate forecasts

$$RV_{t+h} - \hat{\sigma}^2_{t+h|t} = \gamma_0 + \gamma_1 \hat{\sigma}^2_{t+h|t} + \gamma_2 z_{1t} + \ldots + \gamma_{K+1} z_{Kt} + \eta_t$$

  - Also can use GLS version
  - Both RV_{t+h} and ϵ^2_{t+h} are proxies for the variance at $t+h$
    - RV is just better, often 10×+ more precise

Assessing forecasts: Diebold-Mariano

- Relative forecast performance
  - MSE loss

$$\delta_t = \left(\epsilon^2_{t+h} - \hat{\sigma}^2_{A,t+h|t}\right)^2 - \left(\epsilon^2_{t+h} - \hat{\sigma}^2_{B,t+h|t}\right)^2$$

  - $H_0 : \mathrm{E}[\delta_t] = 0$, $H_1^A : \mathrm{E}[\delta_t] < 0$, $H_1^B : \mathrm{E}[\delta_t] > 0$

$$\hat{\bar{\delta}} = R^{-1} \sum_{r=1}^{R} \delta_r$$

  - Standard t-test, 2-sided alternative
  - Newey-West covariance always needed
  - Better DM using QLIK loss (Normal log-likelihood "Kernel")

$$\delta_t = \left(\ln(\hat{\sigma}^2_{A,t+h|t}) + \frac{\epsilon^2_{t+h}}{\hat{\sigma}^2_{A,t+h|t}}\right) - \left(\ln(\hat{\sigma}^2_{B,t+h|t}) + \frac{\epsilon^2_{t+h}}{\hat{\sigma}^2_{B,t+h|t}}\right)$$

  - Patton & Sheppard (2009)

Review

Key Concepts

Mincer-Zarnowitz GLS, QLIK loss

Questions

- Why is GLS useful in forecast evaluation?
- Why is the QLIK loss preferred to MSE in volatility model evaluation?

Realized Variance

- Variance measure computed using ultra-high-frequency data (UHF)
  - Uses all available information to estimate the variance over some period
    - Usually 1 day
  - Variance estimates from RV can be treated as "observable"
    - Standard ARMA modeling
    - Variance estimates are consistent
    - Asymptotically unbiased
    - Variance converges to 0 as the number of samples increases
  - Problems arise when applied to market data
    - Noise (bid-ask bounce)
    - Market closure
    - Prices discrete
    - Prices not continuously observable
    - Data quality

Realized Variance

- Assumptions
  - Log-prices are generated by an arbitrage-free semi-martingale
    - Prices are observable
    - Prices can be sampled often
  - Defined

$$RV_t^{(m)} = \sum_{i=1}^{m} (p_{i,t} - p_{i-1,t})^2 = \sum_{i=1}^{m} r_{i,t}^2.$$

    - m-sample Realized Variance
    - $p_{i,t}$ is the i[th] log-price on day t
    - $r_{i,t}$ is the i[th] return on day t
  - Only uses information on day t to estimate the variance on day t
  - Consistent estimator of the integrated variance

$$\int_t^{t+1} \sigma_s^2 ds$$

  - "Total variance" on day t

Why Realized Variance Works

- Consider a simple Brownian motion

$$dp_t = \mu\,\mathrm{d}t + \sigma\,\mathrm{d}W_t$$

- m-sample Realized Variance

$$RV_t^{(m)} = \sum_{i=1}^{m} r_{i,t}^2$$

- Returns are i.i.d. normal

$$r_{i,t} \overset{\text{i.i.d.}}{\sim} N\left(\frac{\mu}{m}, \frac{\sigma^2}{m}\right)$$

- Nearly unbiased

$$\mathrm{E}\left[RV_t^{(m)}\right] = \frac{\mu^2}{m} + \sigma^2$$

- Variance close to 0

$$\mathrm{V}\left[RV_t^{(m)}\right] = 4\frac{\mu^2\sigma^2}{m^2} + 2\frac{\sigma^4}{m}$$

Why Realized Variance Works

- Works for models with time-varying drift and stochastic volatility

$$dp_t = \mu_t \, dt + \sigma_t \, dW_t$$

  - No arbitrage imposes some restrictions on μ_t
  - Works with price processes with jumps
  - In the general case:

$$RV_t^{(m)} \xrightarrow{p} \int_t^{t+1} \sigma_s^2 ds + \sum_{n=1}^{N} J_n^2$$

  - J_n are jumps

Why Realized Variance Doesn't Work

- Multiple prices at the same time
  - Define the price as the average share price (volume weighted price)
  - Use simple average or median
  - Not a problem
- Prices only observed on a discrete grid
  - $.01 or £.0025
  - Nothing can be done
  - Small problem
- Data quality
  - UHF price data is generally messy
  - Typos
  - Wrong time-stamps
  - Pre-filter to remove obvious errors
  - Often remove “round trips”
- No price available at some point in time
  - Use the last observed price: *last price interpolation*
  - Averaging prices before and after leads to bias

Solutions to bid-ask bounce type noise

- Bid-ask bounce is a critical issue
  - Simple model with "pure" noise

$$p_{i,t} = p^*_{i,t} + \nu_{i,t}$$

    - $p_{i,t}$ is the observed price with noise
    - $p^*_{i,t}$ is the unobserved efficient price
    - $\nu_{i,t}$ is the noise
  - Easy to show

$$r_{i,t} = r^*_{i,t} + \eta_{i,t}$$

    - $r^*_{i,t}$ is the unobserved efficient return
    - $\eta_{i,t} = \nu_{i,t} - \nu_{i-1,t}$ is a MA(1) error
  - RV is badly biased

$$RV^{(m)}_t \approx \widehat{RV}_t + m\tau^2$$

    - Bias is increasing in m
    - Variance is also increasing in m

Simple solution

- Do not sample frequently
  - 5-30 minutes
    - Better than daily but still inefficient
  - Remove MA(1) by filtering
    - $\eta_{i,t}$ is an MA(1)
    - Fit an MA(1) to observed returns

$$r_{i,t} = \theta\epsilon_{i-1,t} + \epsilon_{i,t}$$

    - Use fit residuals $\hat{\epsilon}_{i,t}$ to compute RV
    - Generally biased downward
  - Use mid-quotes
    - A little noise
    - My usual solution

A modified Realized Variance estimator: RV^{AC1}

- Best solution is to use a modified RV estimator
  - RV^{AC1}

$$RV_t^{AC1(m)} = \sum_{i=1}^{m} r_{i,t}^2 + 2\sum_{i=2}^{m} r_{i,t} r_{i-1,t}$$

  - Adds a term to RV to capture the MA(1) noise
  - Looks like a simple Newey-West estimator
  - Unbiased in pure noise model
  - Not consistent
  - Realized Kernel Estimator
    - Adds more weighted cross-products
    - Consistent in the presence of many realistic noise processes
    - Fairly easy to implement

One final problem

- Market closure
  - Markets do not operate 24 hours a day (in general)
  - Add in close-to-open return squared

$$RV_t^{(m)} = r_{\text{CtO},t}^2 + \sum_{i=1}^{m} r_{i,t}^2$$

    - $r_{\text{CtO},t} = p_{\text{Open},t} - p_{\text{Close},t-1}$
  - Compute a modified RV by weighting the overnight and open hour estimates differently

$$\widetilde{RV}_t^{(m)} = \lambda_1 r_{\text{CtO},t}^2 + \lambda_2 RV_t^{(m)}$$

The volatility signature plot

- Hard to know how often to sample
  - Visual inspection may be useful

Definition (Volatility Signature Plot)

The volatility signature plot displays the time-series average of Realized Variance

$$\overline{RV}_t^{(m)} = T^{-1} \sum_{t=1}^{T} RV_t^{(m)}$$

as a function of the number of samples, m. An equivalent representation displays the amount of time, whether in calendar time or tick time (number of trades between observations) along the X-axis.

Review

Key Concepts

Realized Variance, RV^{AC1}, Volatility Signature Plot, Bid-Ask Bounce

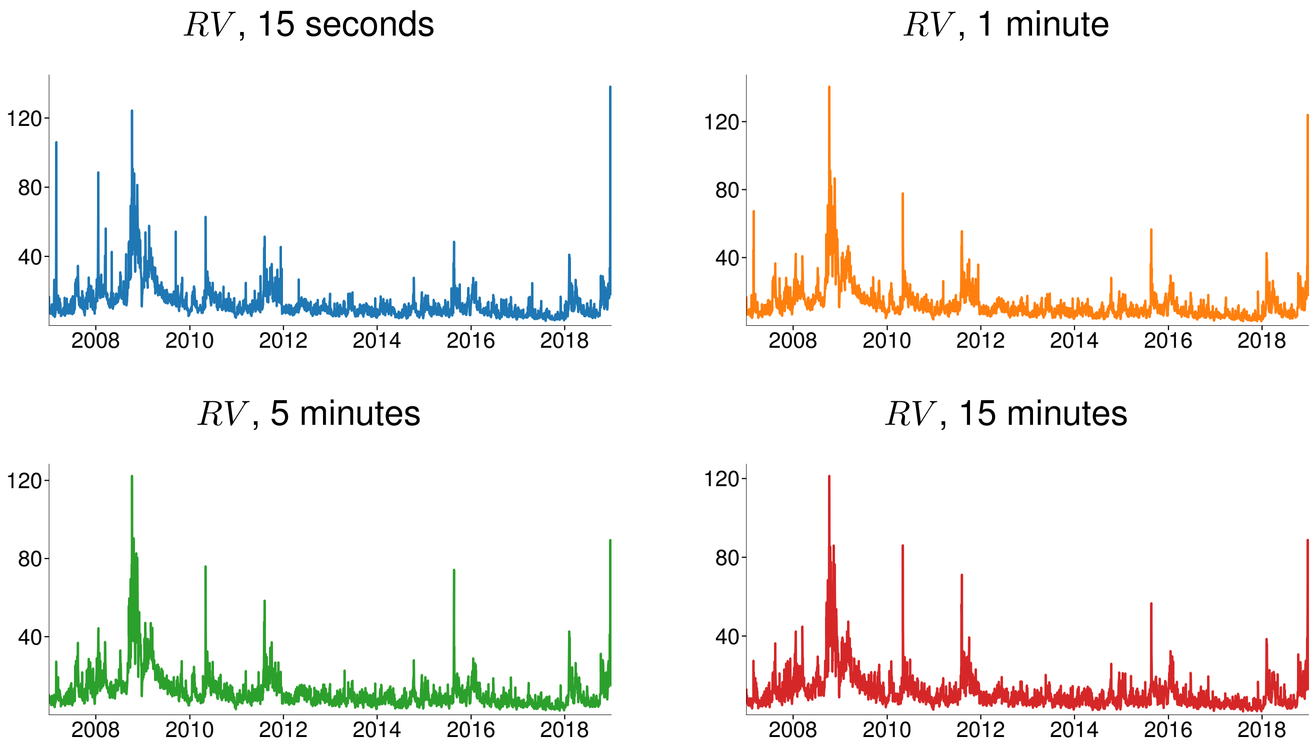
Questions

- What does RV estimate?
- What are the key issues in real data that prevent the literal application of RV to tick data?
- How can RV be modified to account for closed periods even if prices change during these periods?
- How is the volatility signature plot used?

Some empirical results

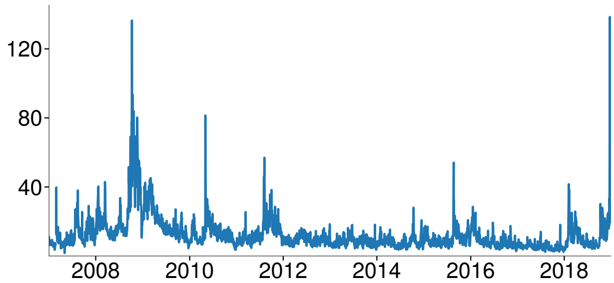
- S&P 500 Depository Receipts
  - SPiDeRs
  - AMEX: SPY
  - Exchange Traded Fund
  - Ultra-liquid
    - 100M shares per day
    - Over 100,000 trades per day
    - 23,400 seconds in a typical trading day
  - January 1, 2007 – December 31, 2018
  - Filtered by daily High-Low data
  - Some cleaning of outliers

SPDR Realized Variance (RV)

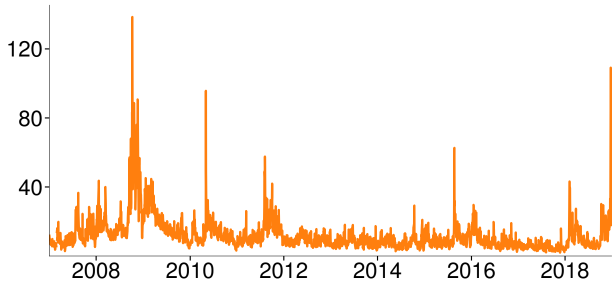


SPDR Realized Variance (RV^{AC1})

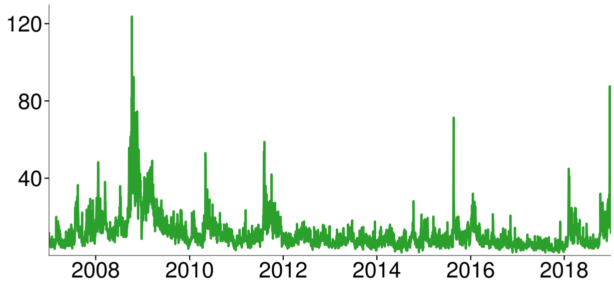
RV^{AC1}, 15 seconds



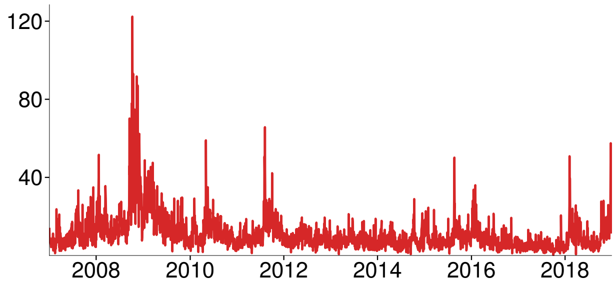
RV^{AC1}, 1 minute



RV^{AC1}, 5 minutes

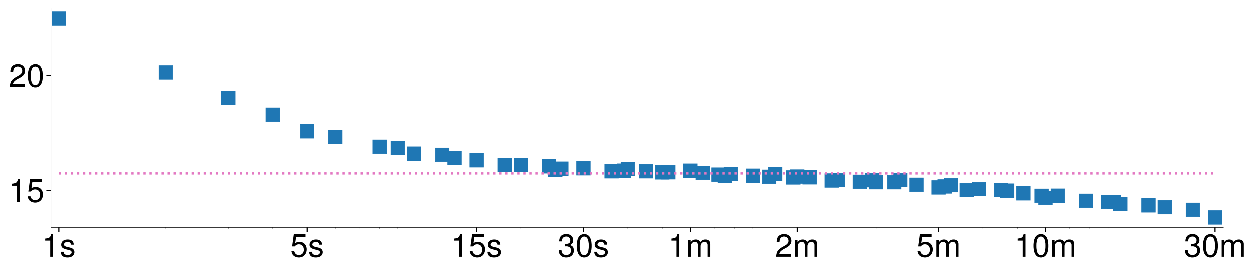


RV^{AC1}, 15 minutes

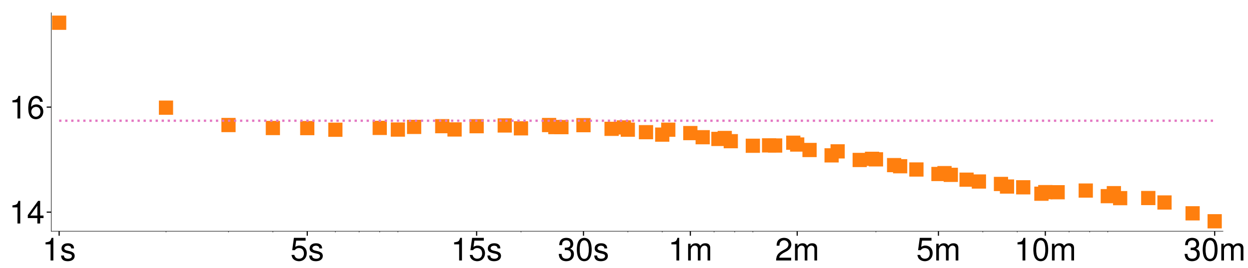


Volatility Signature Plots

Volatility Signature Plot for SPDR RV

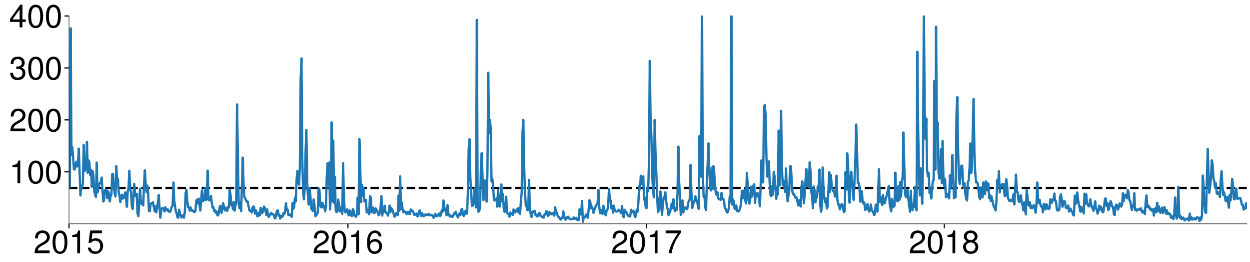


Volatility Signature Plot for SPDR RV^{AC1}

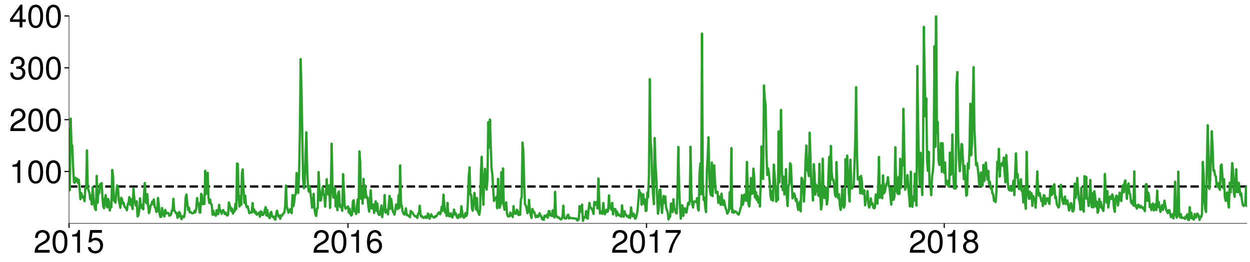


Bitcoin Realized Variance

5-second RV



5-minute RV



Modeling Realized Variance

- Two choices
- Treat volatility as observable and model as ARMA
  - Really simply to do
  - Forecasts are equally simple
  - Theoretical motivation why RV may be well modeled by an ARMA(P,1)
- Continue to treat volatility as latent and use ARCH-type model
  - Realized Variance is still measured with error
  - A more precise measure of conditional variance that daily returns squared, r_t^2, but otherwise similar

Treating σ_t^2 as observable

- *If* RV is σ_t^2,then variance is observable
- Main model used is a Heterogeneous Autoregression
- Restricted AR(22) in levels

$$RV_t = \phi_0 + \phi_1 RV_{t-1} + \phi_5 \overline{RV}_{5,t-1} + \phi_{22} \overline{RV}_{22,t-1} + \epsilon_t$$

- Or in logs

$$\ln RV_t = \phi_0 + \phi_1 \ln RV_{t-1} + \phi_5 \ln \overline{RV}_{5,t-1} + \phi_{22} \ln \overline{RV}_{22,t-1} + \epsilon_t$$

  where $\overline{RV}_{j,t-1} = j^{-1} \sum_{i=1}^{j} RV_{t-i}$ is a j lag moving average
- Model picks up volatility changes at the daily, weekly, and monthly scale
- Fits and forecasts RV fairly well
  - MA term may still be needed

Leaving σ_t^2 latent

- Alternative if to treat RV as a proxy of the latent variance and use a *non-negative multiplicative error model* (MEM)
- MEMs specify the mean of a process as $\mu_t \times \psi_t$ where ψ_t is a mean *1* shock.
- A χ_1^2 is a natural choice here
- ARCH models are special cases of a non-negative MEM model
- Easy to model RV using existing ARCH models
  1. Construct $\tilde{r}_t = \text{sign}(r_t)\sqrt{RV_t}$
  2. Use standard ARCH model building to construct a model for $\tilde{r}_t$

$$\sigma_t^2 = \omega + \alpha_1 \tilde{r}_{t-1}^2 + \gamma_1 \tilde{r}_{t-1}^2 I_{[\tilde{r}_{t-1}<0]} + \beta_1 \sigma_{t-1}^2$$

  becomes

$$\sigma_t^2 = \omega + \alpha_1 RV_{t-1} + \gamma_1 RV_{t-1} I_{[r_{t-1}<0]} + \beta_1 \sigma_{t-1}^2$$

Review

Key Concepts

Heterogeneous Autoregression, Multiplicative Error Model

Questions

- How is a HAR related to an AR?
- What feature does the lag structure of a HAR capture?
- How are forecasts of $\ln RV$ transformed into forecasts of RV?
- What transformation is used to model RV as a MEM (ARCH-type model)?

Implied Volatility and VIX

- Implied volatility is very different from ARCH and Realized measures
- Market based: Level of volatility is calculated from options prices
- Forward looking: Options depend on future price path
- "Classic" implied relies on the Black-Scholes pricing formula
- "Model free" implied volatility exploits a relationship between the second derivative of the call price with respect to the strike and the risk neutral measure
- VIX is a Chicago Board Options Exchange (CBOE) index based on a model free measure
- Allows volatility to be directly traded

Black-Scholes Implied Volatility

- Black-Scholes Options Pricing
- Prices follow a geometric Brownian Motion

$$\mathrm{d}S_t = \mu S_t \mathrm{d}t + \sigma S_t \mathrm{d}W_t$$

- Constant drift and volatility
- Price of a call is

$$C(T,K) = S\Phi(d_1) + Ke^{-rT}\Phi(d_2)$$

where

$$d_1 = \frac{\ln(S/K) + \left(r + \sigma^2/2\right)T}{\sigma\sqrt{T}}$$
$$d_2 = \frac{\ln(S/K) + \left(r - \sigma^2/2\right)T}{\sigma\sqrt{T}}.$$

- Can invert to produce a formula for the volatility given the call price $C(T,K)$

$$\sigma_t^{\text{Implied}} = g\left(C_t(T,K), S_t, K, T, r\right)$$

Model Free Implied Volatility

- Model free uses the relationship between option prices and RN density
- The price of a call option with strike K and maturity t is

$$C(t,K) = \int_K^\infty (S_t - K)\,\phi_t(S_t)\,dS_t$$

- $\phi_t(S_t)$ is the *risk-neutral* density at maturity t
- Differentiating with respect to strike yields

$$\frac{\partial C(t,K)}{\partial K} = -\int_K^\infty \phi_t(S_t)\,dS_t$$

- Differentiating again with respect to strike yields

$$\frac{\partial^2 C(t,K)}{\partial K^2} = \phi_t(K)$$

- The change in an option price as a function of the strike K is the probability of the stock price having value K at time t
- Allows for risk-neutral density to be recovered from a continuum of options *without assuming a model for stock prices*

Model Free Implied Volatility

- The previous result allows a model free IV to be computed from

$$\mathrm{E}_{\mathbb{F}}\left[\int_0^t \left(\frac{\partial F_s}{F_s}\right)^2 ds\right] = 2\int_0^\infty \frac{C^F(t,K)-(F_0-K)^+}{K^2}\mathrm{d}K = 2\int_0^\infty \underbrace{\frac{C^F(t,K)-(F_0-K)^+}{K}}_{\text{Height}}\underbrace{\frac{\mathrm{d}K}{K}}_{\text{Width}}$$

- Devil is in the details
  - Only finitely many calls
  - Thin trading
  - Truncation

$$\sum_{m=1}^{M}\left[g(T,K_m)+g(T,K_{m-1})\right](K_m-K_{m-1})$$

where

$$g(T,K)=\frac{C(t,K/B(0,t))-(S_0-K)^+}{K^2}$$

- See Jiang & Tian (2005, *RFS*) for a very useful discussion

VIX

- VIX is continuously computed by the CBOE
- Uses a model-free style formula
- Uses both calls and puts
- Focuses on out-of-the-money options
  - OOM options are more liquid
- Formula:

$$\sigma^2=\frac{2}{T}e^{rT}\sum_{i=1}^{N}\underbrace{\frac{Q(K_i)}{K_i}}_{\text{Height}}\underbrace{\frac{\Delta K_i}{K_i}}_{\text{Width}}-\frac{1}{T}\left(\frac{F_0}{K_0}-1\right)^2$$

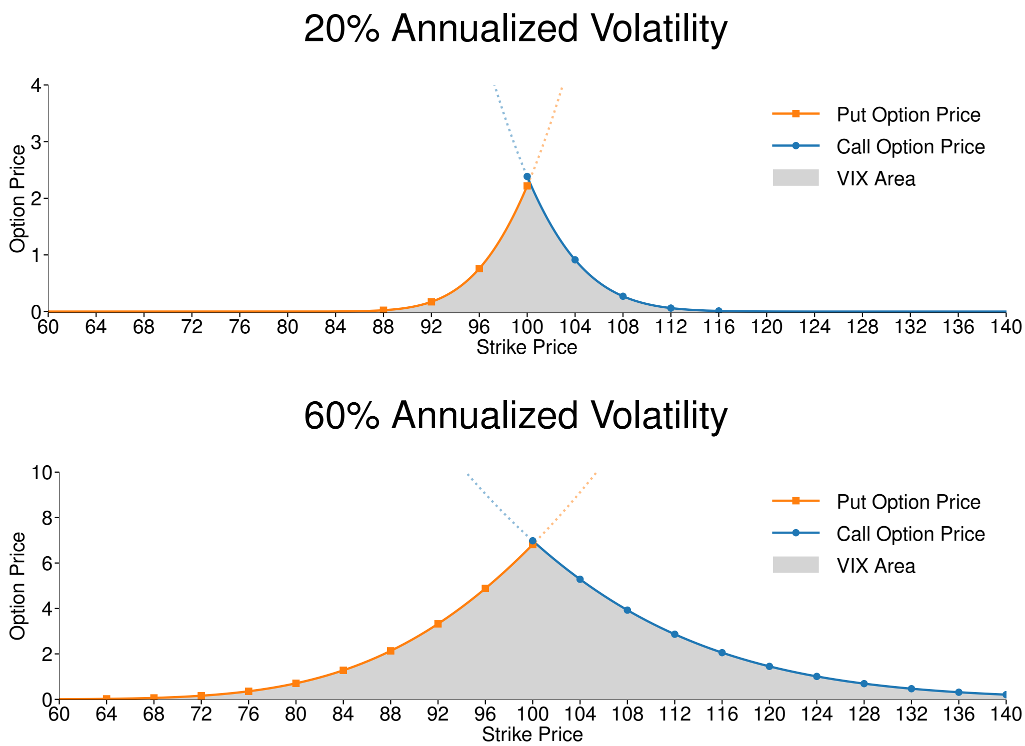
  - $Q(K_i)$ is the mid-quote for a strike of K_i, K_0 is the first strike below the forward index level
  - Only uses out-of-the-money options
  - VIX appears to have information about future *realized* volatility that is not in other backward looking measures (GARCH/RV)

Model-Free Example

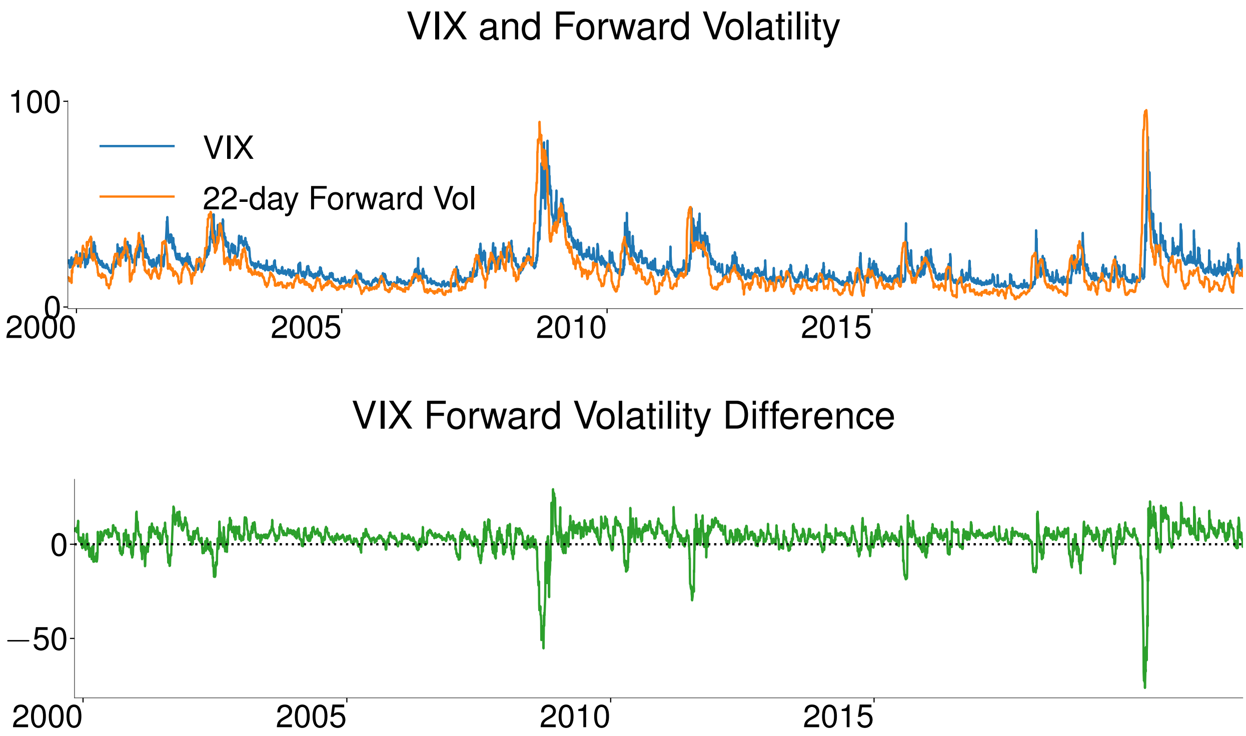
- MFIV works under weak conditions on the underlying price process
  - Geometric Brownian motion is included
- Put and call options prices computed from Black-Scholes
  - Annualized volatility either 20% or 60%
  - Risk-free rate 2%, time-to-maturity 1 month $(T = {}^{1}/{12})$
  - Current price 100 (normalized to moneyness), strikes every 4%
- Contribution is $\frac{2}{T}e^{rT}\frac{\Delta K_i}{K_i^2}Q(K_i)$

Strike	Call	Put	Abs. Diff.	VIX Contrib.
88	12.17	0.02	12.15	0.0002483
92	8.33	0.17	8.15	0.0019314
96	4.92	0.76	4.16	0.0079299
100	2.39	2.22	0.17	0.0221168
104	0.91	4.74	3.83	0.0080904
108	0.27	8.09	7.82	0.0022259
112	0.06	11.88	11.81	0.0004599
116	0.01	15.82	15.81	7.146e-05
Total				0.0430742

Model-Free Example



VIX against TARCH(1,1,1) Forward-vol



Variance Risk Premium

- Difference between VIX and forward volatility is a measure of the return to selling volatility
- Variance Risk Premium is strictly forward looking

$$\mathrm{E}_t^{\mathbb{Q}}\left[\int_0^{t+h}\left(\frac{\partial F_s}{F_s}\right)^2 ds\right]-\mathrm{E}_t^{\mathbb{P}}\left[\int_t^{t+h}\left(\frac{\partial F_s}{F_s}\right)^2 ds\right]$$

- Defined as the difference between RN $\left(\mathrm{E}^{\mathbb{Q}}\right)$ and physical $\left(\mathrm{E}^{\mathbb{P}}\right)$ variance
  - RN variance measured using VIX or other MFIV
  - Physical forecast from HAR or other model based on Realized Variance
    - RV matters, using daily is sufficiently noisy that prediction is not useful

Review

Key Concepts

Black-Scholes Implied Volatility, Model-free Implied Volatility, Variance Risk Premium

Questions

- Why does BSIV curves smile or smirk? What would generate the difference between the two shapes?
- How is MFIV computed, and how does this differ from BSIV?
- What determines the variance risk premium?