Chapter 7

Univariate Volatility Modeling

Note: The primary references for these notes are chapters 10 and 11 in Taylor (2005). Alternative, but less comprehensive, treatments can be found in chapter 21 of Hamilton (1994) or chapter 4 of Enders (2004). Many of the original articles can be found in R. F. Engle (1995).

R. F. Engle (1982) introduced the ARCH model and, in doing so, modern financial econometrics. Since then, measuring and modeling conditional volatility has become the cornerstone of the field. Models used for analyzing conditional volatility can be extended to capture a variety of related phenomena including duration analysis, Value-at-Risk, Expected Shortfall and density forecasting. This chapter begins by examining the meaning of "volatility" - it has many - before turning attention to the ARCH-family of models. The chapter proceeds through estimation, inference, model selection, forecasting and diagnostic testing. The chapter concludes by covering a relatively new method for measuring volatility using ultra-high-frequency data, *realized volatility*, and a market-based measure of volatility, *implied volatility*.

Volatility measurement and modeling is the foundation of financial econometrics. This chapter begins by introducing volatility as a meaningful concept and then describes a widely employed framework for volatility analysis: the ARCH model. The chapter describes the important members of the ARCH family, some of their properties, estimation, inference and model selection. Attention then turns to a new tool in the measurement and modeling of financial econometrics, *realized volatility*, before concluding with a discussion of option-based implied volatility.

7.1 Why does volatility change?

Time-varying volatility is a pervasive empirical regularity of financial time series, so much so that it is difficult to find an asset return series which does *not* exhibit time-varying volatility. This chapter focuses on providing a statistical description of the time-variation of volatility, but does not go into depth on *why* volatility varies over time. A number of explanations have been proffered to explain this phenomenon, although treated individually, none are completely satisfactory.

- *News Announcements*: The arrival of unanticipated news (or "news surprises") forces agents to update beliefs. These new beliefs trigger portfolio rebalancing and high periods of volatility correspond to agents dynamically solving for new asset prices. While certain classes of assets have been shown to react to surprises, in particular government bonds and foreign exchange, many appear to be unaffected by even large surprises (see, *inter alia* R. F. Engle and Li (1998) and Andersen, Bollerslev, Diebold, and Vega (2007)). Additionally, news-induced periods of high volatility are generally short, often on the magnitude of 5 to 30-minutes and the apparent resolution of uncertainty is far too quick to explain the time-variation of volatility seen in asset prices.

- *Leverage*: When firms are financed using both debt and equity, only the equity will reflect the volatility of the firms cash flows. However, as the price of equity falls, a smaller quantity must reflect the same volatility of the firm's cash flows and so negative returns should lead to increases in equity volatility. The leverage effect is pervasive in equity returns, especially in broad equity indices, although alone it is insufficient to explain the time variation of volatility (Christie, 1982; Bekaert and Wu, 2000).

- *Volatility Feedback*: Volatility feedback is motivated by a model where the volatility of an asset is priced. When the price of an asset falls, the volatility must increase to reflect the increased expected return (in the future) of this asset, and an increase in volatility requires an even lower price to generate a sufficient return to compensate an investor for holding a volatile asset. There is evidence that this explanation is empirically supported although it cannot explain the totality of the time-variation of volatility (Bekaert and Wu, 2000).

- *Illiquidity*: Short run spells of illiquidity may produce time varying volatility even when shocks are i.i.d. Intuitively, if the market is oversold (bought), a small negative (positive) shock will cause a small decrease (increase) in demand. However, since there are few participants willing to buy (sell), this shock has a large effect on prices. Liquidity runs tend to last from 20 minutes to a few days and cannot explain the long cycles in present volatility.

- *State Uncertainty*: Asset prices are important instruments that allow agents to express beliefs about the state of the economy. When the state is uncertain, slight changes in beliefs may cause large shifts in portfolio holdings which in turn feedback into beliefs about the state. This feedback loop can generate time-varying volatility and should have the largest effect when the economy is transitioning between periods of growth and contraction.

The actual cause of the time-variation in volatility is likely a combination of these and some not present.

7.1.1 What is volatility?

Volatility comes in many shapes and forms. To be precise when discussing volatility, it is important to be clear what is meant when the term "volatility" used.

Volatility Volatility is simply the standard deviation. Volatility is often preferred to variance as it is measured in the same *units* as the original data. For example, when using returns, the volatility is also in returns, and a volatility of 5% indicates that $\pm$5% is a meaningful quantity.

Realized Volatility Realized volatility has historically been used to denote a measure of the volatility over some arbitrary period of time,

$$\hat{\sigma} = \sqrt{T^{-1}\sum_{t=1}^{T}(r_t - \hat{\mu})^2} \tag{7.1}$$

but is now used to describe a volatility measure constructed using ultra-high-frequency (UHF) data (also known as tick data). See section 7.4 for details.

Conditional Volatility Conditional volatility is the expected volatility at some future time $t+h$ based on all available information up to time t ($\mathcal{F}_t$). The one-period ahead conditional volatility is denoted $E_t[\sigma_{t+1}]$.

Implied Volatility Implied volatility is the volatility that will correctly price an option. The Black-Scholes pricing formula relates the price of a European call option to the current price of the underlying, the strike, the risk-free rate, the time-to-maturity and the *volatility*,

$$BS(S_t, K, r, t, \sigma_t) = C_t \tag{7.2}$$

where C is the price of the call. The implied volatility is the value which solves the Black-Scholes taking the option and underlying prices, the strike, the risk-free and the time-to-maturity as given,

$$\hat{\sigma}_t(S_t, K, r, t, C) \tag{7.3}$$

Annualized Volatility When volatility is measured over an interval other than a year, such as a day, week or month, it can always be scaled to reflect the volatility of the asset over a year. For example, if σ denotes the daily volatility of an asset and there are 252 trading days in a year, the annualized volatility is $\sqrt{252}\sigma$. Annualized volatility is a useful measure that removes the sampling interval from reported volatilities.

Variance All of the uses of volatility can be replaced with variance and most of this chapter is dedicated to modeling *conditional variance*, denoted $\mathrm{E}_t[\sigma^2_{t+h}]$.

7.2 ARCH Models

In financial econometrics, an arch is not an architectural feature of a building; it is a fundamental tool for analyzing the time-variation of conditional variance. The success of the *ARCH*

(*Auto*Regressive *C*onditional *H*eteroskedasticity) family of models can be attributed to three features: ARCH processes are essentially ARMA models and many of the tools of linear time-series analysis can be directly applied, ARCH-family models are easy to estimate and many parsimonious models are capable of providing good descriptions of the dynamics of asset volatility.

7.2.1 The ARCH model

The complete ARCH(P) model (R. F. Engle, 1982) relates the current level of volatility to the past P squared shocks.

Definition 7.1 (P$^{\text{th}}$ Order Autoregressive Conditional Heteroskedasticity (ARCH))**.** A P$^{\text{th}}$ order ARCH process is given by

$$\begin{aligned} r_t &= \mu_t + \epsilon_t \\ \sigma_t^2 &= \omega + \alpha_1 \epsilon_{t-1}^2 + \alpha_2 \epsilon_{t-2}^2 + \ldots + \alpha_P \epsilon_{t-P}^2 \\ \epsilon_t &= \sigma_t e_t \\ e_t &\overset{\text{i.i.d.}}{\sim} N(0,1). \end{aligned} \tag{7.4}$$

where μ_t can be any adapted model for the conditional mean.[1]

The key feature of this model is that the variance of the shock, ϵ_t, is time varying and depends on the past P shocks, $\epsilon_{t-1}, \epsilon_{t-2}, \ldots, \epsilon_{t-P}$ through their squares. σ_t^2 is the time $t-1$ conditional variance and it is in the time $t-1$ information set $\mathcal{F}_{t-1}$. This can be verified by noting that all of the right-hand side variables that determine σ_t^2 are known at time $t-1$. The model for the conditional mean can include own lags, shocks (in a MA model) or exogenous variables such as the default spread or term premium. In practice, the model for the conditional mean should be general enough to capture the dynamics present in the data. For many financial time series, particularly when measured over short intervals - one day to one week - a constant mean, sometimes assumed to be 0, is often sufficient.

An alternative way to describe an ARCH(P) model is

$$\begin{aligned} r_t | \mathcal{F}_{t-1} &\sim N(\mu_t, \sigma_t^2) \\ \sigma_t^2 &= \omega + \alpha_1 \epsilon_{t-1}^2 + \alpha_2 \epsilon_{t-2}^2 + \ldots + \alpha_P \epsilon_{t-P}^2 \\ \epsilon_t &= r_t - \mu_t \end{aligned} \tag{7.5}$$

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

[1] A model is adapted if everything required to model the mean at time t is known at time $t-1$. Standard examples of adapted mean processes include a constant mean or anything in the family of ARMA processes or any exogenous regressors known at time $t-1$.

[2] It is implausible that the unconditional mean return of a risky asset to be zero. However, when using daily equity

The conditional variance, denoted σ_t^2, is

$$\mathrm{E}_{t-1}\left[\epsilon_t^2\right] = \mathrm{E}_{t-1}\left[e_t^2\sigma_t^2\right] = \sigma_t^2\mathrm{E}_{t-1}\left[e_t^2\right] = \sigma_t^2 \tag{7.6}$$

and the *un*conditional variance, denoted $\bar{\sigma}^2$, is

$$\mathrm{E}\left[\epsilon_{t+1}^2\right] = \bar{\sigma}^2. \tag{7.7}$$

The first interesting property of the ARCH(P) model is the unconditional variance. Assuming the unconditional variance exists, $\bar{\sigma}^2 = \mathrm{E}[\sigma_t^2]$ can be derived from

$$\begin{aligned}
\mathrm{E}\left[\sigma_t^2\right] &= \mathrm{E}\left[\omega + \alpha_1\epsilon_{t-1}^2 + \alpha_2\epsilon_{t-2}^2 + \ldots + \alpha_P\epsilon_{t-P}^2\right] \qquad (7.8)\\
&= \omega + \alpha_1\mathrm{E}\left[\epsilon_{t-1}^2\right] + \alpha_2\mathrm{E}\left[\epsilon_{t-2}^2\right] + \ldots + \alpha_P\mathrm{E}\left[\epsilon_{t-P}^2\right]\\
&= \omega + \alpha_1\mathrm{E}\left[\sigma_{t-1}^2\right]\mathrm{E}\left[e_{t-1}^2\right] +\\
&\qquad \alpha_2\mathrm{E}\left[\sigma_{t-2}^2\right]\mathrm{E}\left[e_{t-2}^2\right] + \ldots + \alpha_P\mathrm{E}\left[\sigma_{t-P}^2\right]\mathrm{E}\left[e_{t-P}^2\right]\\
&= \omega + \alpha_1\mathrm{E}\left[\sigma_{t-1}^2\right] + \alpha_2\mathrm{E}\left[\sigma_{t-2}^2\right] + \ldots + \alpha_P\mathrm{E}\left[\sigma_{t-P}^2\right]\\
\mathrm{E}\left[\sigma_t^2\right] - \alpha_1\mathrm{E}\left[\sigma_{t-1}^2\right] - \ldots - \alpha_P\mathrm{E}\left[\sigma_{t-P}^2\right] &= \omega\\
\mathrm{E}\left[\sigma_t^2\right](1 - \alpha_1 - \alpha_2 - \ldots - \alpha_P) &= \omega\\
\bar{\sigma}^2 &= \frac{\omega}{1 - \alpha_1 - \alpha_2 - \ldots - \alpha_P}.
\end{aligned}$$

This derivation makes use of a number of properties of ARCH family models. First, the definition of the shock $\epsilon_t^2 \equiv e_t^2\sigma_t^2$ is used to separate the i.i.d. normal innovation (e_t) from the conditional variance (σ_t^2). e_t and σ_t^2 are independent since σ_t^2 depends on $\epsilon_{t-1}, \epsilon_{t-2}, \ldots, \epsilon_{t-P}$ (and in turn $e_{t-1}, e_{t-2}, \ldots, e_{t-P}$) while e_t is an i.i.d. draw at time t. Using these two properties, the derivation follows by noting that the unconditional expectation of σ_{t-j}^2 is the same in any time period ($\mathrm{E}[\sigma_t^2] = \mathrm{E}[\sigma_{t-p}^2] = \bar{\sigma}^2$) and is assumed to exist. Inspection of the final line in the derivation reveals the condition needed to ensure that the unconditional expectation is finite: $1 - \alpha_1 - \alpha_2 - \ldots - \alpha_P > 0$. As was the case in an AR model, as the persistence (as measured by $\alpha_1, \alpha_2, \ldots$) increases towards a unit root, the process explodes.

7.2.1.1 Stationarity

An ARCH(P) model is covariance stationary as long as the model for the conditional mean corresponds to a stationary process[3] and $1 - \alpha_1 - \alpha_2 - \ldots - \alpha_P > 0$.[4] ARCH models have the property that $\mathrm{E}[\epsilon_t^2] = \bar{\sigma}^2 = \omega/(1 - \alpha_1 - \alpha_2 - \ldots - \alpha_P)$ since

data, the squared mean is typically less than 1% of the variance and there are few ramifications for setting this value to 0 ($\frac{\mu^2}{\sigma^2} < 0.01$). Other assets, such as electricity prices, have non-trivial predictability and so an appropriate model for the conditional mean should be specified.

[3] For example, a constant or a covariance stationary ARMA process.

[4] When $\sum_{i=1}^{P}\alpha_i > 1$, and ARCH(P) may still be strictly stationary although it cannot be covariance stationary since it has infinite variance.

$$\mathrm{E}[\epsilon_t^2] = \mathrm{E}[e_t^2\sigma_t^2] = \mathrm{E}[e_t^2]\mathrm{E}[\sigma_t^2] = 1 \cdot \mathrm{E}[\sigma_t^2] = \mathrm{E}[\sigma_t^2]. \tag{7.9}$$

which exploits the independence of e_t from σ_t^2 and the assumption that e_t is a mean zero process with unit variance and so $\mathrm{E}[e_t^2] = 1$.

One crucial requirement of any covariance stationary ARCH process is that the parameters of the variance evolution, $\alpha_1, \alpha_2, \ldots, \alpha_P$ must all be positive.[5] The intuition behind this requirement is that if one of the αs were negative, eventually a shock would be sufficiently large to produce a negative conditional variance, an undesirable feature. Finally, it is also necessary that $\omega > 0$ to ensure covariance stationarity.

To aid in developing intuition about ARCH-family models consider a simple ARCH(1) with a constant mean of 0,

$$\begin{aligned} r_t &= \epsilon_t \\ \sigma_t^2 &= \omega + \alpha_1 \epsilon_{t-1}^2 \\ \epsilon_t &= \sigma_t e_t \\ e_t &\overset{\text{i.i.d.}}{\sim} N(0,1). \end{aligned} \tag{7.10}$$

While the conditional variance in an ARCH process appears different from anything previously encountered, it can be equivalently expressed as an AR(1) for ϵ_t^2. This transformation allows many properties of ARCH residuals to be directly derived by applying the results of chapter 2. By adding $\epsilon_t^2 - \sigma_t^2$ to both sides of the volatility equation,

$$\begin{aligned} \sigma_t^2 &= \omega + \alpha_1 \epsilon_{t-1}^2 \\ \sigma_t^2 + \epsilon_t^2 - \sigma_t^2 &= \omega + \alpha_1 \epsilon_{t-1}^2 + \epsilon_t^2 - \sigma_t^2 \\ \epsilon_t^2 &= \omega + \alpha_1 \epsilon_{t-1}^2 + \epsilon_t^2 - \sigma_t^2 \\ \epsilon_t^2 &= \omega + \alpha_1 \epsilon_{t-1}^2 + \sigma_t^2\left(e_t^2 - 1\right) \\ \epsilon_t^2 &= \omega + \alpha_1 \epsilon_{t-1}^2 + \nu_t, \end{aligned} \tag{7.11}$$

an ARCH(1) process can be shown to be an AR(1). The error term, ν_t represents the volatility *surprise*, $\epsilon_t^2 - \sigma_t^2$, which can be decomposed as $\sigma_t^2(e_t^2 - 1)$ and is a mean 0 white noise process since e_t is i.i.d. and $\mathrm{E}[e_t^2] = 1$. Using the AR representation, the autocovariances of ϵ_t^2 are simple to derive. First note that $\epsilon_t^2 - \bar{\sigma}^2 = \sum_{i=0}^{\infty} \alpha_1^i \nu_{t-i}$. The first autocovariance can be expressed

[5]Since each $\alpha_j \geq 0$, the roots of the characteristic polynomial associated with $\alpha_1, \alpha_2, \ldots, \alpha_p$ will be less than 1 if and only if $\sum_{p=1}^{P} \alpha_p < 1$.

$$
\begin{aligned}
\mathrm{E}\left[\left(\epsilon_t^2-\bar{\sigma}^2\right)\left(\epsilon_{t-1}^2-\bar{\sigma}^2\right)\right] &= \mathrm{E}\left[\left(\sum_{i=0}^{\infty}\alpha_1^i\nu_{t-i}\right)\left(\sum_{j=1}^{\infty}\alpha_1^{j-1}\nu_{j-i}\right)\right] \\
&= \mathrm{E}\left[\left(\nu_t+\sum_{i=1}^{\infty}\alpha_1^i\nu_{t-i}\right)\left(\sum_{j=1}^{\infty}\alpha_1^{j-1}\nu_{t-j}\right)\right] \\
&= \mathrm{E}\left[\left(\nu_t+\alpha_1\sum_{i=1}^{\infty}\alpha_1^{i-1}\nu_{t-i}\right)\left(\sum_{j=1}^{\infty}\alpha_1^{j-1}\nu_{t-j}\right)\right] \\
&= \mathrm{E}\left[\nu_t\left(\sum_{i=1}^{\infty}\alpha_1^{i-1}\nu_{t-i}\right)\right]+\mathrm{E}\left[\alpha_1\left(\sum_{i=1}^{\infty}\alpha_1^{i-1}\nu_{t-i}\right)\left(\sum_{j=1}^{\infty}\alpha_1^{j-1}\nu_{t-j}\right)\right] \\
&= \sum_{i=1}^{\infty}\alpha_1^{i-1}\mathrm{E}\left[\nu_t\nu_{t-i}\right]+\mathrm{E}\left[\alpha_1\left(\sum_{i=1}^{\infty}\alpha_1^{i-1}\nu_{t-i}\right)\left(\sum_{j=1}^{\infty}\alpha_1^{j-1}\nu_{t-j}\right)\right] \\
&= \sum_{i=1}^{\infty}\alpha_1^{i-1}\cdot 0+\mathrm{E}\left[\alpha_1\left(\sum_{i=1}^{\infty}\alpha_1^{i-1}\nu_{t-i}\right)\left(\sum_{j=1}^{\infty}\alpha_1^{j-1}\nu_{t-j}\right)\right] \\
&= \alpha_1\mathrm{E}\left[\left(\sum_{i=1}^{\infty}\alpha_1^{i-1}\nu_{t-i}\right)^2\right] \\
&= \alpha_1\mathrm{E}\left[\left(\sum_{i=0}^{\infty}\alpha_1^{i}\nu_{t-1-i}\right)^2\right] \\
&= \alpha_1\left(\sum_{i=0}^{\infty}\alpha_1^{2i}\mathrm{E}\left[\nu_{t-1-i}^2\right]+2\sum_{j=0}^{\infty}\sum_{k=j+1}^{\infty}\alpha_1^{jk}\mathrm{E}\left[\nu_{t-1-j}\nu_{t-1-k}\right]\right) \\
&= \alpha_1\left(\sum_{i=0}^{\infty}\alpha_1^{2i}\mathrm{V}\left[\nu_{t-1-i}\right]+2\sum_{j=0}^{\infty}\sum_{k=j+1}^{\infty}\alpha_1^{jk}\cdot 0\right) \\
&= \alpha_1\sum_{i=0}^{\infty}\alpha_1^{2i}\mathrm{V}\left[\nu_{t-1-i}\right] \\
&= \alpha_1\mathrm{V}\left[\epsilon_{t-1}^2\right]
\end{aligned}
\tag{7.12}
$$

where $\mathrm{V}[\epsilon_{t-1}^2]=\mathrm{V}[\epsilon_t^2]$ is the variance of the squared innovations.[6] By repeated substitution, the s$^{\text{th}}$ autocovariance, $\mathrm{E}[(\epsilon_t^2-\bar{\sigma}^2)(\epsilon_{t-s}^2-\bar{\sigma}^2)]$, can be shown to be $\alpha_1^s\mathrm{V}[\epsilon_t^2]$, and so that the autocovariances of an ARCH(1) are identical to those of an AR(1).

[6]For the time being, assume this is finite.

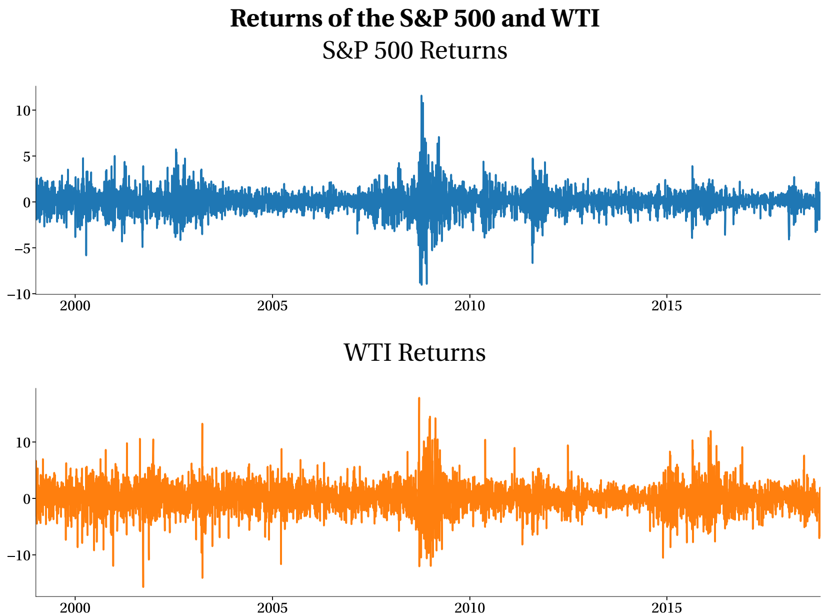

Figure 7.1: Plots of S&P 500 and WTI returns (scaled by 100) from 1997 until 2017. The bulges in the return plots are graphical evidence of time-varying volatility.

7.2.1.2 Autocorrelations

Using the autocovariances, the autocorrelations are

$$\mathrm{Corr}(\epsilon_t^2, \epsilon_{t-s}^2) = \frac{\alpha_1^s \mathrm{V}[\epsilon_t^2]}{\mathrm{V}[\epsilon_t^2]} = \alpha_1^s. \tag{7.13}$$

Further, the relationship between the s[th] autocorrelation of an ARCH process and an AR process holds for ARCH process with other orders. The autocorrelations of an ARCH(P) are identical to those of an AR(P) process with $\{\phi_1, \phi_2, \ldots, \phi_P\} = \{\alpha_1, \alpha_2, \ldots, \alpha_P\}$. One interesting aspect of ARCH(P) processes (and any covariance stationary ARCH-family model) is that the autocorrelations *must* be positive. If one autocorrelation were negative, eventually a shock would be sufficiently large to force the conditional variance negative and so the process would be ill-defined. In practice it is often better to examine the absolute values ($\mathrm{Corr}\left(|\epsilon_t|, |\epsilon_{t-s}|\right)$) rather than the squares since financial returns frequently have outliers that are exacerbated when squared.

7.2.1.3 Kurtosis

The second interesting property of ARCH models is that the kurtosis of shocks (ϵ_t) is strictly greater than the kurtosis of a normal. This may seem strange since all of the shocks $\epsilon_t = \sigma_t e_t$ are normal by assumption. However, an ARCH model is a *variance-mixture* of normals which must produce a kurtosis greater than three. An intuitive proof is simple,

$$\kappa = \frac{\mathrm{E}\left[\epsilon_t^4\right]}{\mathrm{E}\left[\epsilon_t^2\right]^2} = \frac{\mathrm{E}\left[\mathrm{E}_{t-1}\left[\epsilon_t^4\right]\right]}{\mathrm{E}\left[\mathrm{E}_{t-1}\left[e_t^2\sigma_t^2\right]\right]^2} = \frac{\mathrm{E}\left[\mathrm{E}_{t-1}\left[e_t^4\right]\sigma_t^4\right]}{\mathrm{E}\left[\mathrm{E}_{t-1}\left[e_t^2\right]\sigma_t^2\right]^2} = \frac{\mathrm{E}\left[3\sigma_t^4\right]}{\mathrm{E}\left[\sigma_t^2\right]^2} = 3\frac{\mathrm{E}\left[\sigma_t^4\right]}{\mathrm{E}\left[\sigma_t^2\right]^2} \geq 3. \tag{7.14}$$

The key steps in this derivation are that $\epsilon_t^4 = e_t^4\sigma_t^4$ and that $E_t[e_t^4] = 3$ since e_t is a standard normal. The final conclusion that $\mathrm{E}[\sigma_t^4]/\mathrm{E}[\sigma_t^2]^2 > 1$ follows from noting that $\mathrm{V}\left[\epsilon_t^2\right] = \mathrm{E}\left[\epsilon_t^4\right] - \mathrm{E}\left[\epsilon_t^2\right]^2 \geq 0$ and so it must be the case that $\mathrm{E}\left[\epsilon_t^4\right] \geq \mathrm{E}\left[\epsilon_t^2\right]^2$ or $\frac{\mathrm{E}[\epsilon_t^4]}{\mathrm{E}[\epsilon_t^2]^2} \geq 1$. The kurtosis, κ, of an ARCH(1) can be shown to be

$$\kappa = \frac{3(1-\alpha_1^2)}{(1-3\alpha_1^2)} > 3 \tag{7.15}$$

which is greater than 3 since $1-3\alpha_1^2 < 1-\alpha_1^2$ and so $(1-\alpha_1^2)/(1-3\alpha_1^2) > 1$. The formal derivation of the kurtosis is tedious and is presented in Appendix 7.A.

7.2.2 The GARCH model

The ARCH model has been deemed a sufficient contribution to economics to warrant a Nobel prize. Unfortunately, like most models, it has problems. ARCH models typically require 5-8 lags of the squared shock to adequately model conditional variance. The Generalized ARCH (GARCH) process, introduced by Bollerslev (1986), improves the original specification adding lagged conditional variance, which acts as a *smoothing* term. GARCH models typically fit as well as a high-order ARCH yet remain parsimonious.

Definition 7.2 (Generalized Autoregressive Conditional Heteroskedasticity (GARCH) process)**.** A GARCH(P,Q) process is defined as

$$\begin{aligned}
r_t &= \mu_t + \epsilon_t \\
\sigma_t^2 &= \omega + \sum_{p=1}^{P} \alpha_p \epsilon_{t-p}^2 + \sum_{q=1}^{Q} \beta_q \sigma_{t-q}^2 \\
\epsilon_t &= \sigma_t e_t \\
e_t &\overset{\text{i.i.d.}}{\sim} N(0,1)
\end{aligned} \tag{7.16}$$

where μ_t can be any adapted model for the conditional mean.

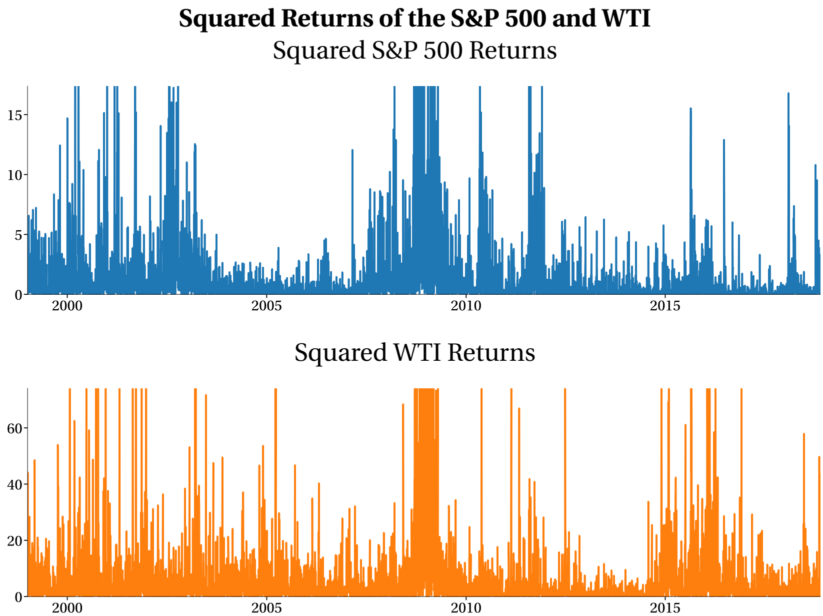

Figure 7.2: Plots of the squared returns of the S&P 500 Index and WTI. Time-variation in the squared returns is evidence of ARCH.

The GARCH(P,Q) model builds on the ARCH(P) model by including Q lags of the conditional variance, $\sigma^2_{t-1}, \sigma^2_{t-2}, \ldots, \sigma^2_{t-Q}$. Rather than focusing on the general specification with all of its complications, consider a simpler GARCH(1,1) model where the conditional mean is assumed to be zero,

$$
\begin{aligned}
r_t &= \epsilon_t \\
\sigma^2_t &= \omega + \alpha_1 \epsilon^2_{t-1} + \beta_1 \sigma^2_{t-1} \\
\epsilon_t &= \sigma_t e_t \\
e_t &\overset{\text{i.i.d.}}{\sim} N(0,1)
\end{aligned}
\tag{7.17}
$$

In this specification, the future variance will be an average of the current shock, ϵ^2_{t-1}, and the current variance, σ^2_{t-1}, plus a constant. The effect of the lagged variance is to produce a model which is actually an ARCH(∞) in disguise. Begin by backward substituting,

$$\sigma^2_t = \omega + \alpha_1 \epsilon^2_{t-1} + \beta_1 \sigma^2_{t-1} \tag{7.18}$$

$$
\begin{aligned}
&= \omega + \alpha_1\epsilon_{t-1}^2 + \beta(\omega + \alpha_1\epsilon_{t-2}^2 + \beta_1\sigma_{t-2}^2) \\
&= \omega + \beta_1\omega + \alpha_1\epsilon_{t-1}^2 + \beta_1\alpha_1\epsilon_{t-2}^2 + \beta_1^2\sigma_{t-2}^2 \\
&= \omega + \beta_1\omega + \alpha_1\epsilon_{t-1}^2 + \beta_1\alpha_1\epsilon_{t-2}^2 + \beta_1^2(\omega + \alpha_1\epsilon_{t-3}^2 + \beta_1\sigma_{t-3}^2) \\
&= \omega + \beta_1\omega + \beta_1^2\omega + \alpha_1\epsilon_{t-1}^2 + \beta_1\alpha_1\epsilon_{t-2}^2 + \beta_1^2\alpha_1\epsilon_{t-3}^2 + \beta_1^3\sigma_{t-3}^2 \\
&= \sum_{i=0}^{\infty}\beta_1^i\omega + \sum_{i=0}^{\infty}\beta_1^i\alpha_1\epsilon_{t-i-1}^2,
\end{aligned}
$$

and the ARCH(∞) representation can be derived.[7] It can be seen that the conditional variance in period t is a constant, $\sum_{i=0}^{\infty}\beta_1^i\omega$, and a weighted average of past squared innovations with weights $\alpha_1, \beta_1\alpha_1, \beta_1^2\alpha_1, \beta_1^3\alpha_1, \ldots$.

As was the case in the ARCH(P) model, the coefficients of a GARCH model must also be restricted to ensure the conditional variances are uniformly positive. In a GARCH(1,1), these restrictions are $\omega > 0$, $\alpha_1 \geq 0$ and $\beta_1 \geq 0$. In a GARCH(P,1) model the restriction change to $\alpha_p \geq 0$, $p = 1, 2, \ldots, P$ with the same restrictions on ω and β_1. However, in a complete GARCH(P,Q) model the parameter restriction are difficult to derive. For example, in a GARCH(2,2), one of the two β's (β_2) can be slightly negative while ensuring that all conditional variances are positive. See D. B. Nelson and Cao (1992) for further details.

As was the case in the ARCH(1) model, the GARCH(1,1) model can be transformed into a standard time series model for ϵ_t^2,

$$
\begin{aligned}
\sigma_t^2 &= \omega + \alpha_1\epsilon_{t-1}^2 + \beta_1\sigma_{t-1}^2 \qquad (7.19)\\
\sigma_t^2 + \epsilon_t^2 - \sigma_t^2 &= \omega + \alpha_1\epsilon_{t-1}^2 + \beta_1\sigma_{t-1}^2 + \epsilon_t^2 - \sigma_t^2 \\
\epsilon_t^2 &= \omega + \alpha_1\epsilon_{t-1}^2 + \beta_1\sigma_{t-1}^2 + \epsilon_t^2 - \sigma_t^2 \\
\epsilon_t^2 &= \omega + \alpha_1\epsilon_{t-1}^2 + \beta_1\sigma_{t-1}^2 - \beta_1\epsilon_{t-1}^2 + \beta_1\epsilon_{t-1}^2 + \epsilon_t^2 - \sigma_t^2 \\
\epsilon_t^2 &= \omega + \alpha_1\epsilon_{t-1}^2 + \beta_1\epsilon_{t-1}^2 - \beta_1(\epsilon_{t-1}^2 - \sigma_{t-1}^2) + \epsilon_t^2 - \sigma_t^2 \\
\epsilon_t^2 &= \omega + \alpha_1\epsilon_{t-1}^2 + \beta_1\epsilon_{t-1}^2 - \beta_1\nu_{t-1} + \nu_t \\
\epsilon_t^2 &= \omega + (\alpha_1 + \beta_1)\epsilon_{t-1}^2 - \beta_1\nu_{t-1} + \nu_t
\end{aligned}
$$

by adding $\epsilon_t^2 - \sigma_t^2$ to both sides. However, unlike the ARCH(1) process which can be transformed into an AR(1), the GARCH(1,1) is transformed into an ARMA(1,1) where $\nu_t = \epsilon_t^2 - \sigma_t^2$ is the volatility surprise. In the general GARCH(P,Q), the ARMA representation takes the form of an ARMA(max(P,Q),Q).

$$
\epsilon_t^2 = \omega + \sum_{i=1}^{\max(P,Q)}(\alpha_i + \beta_i)\epsilon_{t-i}^2 - \sum_{q=1}^{Q}\beta_1\nu_{t-q} + \nu_t \qquad (7.20)
$$

Using the same derivation in the ARCH(1) model, the unconditional variance can be shown

[7] Since the model is assumed to be stationary, it much be the case that $0 \leq \beta < 1$ and so $\lim_{j\rightarrow\infty}\beta^j\sigma_{t-j} = 0$.

to be

$$
\begin{aligned}
E[\sigma_t^2] &= \omega + \alpha_1 E[\epsilon_{t-1}^2] + \beta_1 E[\sigma_{t-1}^2] \\
\bar{\sigma}^2 &= \omega + \alpha_1 \bar{\sigma}^2 + \beta_1 \bar{\sigma}^2 \\
\bar{\sigma}^2 - \alpha_1 \bar{\sigma}^2 - \beta_1 \bar{\sigma}^2 &= \omega \\
\bar{\sigma}^2 &= \frac{\omega}{1 - \alpha_1 - \beta_1}.
\end{aligned} \tag{7.21}
$$

Inspection of the ARMA model leads to an alternative derivation of $\bar{\sigma}^2$ since the AR coefficient is $\alpha_1 + \beta_1$ and the intercept is ω, and the unconditional mean in an ARMA(1,1) is the intercept divided by one minus the AR coefficient, $\omega/(1 - \alpha_1 - \beta_1)$. In a general GARCH(P,Q) the unconditional variance is

$$
\bar{\sigma}^2 = \frac{\omega}{1 - \sum_{p=1}^{P} \alpha_p - \sum_{q=1}^{Q} \beta_q}. \tag{7.22}
$$

As was the case in the ARCH(1) model, the requirements for stationarity are that $1 - \alpha_1 - \beta > 0$ and $\alpha_1 \geq 0$, $\beta_1 \geq 0$ and $\omega > 0$.

The ARMA(1,1) form can be used directly to solve for the autocovariances. Recall the definition of a mean zero ARMA(1,1),

$$
y_t = \phi y_{t-1} + \theta \epsilon_{t-1} + \epsilon_t \tag{7.23}
$$

The 1st autocovariance can be computed as

$$
\begin{aligned}
\mathrm{E}[y_t y_{t-1}] &= \mathrm{E}[(\phi y_{t-1} + \theta \epsilon_{t-1} + \epsilon_t) y_{t-1}] \\
&= \mathrm{E}[\phi y_{t-1}^2] + [\theta \epsilon_{t-1}^2] \\
&= \phi \mathrm{V}[y_{t-1}] + \theta \mathrm{V}[\epsilon_{t-1}] \\
\gamma_1 &= \phi \mathrm{V}[y_{t-1}] + \theta \mathrm{V}[\epsilon_{t-1}]
\end{aligned} \tag{7.24}
$$

and the s$^{\text{th}}$ autocovariance is $\gamma_s = \phi^{s-1}\gamma_1$. In the notation of a GARCH(1,1) model, $\phi = \alpha_1 + \beta_1$, $\theta = -\beta_1$, y_{t-1} is ϵ_{t-1}^2 and η_{t-1} is $\sigma_{t-1}^2 - \epsilon_{t-1}^2$. Thus, $\mathrm{V}[\epsilon_{t-1}^2]$ and $\mathrm{V}[\sigma_t^2 - \epsilon_t^2]$ must be solved for. However, this is tedious and is presented in the appendix. The key to understanding the autocovariance (and autocorrelation) of a GARCH is to use the ARMA mapping. First note that $\mathrm{E}[\sigma_t^2 - \epsilon_t^2] = 0$ so $\mathrm{V}[\sigma_t^2 - \epsilon_t^2]$ is simply $\mathrm{E}[(\sigma_t^2 - \epsilon_t^2)^2]$. This can be expanded to $\mathrm{E}[\epsilon_t^4] - 2\mathrm{E}[\epsilon_t^2 \sigma_t^2] + \mathrm{E}[\sigma_t^4]$ which can be shown to be $2\mathrm{E}[\sigma_t^4]$. The only remaining step is to complete the tedious derivation of the expectation of these fourth powers which is presented in Appendix 7.B.

7.2.2.1 Kurtosis

The kurtosis can be shown to be

$$\kappa = \frac{3(1+\alpha_1+\beta_1)(1-\alpha_1-\beta_1)}{1-2\alpha_1\beta_1-3\alpha_1^2-\beta_1^2} > 3. \tag{7.25}$$

Once again, the kurtosis is greater than that of a normal despite the innovations, e_t, all having normal distributions. The formal derivation is presented in 7.B.

7.2.3 The EGARCH model

The Exponential GARCH (EGARCH) model represents a major shift from the ARCH and GARCH models (D. B. Nelson, 1991). Rather than model the variance directly, EGARCH models the natural logarithm of the variance, and so no parameters restrictions are required to ensure that the conditional variance is positive.

Definition 7.3 (Exponential Generalized Autoregressive Conditional Heteroskedasticity (EGARCH) process)**.** An EGARCH(P,O,Q) process is defined

$$\begin{aligned} r_t &= \mu_t + \epsilon_t \\ \ln(\sigma_t^2) &= \omega + \sum_{p=1}^{P} \alpha_p \left(\left| \frac{\epsilon_{t-p}}{\sigma_{t-p}} \right| - \sqrt{\frac{2}{\pi}} \right) + \sum_{o=1}^{O} \gamma_o \frac{\epsilon_{t-o}}{\sigma_{t-o}} + \sum_{q=1}^{Q} \beta_q \ln(\sigma_{t-q}^2) \\ \epsilon_t &= \sigma_t e_t \\ e_t &\overset{\text{i.i.d.}}{\sim} N(0,1) \end{aligned} \tag{7.26}$$

where μ_t can be any adapted model for the conditional mean. In the original parameterization of D. B. Nelson (1991), P and O were assumed to be equal.

Rather than working with the complete specification, consider a simpler version, an EGARCH(1,1,1) with a constant mean,

$$\begin{aligned} r_t &= \mu + \epsilon_t \\ \ln(\sigma_t^2) &= \omega + \alpha_1 \left(\left| \frac{\epsilon_{t-1}}{\sigma_{t-1}} \right| - \sqrt{\frac{2}{\pi}} \right) + \gamma_1 \frac{\epsilon_{t-1}}{\sigma_{t-1}} + \beta_1 \ln(\sigma_{t-1}^2) \\ \epsilon_t &= \sigma_t e_t \\ e_t &\overset{\text{i.i.d.}}{\sim} N(0,1) \end{aligned} \tag{7.27}$$

which shows that log variance is a constant plus three terms. The first term, $\left|\frac{\epsilon_{t-1}}{\sigma_{t-1}}\right| - \sqrt{\frac{2}{\pi}} = |e_{t-1}| - \sqrt{\frac{2}{\pi}}$, is just the absolute value of a normal random variable, e_{t-1}, minus its expectation, $\sqrt{2/\pi}$, and so it is a mean zero shock. The second term, e_{t-1}, is also a mean zero shock and the last term is the lagged log variance. The two shocks behave differently (the e_{t-1} terms): the first produces a symmetric rise in the log variance while the second creates an asymmetric effect. γ_1 is typically estimated to be less than zero and volatility rises more subsequent to negative shocks than to positive ones. In the usual case where $\gamma_1 < 0$, the magnitude of the shock can be decomposed by conditioning on the sign of e_{t-1}

$$\text{Shock coefficient} = \begin{cases} \alpha_1 + \gamma_1 & \text{when } e_{t-1} < 0 \\ \alpha_1 - \gamma_1 & \text{when } e_{t-1} > 0 \end{cases} \tag{7.28}$$

Since both shocks are mean zero and the current log variance is linearly related to past log variance through β_1, the EGARCH(1,1,1) model is an AR model.

EGARCH models often provide superior fits when compared to standard GARCH models. The presence of the asymmetric term is largely responsible for the superior fit since many asset return series have been found to exhibit a "leverage" effect and the use of standardized shocks (e_{t-1}) in the evolution of the log-variance tend to dampen the effect of large shocks.

7.2.3.1 The S&P 500 and West Texas Intermediate Crude

The application of GARCH models will be demonstrated using daily returns on both the S&P 500 and West Texas Intermediate (WTI) Crude spot prices from January 1, 1997 until December 31, 2017. The S&P 500 data is from Yahoo! finance and the WTI data is from the St. Louis Federal Reserve's FRED database. All returns are scaled by 100. The returns are plotted in figure 7.1, the squared returns are plotted in figure 7.2 and the absolute values of the returns are plotted in figure 7.3. The plots of the squared returns and the absolute values of the returns are useful graphical diagnostics for detecting ARCH. If the residuals are conditionally heteroskedastic, both plots should produce evidence of dynamics in the transformed returns. The absolute value plot is often more useful since the squared returns are often noisy and the dynamics in the data may be obscured by a small number of outliers.

Summary statistics are presented in table 7.1 and estimates from an ARCH(5), and GARCH(1,1) and an EGARCH(1,1,1) are presented in table 7.2. The summary statistics are typical of financial data where both series are negatively skewed and heavy-tailed (leptokurtotic).

Definition 7.4 (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 7.5 (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".

Summary Statistics

	S&P 500	WTI
Ann. Mean	14.03	5.65
Ann. Volatility	38.57	19.04
Skewness	0.063	-0.028
Kurtosis	7.22	11.45

Table 7.1: Summary statistics for the S&P 500 and WTI. Means and volatilities are reported in annualized terms using 100 $\times$ returns. Skewness and kurtosis are estimated directly on the daily data and not scaled.

Table 7.2 contains estimates from an ARCH(5), a GARCH(1,1) and an EGARCH(1,1,1) model. All estimates were computed using maximum likelihood assuming the innovations are conditionally normally distributed. Examining the table, there is strong evidence of time varying variances since most p-values are near 0. The highest log-likelihood (a measure of fit) is produced by the EGARCH model in both series. This is likely due to the EGARCH's inclusion of asymmetries, a feature excluded from both the ARCH and GARCH models.

7.2.4 Alternative Specifications

Many extensions to the basic ARCH model have been introduced to capture important empirical features. This section outlines three of the most useful extensions in the ARCH-family.

7.2.4.1 GJR-GARCH

The GJR-GARCH model was named after the authors who introduced it, Glosten, Jagannathan, and Runkle (1993). It extends the standard GARCH(P,Q) to include asymmetric terms that capture an important phenomenon in the conditional variance of equities: the propensity for the volatility to rise more subsequent to large negative shocks than to large positive shocks (known as the "leverage effect").

Definition 7.6 (GJR-Generalized Autoregressive Conditional Heteroskedasticity (GJR-GARCH) process)**.** A GJR-GARCH(P,O,Q) process is defined as

$$
\begin{aligned}
r_t &= \mu_t + \epsilon_t \qquad (7.29)\\
\sigma_t^2 &= \omega + \sum_{p=1}^{P} \alpha_p \epsilon_{t-p}^2 + \sum_{o=1}^{O} \gamma_o \epsilon_{t-o}^2 I_{[\epsilon_{t-o}<0]} + \sum_{q=1}^{Q} \beta_q \sigma_{t-q}^2\\
\epsilon_t &= \sigma_t e_t\\
e_t &\stackrel{\text{i.i.d.}}{\sim} N(0,1)
\end{aligned}
$$

S&P 500

ARCH(5)						
ω	α_1	α_2	α_3	α_4	α_5	Log Lik.
0.294 (0.000)	0.095 (0.000)	0.204 (0.000)	0.189 (0.000)	0.193 (0.000)	0.143 (0.000)	−7008

GARCH(1,1)			
ω	α_1	β_1	Log Lik.
0.018 (0.000)	0.102 (0.000)	0.885 (0.000)	−6888

EGARCH(1,1,1)				
ω	α_1	γ_1	β_1	Log Lik.
0.000 (0.909)	0.136 (0.000)	−0.153 (0.000)	0.975 (0.000)	−6767

WTI

ARCH(5)						
ω	α_1	α_2	α_3	α_4	α_5	Log Lik.
2.282 (0.000)	0.138 (0.000)	0.129 (0.000)	0.131 (0.000)	0.094 (0.000)	0.130 (0.000)	−11129

GARCH(1,1)			
ω	α_1	β_1	Log Lik.
0.047 (0.034)	0.059 (0.000)	0.934 (0.000)	−11030

EGARCH(1,1,1)				
ω	α_1	γ_1	β_1	Log Lik.
0.020 (0.002)	0.109 (0.000)	−0.050 (0.000)	0.990 (0.000)	−11001

Table 7.2: Parameter estimates, p-values and log-likelihoods from ARCH(5), GARCH(1,1) and EGARCH(1,1,1) models for the S&P 500 and WTI. These parameter values are typical of models estimated on daily data. The persistence of conditional variance, as measures by the sum of the αs in the ARCH(5), $\alpha_1 + \beta_1$ in the GARCH(1,1) and β_1 in the EGARCH(1,1,1), is high in all models. The log-likelihoods indicate the EGARCH model is preferred for both return series.

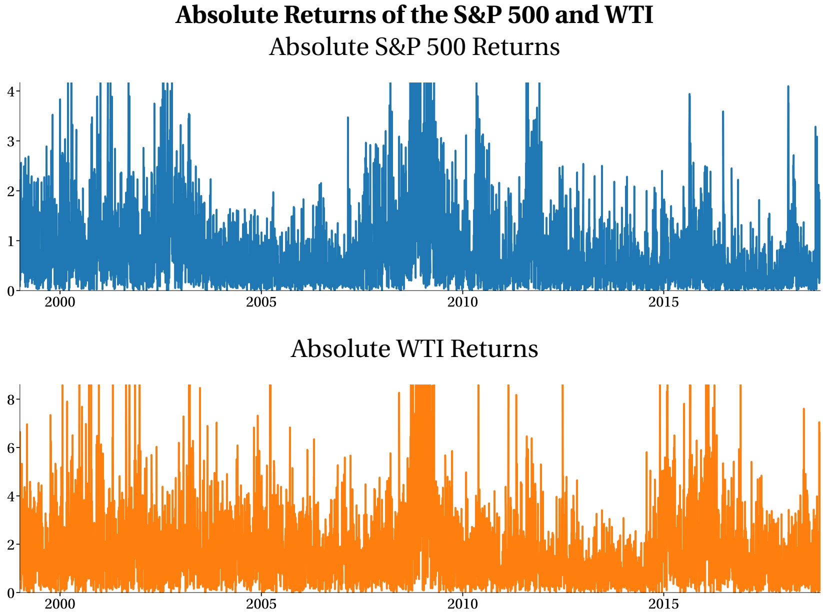

Figure 7.3: Plots of the absolute returns of the S&P 500 and WTI. Plots of the absolute value are often more useful in detecting ARCH as they are less noisy than squared returns yet still show changes in conditional volatility.

---

where μ_t can be any adapted model for the conditional mean and $I_{[\epsilon_{t-o}<0]}$ is an indicator function that takes the value 1 if $\epsilon_{t-o} < 0$ and 0 otherwise.

The parameters of the GJR-GARCH, like the standard GARCH model, must be restricted to ensure that the fit variances are always positive. This set is difficult to describe for a complete GJR-GARCH(P,O,Q) model although it is simple of a GJR-GARCH(1,1,1). The dynamics in a GJR-GARCH(1,1,1) evolve according to

$$\sigma_t^2 = \omega + \alpha_1 \epsilon_{t-1}^2 + \gamma_1 \epsilon_{t-1}^2 I_{[\epsilon_{t-1}<0]} + \beta_1 \sigma_{t-1}^2. \tag{7.30}$$

and it must be the case that $\omega > 0$, $\alpha_1 \geq 0$, $\alpha_1 + \gamma \geq 0$ and $\beta_1 \geq 0$. If the innovations are conditionally normal, a GJR-GARCH model will be covariance stationary as long as the parameter restriction are satisfied and $\alpha_1 + \frac{1}{2}\gamma_1 + \beta_1 < 1$.

7.2.4.2 AVGARCH/TARCH/ZARCH

The Threshold ARCH (TARCH) model (also known as AVGARCH and ZARCH) makes one fundamental change to the GJR-GARCH model (Taylor, 1986; Zakoian, 1994). Rather than modeling the variance directly using squared innovations, a TARCH model parameterizes the *conditional standard deviation* as a function of the lagged absolute value of the shocks. It also captures asymmetries using an asymmetric term similar in a manner similar to the asymmetry in the GJR-GARCH model.

Definition 7.7 (Threshold Autoregressive Conditional Heteroskedasticity (TARCH) process)**.** A TARCH(P, O, Q) process is defined as

$$\begin{aligned}
r_t &= \mu_t + \epsilon_t \\
\sigma_t &= \omega + \sum_{p=1}^{P} \alpha_p |\epsilon_{t-p}| + \sum_{o=1}^{O} \gamma_o |\epsilon_{t-o}| I_{[\epsilon_{t-o}<0]} + \sum_{q=1}^{Q} \beta_q \sigma_{t-q} \\
\epsilon_t &= \sigma_t e_t \\
e_t &\stackrel{\text{i.i.d.}}{\sim} N(0,1)
\end{aligned} \tag{7.31}$$

where μ_t can be any adapted model for the conditional mean. TARCH models are also known as ZARCH due to Zakoian (1994) or AVGARCH when no asymmetric terms are included ($O = 0$, Taylor (1986)).

Below is an example of a TARCH(1,1,1) model.

$$\sigma_t = \omega + \alpha_1 |\epsilon_{t-1}| + \gamma_1 |\epsilon_{t-1}| I_{[\epsilon_{t-1}<0]} + \beta_1 \sigma_{t-1}, \quad \alpha_1 + \gamma_1 \geq 0 \tag{7.32}$$

where $I_{[\epsilon_{t-1}<0]}$ is an indicator variable which takes the value 1 if $\epsilon_{t-1} < 0$. Models of the conditional standard deviation often outperform models that directly parameterize the conditional variance and the gains arise since the absolute shocks are less responsive then the squared shocks, an empirically relevant feature.

7.2.4.3 APARCH

The third model extends the TARCH and GJR-GARCH models by directly parameterizing the non-linearity in the conditional variance. Where the GJR-GARCH model uses 2 and the TARCH model uses 1, the Asymmetric Power ARCH (APARCH) of Zhuanxin Ding, Granger, and R. F. Engle (1993) parameterizes this value directly (using δ). This form provides greater flexibility in modeling the memory of volatility while remaining parsimonious.

Definition 7.8 (Asymmetric Power Autoregressive Conditional Heteroskedasticity (APARCH) process)**.** An APARCH(P,O,Q) process is defined as

$$\begin{aligned}
r_t &= \mu_t + \epsilon_t \\
\sigma_t^{\delta} &= \omega + \sum_{j=1}^{\max(P,O)} \alpha_j \left(|\epsilon_{t-j}| + \gamma_j \epsilon_{t-j} \right)^{\delta} + \sum_{q=1}^{Q} \beta_q \sigma_{t-q}^{\delta}
\end{aligned} \tag{7.33}$$

$$\epsilon_t = \sigma_t e_t$$
$$e_t \overset{\text{i.i.d.}}{\sim} N(0,1)$$

where μ_t can be any adapted model for the conditional mean. In this specification it must be the case that $P \geq O$. When $P > O$, $\gamma_j = 0$ if $j > O$. To ensure the conditional variances are non-negative, it is necessary that $\omega > 0$, $\alpha_k \geq 0$ and $-1 \leq \gamma_j \leq 1$.

It is not obvious that the APARCH model nests the GJR-GARCH and TARCH models as special cases. To examine how an APARCH nests a GJR-GARCH, consider an APARCH(1,1,1) model.

$$\sigma_t^\delta = \omega + \alpha_1 \left(|\epsilon_{t-1}| + \gamma_1 \epsilon_{t-1}\right)^\delta + \beta_1 \sigma_{t-1}^\delta \tag{7.34}$$

Suppose $\delta = 2$, then

$$\begin{aligned}
\sigma_t^2 &= \omega + \alpha_1 \left(|\epsilon_{t-1}| + \gamma_1 \epsilon_{t-1}\right)^2 + \beta_1 \sigma_{t-1}^2 \\
&= \omega + \alpha_1 |\epsilon_{t-1}|^2 + 2\alpha_1 \gamma_1 \epsilon_{t-1} |\epsilon_{t-1}| + \alpha_1 \gamma_1^2 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \\
&= \omega + \alpha_1 \epsilon_{t-1}^2 + \alpha_1 \gamma_1^2 \epsilon_{t-1}^2 + 2\alpha_1 \gamma_1 \epsilon_{t-1}^2 \operatorname{sign}(\epsilon_{t-1}) + \beta_1 \sigma_{t-1}^2
\end{aligned} \tag{7.35}$$

where sign$(\cdot)$ is a function that returns 1 if its argument is positive and -1 if its argument is negative. Consider the total effect of ϵ_{t-1}^2 as it depends on the sign of ϵ_{t-1},

$$\text{Shock coefficient } = \begin{cases} \alpha_1 + \alpha_1 \gamma_1^2 + 2\alpha_1 \gamma_1 & \text{when } \epsilon_t > 0 \\ \alpha_1 + \alpha_1 \gamma_1^2 - 2\alpha_1 \gamma_1 & \text{when } \epsilon_t < 0 \end{cases} \tag{7.36}$$

γ is usually estimated to be less than zero which corresponds to the typical "leverage effect" in GJR-GARCH models.[8] The relationship between a TARCH model and an APARCH model works analogously setting $\delta = 1$. The APARCH model also nests the ARCH(P), GARCH(P,Q) and AVGARCH(P,Q) models as special cases.

7.2.5 The News Impact Curve

With a wide range of volatility models, each with a different specification for the dynamics of conditional variances, it can be difficult to determine the precise effect of a shock to the conditional variance. *News impact curves* which measure the effect of a shock in the current period on the conditional variance in the subsequent period facilitate comparison between models.

Definition 7.9 (News Impact Curve (NIC))**.** The news impact curve of an ARCH-family model is defined as the difference between the variance with a shock e_t and the variance with no shock

[8] The explicit relationship between an APARCH and a GJR-GARCH can be derived when $\delta = 2$ by solving a system of two equation in two unknowns where eq. (7.36) is equated with the effect in a GJR-GARCH model.

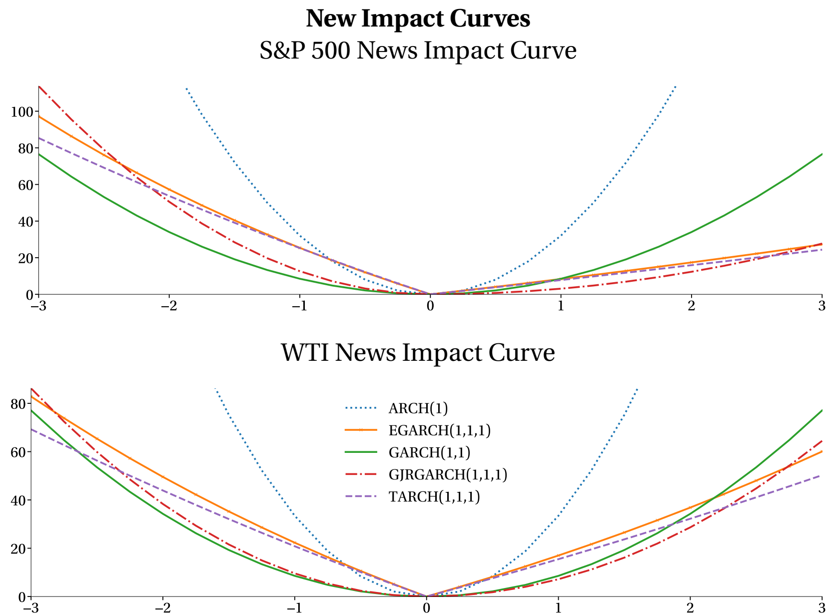

Figure 7.4: News impact curves for returns on both the S&P 500 and WTI. While the ARCH and GARCH curves are symmetric, the others show substantial asymmetries to negative news. Additionally, the fit APARCH models chose $\hat{\hat{\delta}} \approx 1$ and so the NIC of the APARCH and the TARCH models appear similar.

$(e_t = 0)$. To ensure that the NIC does not depend on the level of variance, the variance in all previous periods is set to the unconditional expectation of the variance, $\bar{\sigma}^2$,

$$n(e_t) = \sigma^2_{t+1}(e_t | \sigma^2_t = \bar{\sigma}^2_t) \tag{7.37}$$

$$NIC(e_t) = n(e_t) - n(0) \tag{7.38}$$

To facilitate comparing both linear and non-linear specification (e.g. EGARCH) NICs are normalized by setting the variance in the current period to the unconditional variance.

News impact curves for ARCH and GARCH models are simply the terms which involve ϵ_t^2.

GARCH(1,1)

$$n(e_t) = \omega + \alpha_1 \bar{\sigma}^2 e_t^2 + \beta_1 \bar{\sigma}^2 \tag{7.39}$$

$$NIC(e_t) = \alpha_1 \bar{\sigma}^2 e_t^2 \tag{7.40}$$

The news impact curve can be fairly complicated if a model is not linear in ϵ_t^2, as this example from a TARCH(1,1,1) shows.

$$\sigma_t = \omega + \alpha_1|\epsilon_t| + \gamma_1|\epsilon_t|I_{[\epsilon_t<0]} + \beta_1\sigma_{t-1} \tag{7.41}$$

$$n(e_t) = \omega^2 + 2\omega(\alpha_1+\gamma_1 I_{[\epsilon_t<0]})|\epsilon_t| + 2\beta(\alpha_1+\gamma_1 I_{[\epsilon_t<0]})|\epsilon_t|\bar{\sigma} + \beta_1^2\bar{\sigma}^2 + 2\omega\beta_1\bar{\sigma} + (\alpha_1+\gamma_1 I_{[\epsilon_t<0]})^2\epsilon_t^2 \tag{7.42}$$

$$NIC(e_t) = (\alpha_1+\gamma_1 I_{[\epsilon_t<0]})^2\epsilon_t^2 + (2\omega + 2\beta_1\bar{\sigma})(\alpha_1+\gamma_1 I_{[\epsilon_t<0]})|\epsilon_t| \tag{7.43}$$

While deriving explicit expressions for NICs can be tedious practical implementation only requires computing the conditional variance for a shock of 0 ($n(0)$) and for a set of shocks between -3 and 3 ($n(z)$ for $z \in (-3,3)$). The difference between the conditional variance with a shock and the conditional variance without a shock is the NIC.

7.2.5.1 The S&P 500 and WTI

Figure 7.4 contains plots of the news impact curves for both the S&P 500 and WTI. When the models include asymmetries, the news impact curves are asymmetric and show a much larger response to negative shocks than to positive shocks, although the asymmetry is stronger in the volatility of the returns of the S&P 500 than it is in the volatility of WTI's returns.

7.2.6 Estimation

Consider a simple GARCH(1,1) specification,

$$\begin{aligned} r_t &= \mu_t + \epsilon_t \\ \sigma_t^2 &= \omega + \alpha_1\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2 \\ \epsilon_t &= \sigma_t e_t \\ e_t &\overset{\text{i.i.d.}}{\sim} N(0,1) \end{aligned} \tag{7.44}$$

Since the errors are assumed to be conditionally i.i.d. normal[9], maximum likelihood is a natural choice to estimate the unknown parameters, $\boldsymbol{\theta}$ which contain both the mean and variance parameters. The normal likelihood for T independent variables is

$$f(\mathbf{r};\boldsymbol{\theta}) = \prod_{t=1}^{T}(2\pi\sigma_t^2)^{-\frac{1}{2}}\exp\left(-\frac{(r_t-\mu_t)^2}{2\sigma_t^2}\right) \tag{7.45}$$

and the normal log-likelihood function is

[9] The use of conditional is to denote the dependence on σ_t^2, which is in $\mathcal{F}_{t-1}$.

$$l(\boldsymbol{\theta};\mathbf{r})=\sum_{t=1}^{T}-\frac{1}{2}\log(2\pi)-\frac{1}{2}\log(\sigma_t^2)-\frac{(r_t-\mu_t)^2}{2\sigma_t^2}. \tag{7.46}$$

If the mean is set to 0, the log-likelihood simplifies to

$$l(\boldsymbol{\theta};\mathbf{r})=\sum_{t=1}^{T}-\frac{1}{2}\log(2\pi)-\frac{1}{2}\log(\sigma_t^2)-\frac{r_t^2}{2\sigma_t^2} \tag{7.47}$$

and is maximized by solving the first order conditions.

$$\frac{\partial l(\boldsymbol{\theta};\mathbf{r})}{\partial\sigma_t^2}=\sum_{t=1}^{T}-\frac{1}{2\sigma_t^2}+\frac{r_t^2}{2\sigma_t^4}=0 \tag{7.48}$$

which can be rewritten to provide some insight into the estimation of ARCH models,

$$\frac{\partial l(\boldsymbol{\theta};\mathbf{r})}{\partial\sigma_t^2}=\frac{1}{2}\sum_{t=1}^{T}\frac{1}{\sigma_t^2}\left(\frac{r_t^2}{\sigma_t^2}-1\right). \tag{7.49}$$

This expression clarifies that the parameters of the volatility are chosen to make $\left(\frac{r_t^2}{\sigma_t^2}-1\right)$ as close to zero as possible.[10] These first order conditions are not complete since ω, α_1 and β_1, not σ_t^2, are the parameters of a GARCH(1,1) model and

$$\frac{\partial l(\boldsymbol{\theta};\mathbf{r})}{\partial\theta_i}=\frac{\partial l(\boldsymbol{\theta};\mathbf{r})}{\partial\sigma_t^2}\frac{\partial\sigma_t^2}{\partial\theta_i} \tag{7.50}$$

The derivatives take an recursive form not previously encountered,

$$\begin{aligned}\frac{\partial\sigma_t^2}{\partial\omega}&=1+\beta_1\frac{\partial\sigma_{t-1}^2}{\partial\omega}\\ \frac{\partial\sigma_t^2}{\partial\alpha_1}&=\epsilon_{t-1}^2+\beta_1\frac{\partial\sigma_{t-1}^2}{\partial\alpha_1}\end{aligned} \tag{7.51}$$

[10] If $E_{t-1}\left[\frac{r_t^2}{\sigma_t^2}-1\right]=0$, and so the volatility is correctly specified, then the scores of the log-likelihood have expectation zero since

$$\begin{aligned}\mathrm{E}\left[\frac{1}{\sigma_t^2}\left(\frac{r_t^2}{\sigma_t^2}-1\right)\right]&=\mathrm{E}\left[\mathrm{E}_{t-1}\left[\frac{1}{\sigma_t^2}\left(\frac{r_t^2}{\sigma_t^2}-1\right)\right]\right]\\ &=\mathrm{E}\left[\frac{1}{\sigma_t^2}\left(\mathrm{E}_{t-1}\left[\frac{r_t^2}{\sigma_t^2}-1\right]\right)\right]\\ &=\mathrm{E}\left[\frac{1}{\sigma_t^2}(0)\right]\\ &=0.\end{aligned}$$

$$\frac{\partial \sigma_t^2}{\partial \beta_1} = \sigma_{t-1}^2 + \beta_1 \frac{\partial \sigma_{t-1}^2}{\partial \beta_1},$$

although the recursion in the first order condition for ω can be removed noting that

$$\frac{\partial \sigma_t^2}{\partial \omega} = 1 + \beta_1 \frac{\partial \sigma_{t-1}^2}{\partial \omega} \approx \frac{1}{1-\beta_1}. \tag{7.52}$$

Eqs. (7.50) – (7.52) provide the necessary formulas to implement the scores of the log-likelihood although they are not needed to estimate a GARCH model.[11]

The use of the normal likelihood has one strong justification; estimates produced by maximizing the log-likelihood of a normal are *strongly consistent*. Strong consistency is a property of an estimator that ensures parameter estimates converge to the true parameters *even if the wrong conditional distribution is assumed*. For example, in a standard GARCH(1,1), the parameter estimates would still converge to their true value if estimated with the normal likelihood as long as the volatility model was correctly specified. The intuition behind this result comes form the *generalized error*

$$\left(\frac{\epsilon_t^2}{\sigma_t^2} - 1\right). \tag{7.53}$$

Whenever $\sigma_t^2 = \mathrm{E}_{t-1}[\epsilon_t^2]$,

$$\mathrm{E}\left[\left(\frac{\epsilon_t^2}{\sigma_t^2} - 1\right)\right] = \mathrm{E}\left[\left(\frac{\mathrm{E}_{t-1}[\epsilon_t^2]}{\sigma_t^2} - 1\right)\right] = \mathrm{E}\left[\left(\frac{\sigma_t^2}{\sigma_t^2} - 1\right)\right] = 0. \tag{7.54}$$

Thus, as long as the GARCH model nests the true DGP, the parameters will be chosen to make the conditional expectation of the generalized error 0; these parameters correspond to those of the original DGP even if the conditional distribution is misspecified.[12] This is a very special property of the normal distribution and is not found in any other common distribution.

7.2.7 Inference

Under some regularity conditions, parameters estimated using maximum likelihood have been shown to be asymptotically normally distributed,

$$\sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N(0, \mathcal{I}^{-1}) \tag{7.55}$$

[11] MATLAB and many other econometric packages are capable of estimating the derivatives using a numerical approximation that only requires the log-likelihood. Numerical derivatives use the definition of a derivative, $f'(x) = \lim_{h\to 0} \frac{f(x+h)-f(x)}{h}$ to approximate the derivative using $f'(x) \approx \frac{f(x+h)-f(x)}{h}$ for some small h.

[12] An assumption that a GARCH specification nests the DGP is extremely strong and almost certainly wrong. However, this property of the normal provides a justification even though the standardized residuals of most asset return series are leptokurtotic and skewed.

where

$$\mathcal{I} = -\mathrm{E}\left[\frac{\partial^2 l(\boldsymbol{\theta}_0; r_t)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right] \tag{7.56}$$

is the negative of the expected Hessian. The Hessian measures how much curvature there is in the log-likelihood at the optimum just like the second-derivative measures the rate-of-change in the rate-of-change of the function in a standard calculus problem. To estimate $\mathcal{I}$, the sample analogue employing the time-series of Hessian matrices computed at $\hat{\boldsymbol{\theta}}$ is used,

$$\hat{\mathcal{I}} = T^{-1}\sum_{t=1}^{T}\frac{\partial^2 l(\hat{\boldsymbol{\theta}}; r_t)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}. \tag{7.57}$$

The chapter 1 notes show that the Information Matrix Equality (IME) generally holds for MLE problems, so

$$\mathcal{I} = \mathcal{J} \tag{7.58}$$

where

$$\mathcal{J} = \mathrm{E}\left[\frac{\partial l(r_t; \boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\frac{\partial l(r_t; \boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}'}\right] \tag{7.59}$$

is the covariance of the scores, which measures how much information there is in the data to pin down the parameters. Large score variance indicate that small parameter changes have a large impact and so the parameters are precisely estimated. The estimator of $\mathcal{J}$ is the sample analogue using the scores evaluated at the estimated parameters,

$$\hat{\mathcal{J}} = T^{-1}\sum_{t=1}^{T}\frac{\partial l(\hat{\boldsymbol{\theta}}; r_t)}{\partial\boldsymbol{\theta}}\frac{\partial l(\hat{\boldsymbol{\theta}}; r_t)}{\partial\boldsymbol{\theta}'}. \tag{7.60}$$

The conditions for the IME to hold require that the parameter estimates *are maximum likelihood estimates* which in turn requires both the likelihood used in estimation to be correct as well as the specification for the conditional variance. When one specification is used for estimation (e.g. normal) but the data follow a different conditional distribution, these estimators are known as Quasi Maximum Likelihood Estimators (QMLE) and the IME generally fails to hold. Under some regularity conditions, the estimated parameters are still asymptotically normal but with a different covariance,

$$\sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N(0, \mathcal{I}^{-1}\mathcal{J}\mathcal{I}^{-1}) \tag{7.61}$$

If the IME was valid, $\mathcal{I} = \mathcal{J}$ and so this covariance would simplify to the usual MLE variance estimator.

In most applications of ARCH models, the conditional distribution of shocks is decidedly not normal, exhibiting both excess kurtosis and skewness. Bollerslev and Wooldridge (1992) were the first to show that the IME does not generally hold for GARCH models when the distribution

WTI

	ω	α_1	γ_1	β_1
Coefficient	0.031	0.030	0.055	0.942
Std. T-stat	3.62	4.03	7.67	102.94
Robust T-stat	1.85	2.31	4.45	49.66

S&P 500

	ω	α_1	γ_1	β_1
Coefficient	0.026	0.000	0.172	0.909
Std. T-stat	9.63	0.000	14.79	124.92
Robust T-stat	6.28	0.000	10.55	93.26

Table 7.3: Estimates from a TARCH(1,1,1) for the S&P 500 and WTI using alternative paramter covariance estimators.

---

is misspecified and the "sandwich" form

$$\hat{\mathcal{I}}^{-1}\hat{\mathcal{J}}\hat{\mathcal{I}}^{-1} \tag{7.62}$$

of the covariance estimator is often referred to as the *Bollerslev-Wooldridge* covariance matrix or simply a robust covariance matrix. Standard Wald tests can be used to test hypotheses of interest, such as whether an asymmetric term is statistically significant, although likelihood ratio tests are not reliable since they do not have the usual χ^2_m distribution.

7.2.7.1 The S&P 500 and WTI

To demonstrate the different covariance estimators, TARCH(1,1,1) models were estimated on both the S&P 500 and WTI returns. Table 7.3 contains the estimated parameters and t-stats using both the MLE covariance matrix and the Bollerslev-Wooldridge covariance matrix. The robust t-stats are substantially smaller than contentional ones, although conculsions about statistical significance are not affected except for ω in the WTI model. These changes are due to the heavy-tailed nature of the standardized residuals in these series.

7.2.7.2 Independence of the mean and variance parameters

One important but not obvious issue when using MLE assuming conditionally normal errors - or QMLE when conditional normality is wrongly assumed - is that *inference on the parameters of the ARCH model is still correct as long as the model for the mean is general enough to nest the true form.* As a result, it is safe to estimate the mean and variance parameters separately without correcting the covariance matrix of the estimated parameters.[13] This surprising feature of QMLE estimators employing a normal log-likelihood comes form the cross-partial derivative of the log-likelihood with respect to the mean and variance parameters,

$$l(\boldsymbol{\theta}; r_t) = -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma_t^2) - \frac{(r_t - \mu_t)^2}{2\sigma_t^2}. \tag{7.63}$$

The first order condition is,

$$\frac{\partial l(\boldsymbol{\theta}; \mathbf{r})}{\partial \mu_t}\frac{\partial \mu_t}{\partial \phi} = -\sum_{t=1}^{T}\frac{(r_t - \mu_t)}{\sigma_t^2}\frac{\partial \mu_t}{\partial \phi} \tag{7.64}$$

and the second order condition is

$$\frac{\partial^2 l(\boldsymbol{\theta}; \mathbf{r})}{\partial \mu_t \partial \sigma_t^2}\frac{\partial \mu_t}{\partial \phi}\frac{\partial \sigma_t^2}{\partial \psi} = \sum_{t=1}^{T}\frac{(r_t - \mu_t)}{\sigma_t^4}\frac{\partial \mu_t}{\partial \phi}\frac{\partial \sigma_t^2}{\partial \psi} \tag{7.65}$$

where ϕ is a parameter of the conditional mean and ψ is a parameter of the conditional variance. For example, in a simple ARCH(1) model with a constant mean,

$$\begin{aligned} r_t &= \mu + \epsilon_t \\ \sigma_t^2 &= \omega + \alpha_1 \epsilon_{t-1}^2 \\ \epsilon_t &= \sigma_t e_t \\ e_t &\overset{\text{i.i.d.}}{\sim} N(0,1), \end{aligned} \tag{7.66}$$

$\phi = \mu$ and ψ can be either ω or α_1. Taking expectations of the cross-partial,

$$\begin{aligned} \mathrm{E}\left[\frac{\partial^2 l(\boldsymbol{\theta}; \mathbf{r})}{\partial \mu_t \partial \sigma_t^2}\frac{\partial \mu_t}{\partial \phi}\frac{\partial \sigma_t^2}{\partial \psi}\right] &= \mathrm{E}\left[\sum_{t=1}^{T}\frac{r_t - \mu_t}{\sigma_t^4}\frac{\partial \mu_t}{\partial \phi}\frac{\partial \sigma_t^2}{\partial \psi}\right] \\ &= \mathrm{E}\left[\mathrm{E}_{t-1}\left[\sum_{t=1}^{T}\frac{r_t - \mu_t}{\sigma_t^4}\frac{\partial \mu_t}{\partial \phi}\frac{\partial \sigma_t^2}{\partial \psi}\right]\right] \\ &= \mathrm{E}\left[\sum_{t=1}^{T}\frac{\mathrm{E}_{t-1}\left[r_t - \mu_t\right]}{\sigma_t^4}\frac{\partial \mu_t}{\partial \phi}\frac{\partial \sigma_t^2}{\partial \psi}\right] \\ &= \mathrm{E}\left[\sum_{t=1}^{T}\frac{0}{\sigma_t^4}\frac{\partial \mu_t}{\partial \phi}\frac{\partial \sigma_t^2}{\partial \psi}\right] \end{aligned} \tag{7.67}$$

[13]The estimated covariance for the mean should use a White covariance estimator. If the mean parameters are of particular interest, it may be more efficient to jointly estimate the parameters of the mean and volatility equations as a form of GLS (see Chapter 3).

$$= 0$$

it can be seen that the expectation of the cross derivative is 0. The intuition behind this result is also simple: if the mean model is correct for the conditional expectation of r_t, the term $r_t - \mu_t$ has conditional expectation 0 and knowledge of the variance is not needed, and is a similar argument to the validity of least squares estimation when the errors are heteroskedastic.

7.2.8 GARCH-in-Mean

The GARCH-in-mean model (GIM) makes a significant change to the role of time-varying volatility by explicitly relating the level of volatility to the expected return (R. F. Engle, Lilien, and Robins, 1987). A simple GIM model can be specified as

$$\begin{aligned} r_t &= \mu + \delta\sigma_t^2 + \epsilon_t \\ \sigma_t^2 &= \omega + \alpha_1\epsilon_{t-1}^2 + \beta_1\sigma_{t-1}^2 \\ \epsilon_t &= \sigma_t e_t \\ e_t &\overset{\text{i.i.d.}}{\sim} N(0,1) \end{aligned} \tag{7.68}$$

although virtually any ARCH-family model could be used for the conditional variance. The obvious difference between the GIM and a standard GARCH(1,1) is that the variance appears in the mean of the return. Note that the shock driving the changes in variance is not the mean return but still ϵ_{t-1}^2, and so the ARCH portion of a GIM is unaffected. Other forms of the GIM model have been employed where the conditional standard deviation or the log of the conditional variance are used in the mean equation[14],

$$r_t = \mu + \delta\sigma_t + \epsilon_t \tag{7.69}$$

or

$$r_t = \mu + \delta\ln(\sigma_t^2) + \epsilon_t \tag{7.70}$$

Because the variance appears in the mean equation for r_t, the mean and variance parameters cannot be separately estimated. Despite the apparent feedback, processes that follow a GIM will be stationary as long as the variance process is stationary. This result follows from noting that the conditional variance (σ_t^2) in the conditional mean does not feedback into the conditional variance process.

7.2.8.1 The S&P 500

Asset pricing theorizes that there is a risk-return trade off and GARCH-in-mean models provide a natural method to test whether this is the case. Using the S&P 500 data, three GIM models

[14] The model for the conditional mean can be extended to include ARMA terms or any other predetermined regressor.

S&P 500 Garch-in-Mean Estimates

	μ	δ	ω	α	γ	β	Log Lik.
σ^2	0.004 (0.753)	0.022 (0.074)	0.022 (0.000)	0.000 (0.999)	0.183 (0.000)	0.888 (0.000)	−6773.7
σ	−0.034 (0.304)	0.070 (0.087)	0.022 (0.000)	0.000 (0.999)	0.182 (0.000)	0.887 (0.000)	−6773.4
$\ln\sigma^2$	0.038 (0.027)	0.030 (0.126)	0.022 (0.000)	0.000 (0.999)	0.183 (0.000)	0.888 (0.000)	−6773.8

Table 7.4: GARCH-in-mean estimates for the S&P 500 series. δ, the parameter which measures the GIM effect, is the most interesting parameter and it is significant in both the log variance specification and the variance specification. The GARCH model estimated was a standard GARCH(1,1). P-values are in parentheses.

were estimated (one for each transformation of the variance in the mean equation) and the results are presented in table 7.4. Based on these estimates, there does appear to be a trade off between mean and variance and higher variances produce higher expected means, although the magnitude is economically small.

7.2.9 Alternative Distributional Assumptions

Despite the strengths of the assumption that the errors are conditionally normal (estimation is simple and parameters are *strongly consistent* for the true parameters), GARCH models can be specified and estimated with alternative distributional assumptions. The motivation for using something other than the normal distribution is two-fold. First, a better approximation to the conditional distribution of the standardized returns may improve the precision of the volatility process parameter estimates and, in the case of MLE, the estimates will be fully efficient. Second, GARCH models are often used in situations where the choice of the density matters such as Value-at-Risk and option pricing.

Three distributions stand among the myriad that have been used to estimate the parameters of GARCH processes. The first is a standardized Student's t (to have unit variance for any value ν, see Bollerslev (1987)) with ν degrees of freedom,

Standardized Student's t

$$f(r_t;\mu,\sigma_t^2,\nu)=\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}\frac{1}{\sqrt{\pi(\nu-2)}}\frac{1}{\sigma_t}\frac{1}{\left(1+\frac{(r_t-\mu)^2}{\sigma_t^2(\nu-2)}\right)^{\frac{\nu+1}{2}}} \tag{7.71}$$

where $\Gamma(\cdot)$ is the gamma function.[15] This distribution is always fat-tailed and produces a better

[15]The standardized Student's t differs from the usual Student's t so that the it is necessary to scale data by $\sqrt{\frac{\nu}{\nu-2}}$ if using functions (such as the CDF) for the regular Student's t distribution.

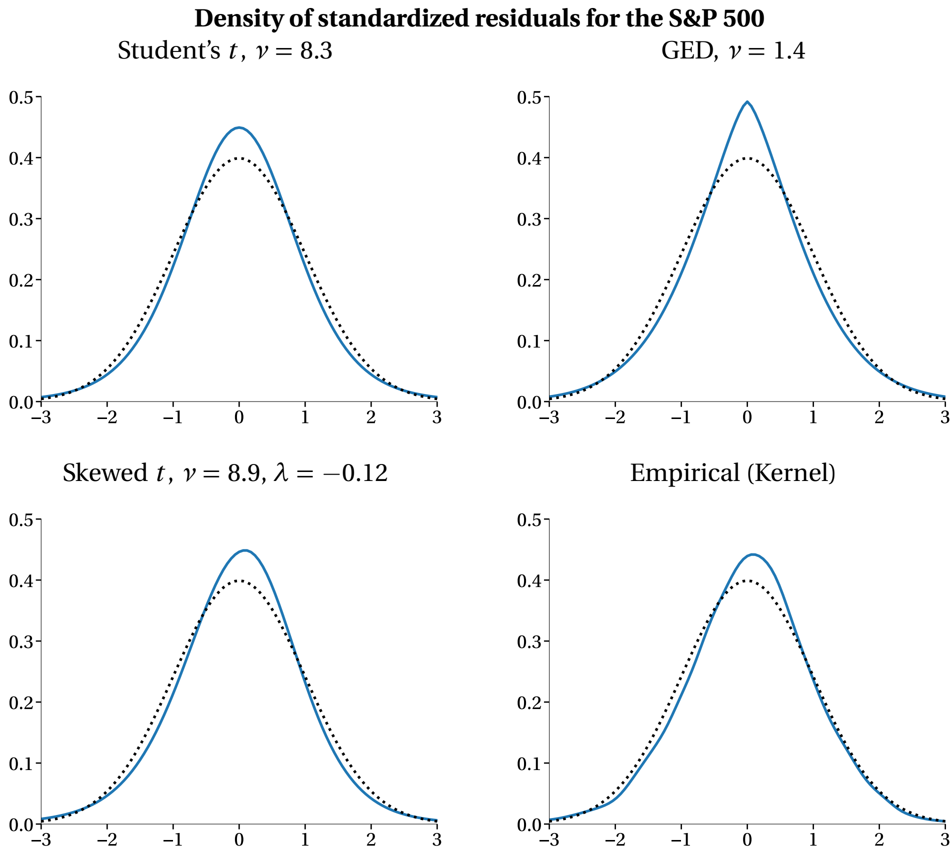

Figure 7.5: The four panels of this figure contain the estimated density for the S&P 500 and the density implied by the distributions: Student's t, GED, Hansen's Skew t and a kernel density plot of the standardized residuals, $\hat{e}_t = \epsilon_t/\hat{\sigma}_t$, along with the PDF of a normal (dotted line) for comparison. The shape parameters in the Student's t, GED and skewed t, ν and λ, were jointly estimated with the variance parameters.

---

fit than the normal for most asset return series. This distribution is only well defined if $\nu > 2$ since the variance of a Student's t with $\nu \leq 2$ is infinite. The second is the generalized error distribution (GED, see D. B. Nelson (1991)),

Generalized Error Distribution

$$f(r_t;\mu,\sigma_t^2,\nu) = \frac{\nu \exp\left(-\frac{1}{2}\left|\frac{r_t-\mu}{\sigma_t\lambda}\right|^{\nu}\right)}{\sigma_t \lambda 2^{\frac{\nu+1}{\nu}}\Gamma(\frac{1}{\nu})} \tag{7.72}$$

$$\lambda = \sqrt{\frac{2^{-\frac{2}{\nu}}\Gamma(\frac{1}{\nu})}{\Gamma(\frac{3}{\nu})}} \tag{7.73}$$

which nests the normal when $\nu = 2$. The GED is fat-tailed when $\nu < 2$ and thin-tailed when $\nu > 2$. In order for this distribution to be used for estimating GARCH parameters, it is necessary that $\nu \geq 1$ since the variance is infinite when $\nu < 1$. The third useful distribution introduced in B. E. Hansen (1994) extends the standardized Student's t to allow for skewness of returns

Hansen's skewed t

$$f(\epsilon_t;\mu,\sigma_t,\nu,\lambda) = \begin{cases} bc\left(1+\frac{1}{\nu-2}\left(\frac{b\left(\frac{r_t-\mu}{\sigma_t}\right)+a}{(1-\lambda)}\right)^2\right)^{-(\nu+1)/2} & , \quad \frac{r_t-\mu}{\sigma_t} < -a/b \\ bc\left(1+\frac{1}{\nu-2}\left(\frac{b\left(\frac{r_t-\mu}{\sigma_t}\right)+a}{(1+\lambda)}\right)^2\right)^{-(\nu+1)/2} & , \quad \frac{r_t-\mu}{\sigma_t} \geq -a/b \end{cases} \tag{7.74}$$

where

$$a = 4\lambda c\left(\frac{\nu-2}{\nu-1}\right),$$
$$b = \sqrt{1+3\lambda^2-a^2},$$

and

$$c = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi(\nu-2)}\Gamma\left(\frac{\nu}{2}\right)}.$$

The two shape parameters, ν and λ, control the kurtosis and the skewness, respectively.

These distributions may be better approximations to the true distribution since they allow for kurtosis greater than that of the normal, an important empirical fact, and, in the case of the skewed t, skewness in the standardized returns. Chapter 8 will return to these distributions in the context of Value-at-Risk and density forecasting.

7.2.9.1 Alternative Distribution in Practice

To explore the role of alternative distributional assumptions in the estimation of GARCH models, a TARCH(1,1,1) was fit to the S&P 500 returns using the conditional normal, the Student's t, the GED and Hansen's skewed t. Figure 7.5 contains the empirical density (constructed with a kernel) and the fit density of the three distributions. Note that the shape parameters, ν and λ, were jointly estimated with the conditional variance parameters. Figure 7.6 shows plots of the estimated conditional variance for both the S&P 500 and WTI using three distributional assumptions. The

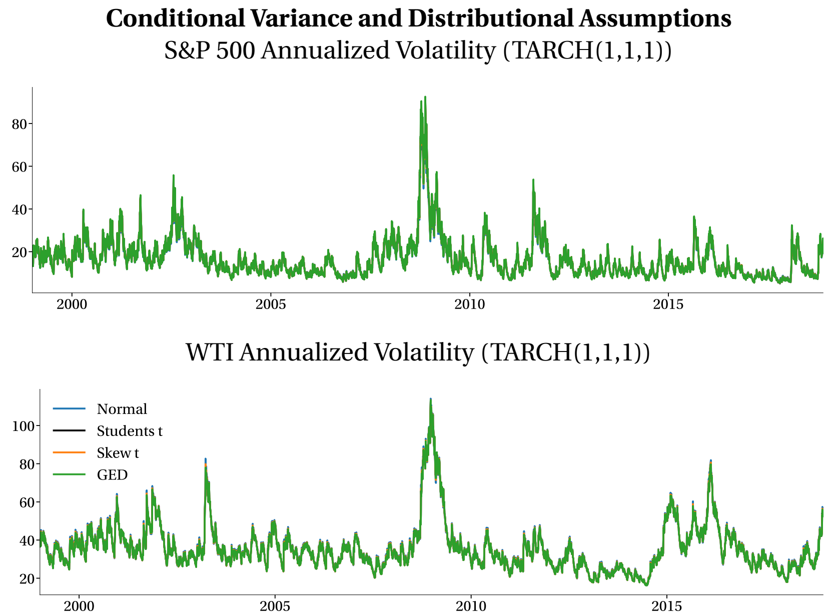

Figure 7.6: The choice of the distribution for the standardized innovation makes little difference to the fit variances or the estimated parameters in most models. The alternative distributions are more useful in application to Value-at-Risk and Density forecasting in which case the choice of density may make a large difference.

---

most important aspect of this figure is that the fit variances are virtually identical. This is a common finding estimating models using alternative distributional assumptions where it is often found that there is little meaningful difference in the fit conditional variances or estimated parameters.

7.2.10 Model Building

Since ARCH and GARCH models are similar to AR and ARMA models, the Box-Jenkins methodology is a natural way to approach the problem. The first step is to analyze the sample ACF and PACF of the *squared* returns, or if the model for the conditional mean is non-trivial, the sample ACF and PACF of the estimated residuals, $\hat{\epsilon}_t$ should be examined for heteroskedasticity. Figures 7.7 and 7.8 contains the ACF and PACF for the squared returns of the S&P 500 and WTI respectively. The models used in selecting the final model are reproduced in tables 7.5 and 7.6 respectively. Both selections began with a simple GARCH(1,1). The next step was to check if more lags were needed for either the squared innovation or the lagged variance by fitting a GARCH(2,1) and a GARCH(1,2) to each series. Neither of these meaningfully improved the fit

	α_1	α_2	γ_1	γ_2	β_1	β_2	Log Lik.
GARCH(1,1)	0.102 (0.000)				0.885 (0.000)		−6887.6
GARCH(1,2)	0.102 (0.000)				0.885 (0.000)	0.000 (0.999)	−6887.6
GARCH(2,1)	0.067 (0.003)	0.053 (0.066)			0.864 (0.000)		−6883.5
GJR-GARCH(1,1,1)	0.000 (0.999)		0.185 (0.000)		0.891 (0.000)		−6775.1
GJR-GARCH(1,2,1)	0.000 (0.999)		0.158 (0.000)	0.033 (0.460)	0.887 (0.000)		−6774.5
TARCH(1,1,1)*	0.000 (0.999)		0.172 (0.000)		0.909 (0.000)		−6751.9
TARCH(1,2,1)	0.000 (0.999)		0.165 (0.000)	0.009 (0.786)	0.908 (0.000)		−6751.8
TARCH(2,1,1)	0.000 (0.999)	0.003 (0.936)	0.171 (0.000)		0.907 (0.000)		−6751.9
EGARCH(1,0,1)	0.211 (0.000)				0.979 (0.000)		−6908.4
EGARCH(1,1,1)	0.136 (0.000)		−0.153 (0.000)		0.975 (0.000)		−6766.7
EGARCH(1,2,1)	0.129 (0.000)		−0.213 (0.000)	0.067 (0.045)	0.977 (0.000)		−6761.7
EGARCH(2,1,1)	0.020 (0.651)	0.131 (0.006)	−0.162 (0.000)		0.970 (0.000)		−6757.6

Table 7.5: The models estimated in selecting a final model for the conditional variance of the S&P 500 Index. ⋆ indicates the selected model.

and a GARCH(1,1) was assumed to be sufficient to capture the symmetric dynamics.

The next step in model building is to examine whether the data exhibit any evidence of asymmetries using a GJR-GARCH(1,1,1). The asymmetry term was significant and so other forms of the GJR model were explored. All were found to provide little improvement in the fit. Once a GJR-GARCH(1,1,1) model was decided upon, a TARCH(1,1,1) was fit to test whether evolution in variances or standard deviations was more appropriate. Both series preferred the TARCH to the GJR-GARCH (compare the log-likelihoods), and the TARCH(1,1,1) was selected. In comparing alternative specifications, an EGARCH was fit and found to also provide a good description of the data. In both cases the EGARCH was expanded to include more lags of the shocks or lagged log volatility. The EGARCH did not improve over the TARCH for the S&P 500 and so the TARCH(1,1,1) was selected. The EGARCH did fit the WTI data better and so the preferred model is an EGARCH(1,1,1), although a case could be made for the EGARCH(2,1,1) which provided a better fit. Overfitting is always a concern and the opposite signs on α_1 and α_2 in the EGARCH(2,1,1) are suspisious.

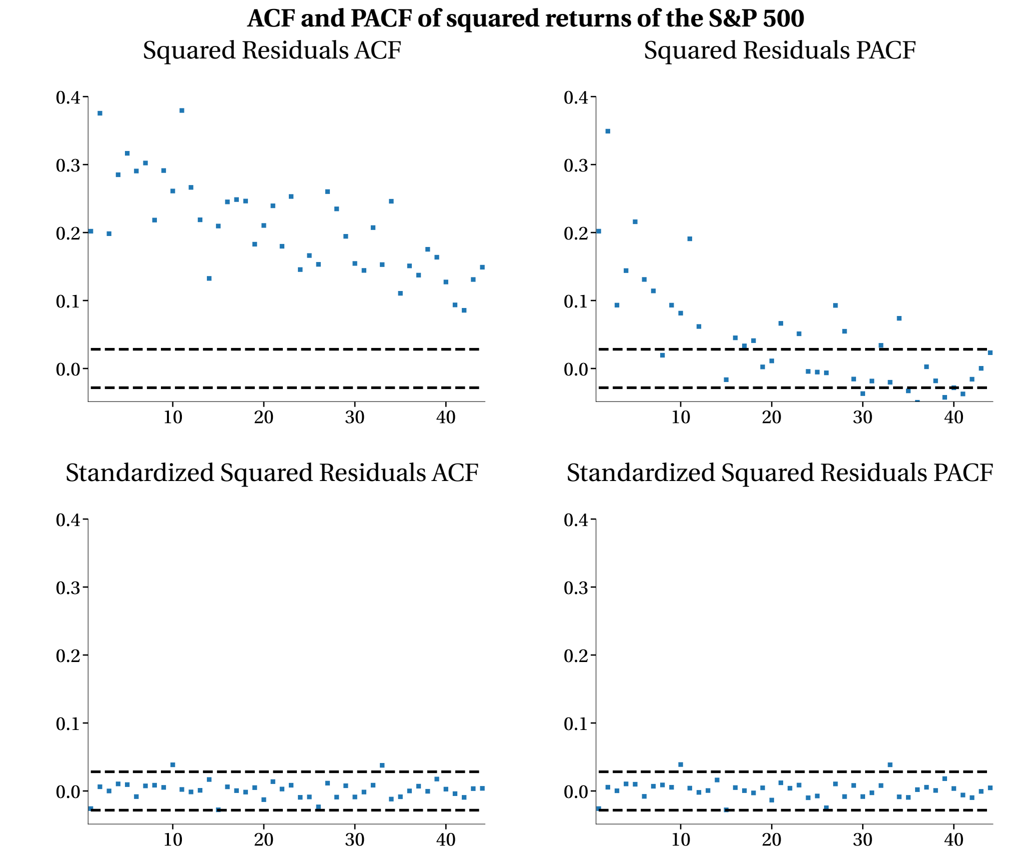

Figure 7.7: ACF and PACF of the squared returns for the S&P 500. The bottom two panels contain the ACF and PACF of $\hat{e}_t^2 = \hat{\epsilon}_t^2/\hat{\sigma}_t^2$. The top panels show persistence in both the ACF and PACF which indicate an ARMA model is needed (hence a GARCH model) while the ACF and PACF of the standardized residuals appear to be compatible with an assumption of white noise.

7.2.10.1 Testing for (G)ARCH

Although conditional heteroskedasticity can often be identified by graphical inspection, a formal test of conditional homoskedasticity is also useful. The standard method to test for ARCH is to use the ARCH-LM test which is implemented as a regression of *squared* residuals on lagged squared residuals and it directly exploits the AR representation of an ARCH process (R. F. Engle, 1982). The test is computed by estimating

$$\hat{\epsilon}_t^2 = \phi_0 + \phi_1 \hat{\epsilon}_{t-1}^2 + \ldots + \phi_P \hat{\epsilon}_{t-P}^2 + \eta_t \tag{7.75}$$

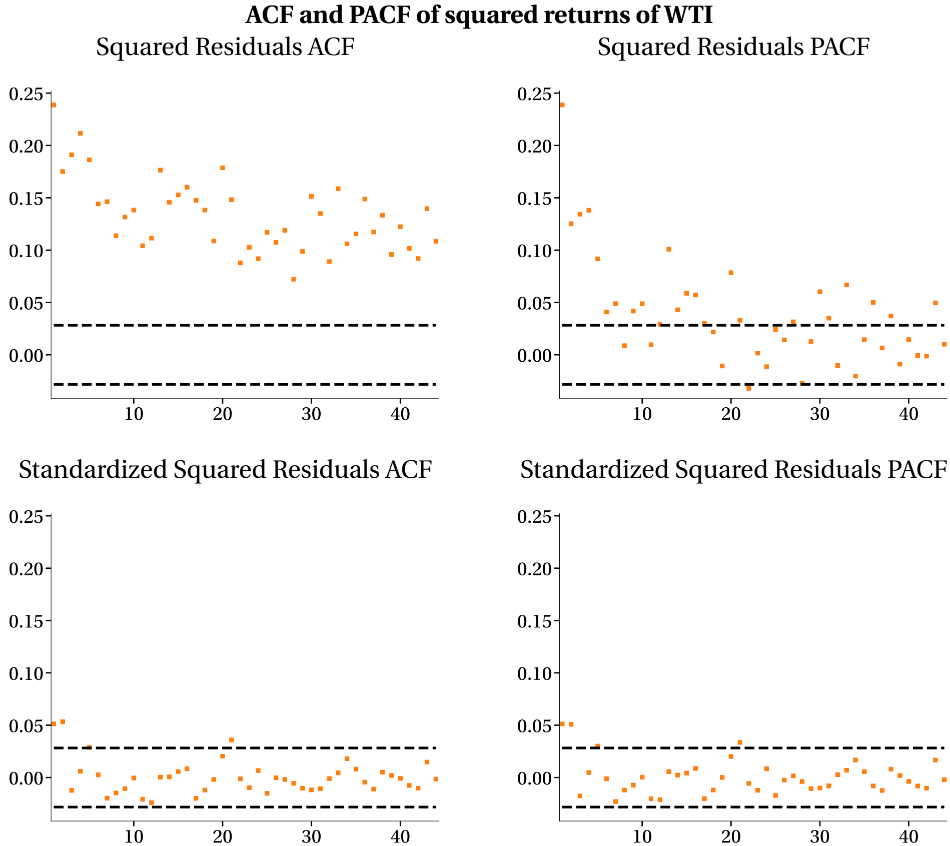

Figure 7.8: ACF and PACF of the squared returns for WTI. The bottom two panels contain the ACF and PACF of $\hat{e}_t^2 = \hat{\epsilon}_t^2/\hat{\sigma}_t^2$. The top panels show persistence in both the ACF and PACF which indicate an ARMA model is needed (hence a GARCH model) while the ACF and PACF of the standardized residuals appear to be compatible with an assumption of white noise. Compared to the S&P 500 ACF and PACF, WTI appears to slightly have weaker volatility dynamics.

---

and then computing a test statistic as T times the R^2 of the regression ($LM = T \times R^2$), and is asymptotically distributed χ^2_P where $\hat{\epsilon}_t$ are residuals from a conditional mean model. The null hypothesis is $H_0 : \phi_1 = \ldots = \phi_P = 0$ which corresponds to no persistence in the conditional variance.

	α_1	α_2	γ_1	γ_2	β_1	β_2	Log Lik.
GARCH(1,1)	0.059 (0.000)				0.934 (0.000)		−11030.1
GARCH(1,2)	0.075 (0.000)				0.585 (0.000)	0.331 (0.027)	−11027.4
GARCH(2,1)	0.059 (0.001)	0.000 (0.999)			0.934 (0.000)		−11030.1
GJR-GARCH(1,1,1)	0.026 (0.008)		0.049 (0.000)		0.945 (0.000)		−11011.9
GJR-GARCH(1,2,1)	0.026 (0.010)		0.049 (0.102)	0.000 (0.999)	0.945 (0.000)		−11011.9
TARCH(1,1,1)	0.030 (0.021)		0.055 (0.000)		0.942 (0.000)		−11005.6
TARCH(1,2,1)	0.030 (0.038)		0.055 (0.048)	0.000 (0.999)	0.942 (0.000)		−11005.6
TARCH(2,1,1)	0.030 (0.186)	0.000 (0.999)	0.055 (0.000)		0.942 (0.000)		−11005.6
EGARCH(1,0,1)	0.148 (0.000)				0.986 (0.000)		−11029.5
EGARCH(1,1,1)†	0.109 (0.000)		−0.050 (0.000)		0.990 (0.000)		−11000.6
EGARCH(1,2,1)	0.109 (0.000)		−0.056 (0.043)	0.006 (0.834)	0.990 (0.000)		−11000.5
EGARCH(2,1,1)*	0.195 (0.000)	−0.101 (0.019)	−0.049 (0.000)		0.992 (0.000)		−10994.4

Table 7.6: The models estimated in selecting a final model for the conditional variance of WTI. $\star$ indicates the selected model. † indicates a model that could be considered for model selection.

7.3 Forecasting Volatility

Forecasting conditional variances with GARCH models ranges from trivial for simple ARCH and GARCH specifications to difficult for non-linear specifications. Consider the simple ARCH(1) process,

$$\begin{aligned}\epsilon_t &= \sigma_t e_t \\ e_t &\overset{\text{i.i.d.}}{\sim} N(0,1) \\ \sigma_t^2 &= \omega + \alpha_1 \epsilon_{t-1}^2\end{aligned} \tag{7.76}$$

Iterating forward, $\sigma_{t+1}^2 = \omega + \alpha_1 \epsilon_t^2$, and taking conditional expectations, $\mathrm{E}_t[\sigma_{t+1}^2] = \mathrm{E}_t[\omega + \alpha_1 \epsilon_t^2] = \omega + \alpha_1 \epsilon_t^2$ since all of these quantities are known at time t. This is a property common to *all* ARCH-family models[16]: *forecasts of σ_{t+1}^2 are always known at time t.*

The 2-step ahead forecast follows from the law of iterated expectations,

$$\mathrm{E}_t[\sigma_{t+2}^2] = \mathrm{E}_t[\omega + \alpha_1 \epsilon_{t+1}^2] = \omega + \alpha_1 \mathrm{E}_t[\epsilon_{t+1}^2] = \omega + \alpha_1(\omega + \alpha_1 \epsilon_t^2) = \omega + \alpha_1\omega + \alpha_1^2 \epsilon_t^2. \tag{7.77}$$

[16]Not only is this property common to all ARCH-family members, it is the defining characteristic of an ARCH model.

A generic expression for a h-step ahead forecast can be constructed by repeatedly substitution and is given by

$$E_t[\sigma^2_{t+h}] = \sum_{i=0}^{h-1} \alpha_1^i \omega + \alpha_1^h \epsilon_t^2. \tag{7.78}$$

This form should look familiar since it is the multi-step forecasting formula for an AR(1). This should not be surprising since an ARCH(1) *is* an AR(1).

Forecasts from GARCH(1,1) models can be derived in a similar fashion,

$$\begin{aligned}
E_t[\sigma^2_{t+1}] &= E_t[\omega + \alpha_1 \epsilon_t^2 + \beta_1 \sigma_t^2] \\
&= \omega + \alpha_1 \epsilon_t^2 + \beta_1 \sigma_t^2 \\
E_t[\sigma^2_{t+2}] &= E_t[\omega + \alpha_1 \epsilon_{t+1}^2 + \beta_1 \sigma_{t+1}^2] \\
&= \omega + \alpha_1 E_t[\epsilon_{t+1}^2] + \beta_1 E_t[\sigma_{t+1}^2] \\
&= \omega + \alpha_1 E_t[e_{t+1}^2 \sigma_{t+1}^2] + \beta_1 E_t[\sigma_{t+1}^2] \\
&= \omega + \alpha_1 E_t[e_{t+1}^2] E_t[\sigma_{t+1}^2] + \beta_1 E_t[\sigma_{t+1}^2] \\
&= \omega + \alpha_1 \cdot 1 \cdot E_t[\sigma_{t+1}^2] + \beta_1 E_t[\sigma_{t+1}^2] \\
&= \omega + \alpha_1 E_t[\sigma_{t+1}^2] + \beta_1 E_t[\sigma_{t+1}^2] \\
&= \omega + (\alpha_1 + \beta_1) E_t[\sigma_{t+1}^2]
\end{aligned} \tag{7.79}$$

and substituting the one-step ahead forecast, $E_t[\sigma^2_{t+1}]$, produces

$$\begin{aligned}
E_t[\sigma^2_{t+2}] &= \omega + (\alpha_1 + \beta_1)(\omega + \alpha_1 \epsilon_t^2 + \beta_1 \sigma_t^2) \\
&= \omega + (\alpha_1 + \beta_1)\omega + (\alpha_1 + \beta_1)\alpha_1 \epsilon_t^2 + (\alpha_1 + \beta_1)\beta_1 \sigma_t^2
\end{aligned} \tag{7.80}$$

Note that $E_t[\sigma^2_{t+3}] = \omega + (\alpha_1 + \beta_1) E_t[\sigma^2_{t+2}]$, and so

$$\begin{aligned}
E_t[\sigma^2_{t+3}] &= \omega + (\alpha_1 + \beta_1)(\omega + (\alpha_1 + \beta_1)\omega + (\alpha_1 + \beta_1)\alpha_1 \epsilon_t^2 + (\alpha_1 + \beta_1)\beta_1 \sigma_t^2) \\
&= \omega + (\alpha_1 + \beta_1)\omega + (\alpha_1 + \beta_1)^2 \omega + (\alpha_1 + \beta_1)^2 \alpha_1 \epsilon_t^2 + (\alpha_1 + \beta_1)^2 \beta_1 \sigma_t^2.
\end{aligned} \tag{7.81}$$

Continuing in this manner produces a pattern which can be compactly expressed

$$E_t[\sigma^2_{t+h}] = \sum_{i=0}^{h-1} (\alpha_1 + \beta_1)^i \omega + (\alpha_1 + \beta_1)^{h-1} (\alpha_1 \epsilon_t^2 + \beta_1 \sigma_t^2). \tag{7.82}$$

Despite similarities to ARCH and GARCH models, forecasts from GJR-GARCH are less simple since the presence of the asymmetric term results in the probability that $e_{t-1} < 0$ appearing in the forecasting formula. If the standardized residuals were normal (or any other symmetric distribution), then the probability would be $\frac{1}{2}$. If the density is unknown, this quantity would need to be estimated from the standardized residuals.

In the GJR-GARCH model, the one-step ahead forecast is just

$$\mathrm{E}_t[\sigma^2_{t+1}] = \omega + \alpha_1\epsilon^2_t + \alpha_1\epsilon^2_t I_{[\epsilon_t<0]} + \beta_1\sigma^2_t \tag{7.83}$$

The two-step ahead forecast can be computed following

$$\begin{aligned}
\mathrm{E}_t[\sigma^2_{t+2}] &= \omega + \alpha_1\mathrm{E}_t[\epsilon^2_{t+1}] + \alpha_1\mathrm{E}_t[\epsilon^2_{t+1} I_{[\epsilon_{t+1}<0]}] + \beta_1\mathrm{E}_t[\sigma^2_{t+1}] && (7.84)\\
&= \omega + \alpha_1\mathrm{E}_t[\sigma^2_{t+1}] + \alpha_1\mathrm{E}_t[\epsilon^2_{t+1}|\epsilon_{t+1} < 0] + \beta_1\mathrm{E}_t[\sigma^2_{t+1}] && (7.85)
\end{aligned}$$

and if normality is assumed, $\mathrm{E}_t[\epsilon^2_{t+1}|\epsilon_{t+1} < 0] = \Pr(\epsilon_{t+1} < 0)\mathrm{E}[\sigma^2_{t+1}] = 0.5\mathrm{E}[\sigma^2_{t+1}]$ since the probability $\epsilon_{t+1} < 0$ is independent of $\mathrm{E}_t[\sigma^2_{t+1}]$.

Multi-step forecasts from other models in the ARCH-family, particularly those which are not linear combinations of ϵ^2_t, are nontrivial and generally do not have simple recursive formulas. For example, consider forecasting the variance from the simplest nonlinear ARCH-family member, a TARCH(1,0,0) model,

$$\sigma_t = \omega + \alpha_1|\epsilon_{t-1}| \tag{7.86}$$

As is *always* the case, the 1-step ahead forecast is known at time t,

$$\begin{aligned}
\mathrm{E}_t[\sigma^2_{t+1}] &= \mathrm{E}_t[(\omega + \alpha_1|\epsilon_t|)^2] && (7.87)\\
&= \mathrm{E}_t[\omega^2 + 2\omega\alpha_1|\epsilon_t| + \alpha_1^2\epsilon^2_t]\\
&= \omega^2 + 2\omega\alpha_1\mathrm{E}_t[|\epsilon_t|] + \alpha_1^2\mathrm{E}_t[\epsilon^2_t]\\
&= \omega^2 + 2\omega\alpha_1|\epsilon_t| + \alpha_1^2\epsilon^2_t
\end{aligned}$$

The 2-step ahead forecast is more complicated, and is given by

$$\begin{aligned}
\mathrm{E}_t[\sigma^2_{t+2}] &= \mathrm{E}_t[(\omega + \alpha_1|\epsilon_{t+1}|)^2] && (7.88)\\
&= \mathrm{E}_t[\omega^2 + 2\omega\alpha_1|\epsilon_{t+1}| + \alpha_1^2\epsilon^2_{t+1}]\\
&= \omega^2 + 2\omega\alpha_1\mathrm{E}_t[|\epsilon_{t+1}|] + \alpha_1^2\mathrm{E}_t[\epsilon^2_{t+1}]\\
&= \omega^2 + 2\omega\alpha_1\mathrm{E}_t[|e_{t+1}|\sigma_{t+1}] + \alpha_1^2\mathrm{E}_t[e^2_t\sigma^2_{t+1}]\\
&= \omega^2 + 2\omega\alpha_1\mathrm{E}_t[|e_{t+1}|]\mathrm{E}_t[\sigma_{t+1}] + \alpha_1^2\mathrm{E}_t[e^2_t]\mathrm{E}_t[\sigma^2_{t+1}]\\
&= \omega^2 + 2\omega\alpha_1\mathrm{E}_t[|e_{t+1}|](\omega + \alpha_1|\epsilon_t|) + \alpha_1^2\cdot 1\cdot(\omega^2 + 2\omega\alpha_1|\epsilon_t| + \alpha_1^2\epsilon^2_t)
\end{aligned}$$

The issues in multi-step ahead forecasting arise because the forecast depends on more than just $\mathrm{E}_t[e^2_{t+h}] \equiv 1$. In the above example, the forecast depends on both $\mathrm{E}_t[e^2_{t+1}] = 1$ and

$\mathrm{E}_t[|e_{t+1}|]$. When returns are normal $\mathrm{E}_t[|e_{t+1}|] == \sqrt{\frac{2}{\pi}}$ but if the driving innovations have a different distribution, this expectation will differ. The final form, assuming the conditional distribution is normal, is

$$\mathrm{E}_t[\sigma^2_{t+2}] = \omega^2 + 2\omega\alpha_1\sqrt{\frac{2}{\pi}}(\omega + \alpha_1|\epsilon_t|) + \alpha_1^2(\omega^2 + 2\omega\alpha_1|\epsilon_t| + \alpha_1^2\epsilon_t^2). \tag{7.89}$$

The difficulty in multi-step forecasting using "nonlinear" GARCH models follows directly from Jensen's inequality. In the case of TARCH,

$$\mathrm{E}_t[\sigma_{t+h}]^2 \neq \mathrm{E}_t[\sigma^2_{t+h}] \tag{7.90}$$

and in the general case

$$\mathrm{E}_t[\sigma^\delta_{t+h}]^{\frac{2}{\delta}} \neq \mathrm{E}_t[\sigma^2_{t+h}]. \tag{7.91}$$

7.3.1 Evaluating Volatility Forecasts

Evaluation of volatility forecasts is similar to the evaluation of forecasts from conditional mean models with one caveat. In standard time series models, once time $t+h$ has arrived, the value of the variable being forecast is known. However, in volatility models the value of σ^2_{t+h} is always unknown and it must be replaced with a proxy. The standard choice is to use the squared return, r_t^2. This is reasonable if the squared conditional mean is small relative to the variance, a reasonable assumption for short-horizon problems (daily or possibly weekly returns). If using longer horizon measurements of returns, such as monthly, squared residuals (ϵ_t^2) produced from a model for the conditional mean can be used instead. An alternative, and likely better choice if to use the *realized variance*, $RV_t^{(m)}$, to proxy for the unobserved volatility (see section 7.4). Once a choice of proxy has been made, Generalized Mincer-Zarnowitz regressions can be used to assess forecast optimality,

$$r^2_{t+h} - \hat{\sigma}^2_{t+h|t} = \gamma_0 + \gamma_1\hat{\sigma}^2_{t+h|t} + \gamma_2 z_{1t} + \ldots + \gamma_{K+1} z_{Kt} + \eta_t \tag{7.92}$$

where z_{jt} are any instruments known at time t. Common choices for z_{jt} include r_t^2, $|r_t|$, r_t or indicator variables for the sign of the lagged return. The GMZ regression in equation 7.92 has a heteroskedastic variance and that a better estimator, GMZ-GLS, can be constructed as

$$\frac{r^2_{t+h} - \hat{\sigma}^2_{t+h|t}}{\hat{\sigma}^2_{t+h|t}} = \gamma_0\frac{1}{\hat{\sigma}^2_{t+h|t}} + \gamma_1 1 + \gamma_2\frac{z_{1t}}{\hat{\sigma}^2_{t+h|t}} + \ldots + \gamma_{K+1}\frac{z_{Kt}}{\hat{\sigma}^2_{t+h|t}} + \nu_t \tag{7.93}$$

$$\frac{r^2_{t+h}}{\hat{\sigma}^2_{t+h|t}} - 1 = \gamma_0\frac{1}{\hat{\sigma}^2_{t+h|t}} + \gamma_1 1 + \gamma_2\frac{z_{1t}}{\hat{\sigma}^2_{t+h|t}} + \ldots + \gamma_{K+1}\frac{z_{Kt}}{\hat{\sigma}^2_{t+h|t}} + \nu_t \tag{7.94}$$

by dividing both sized by the time t forecast, $\hat{\sigma}^2_{t+h|t}$ where $\nu_t = \eta_t/\hat{\sigma}^2_{t+h|t}$. Equation 7.94 shows that the GMZ-GLS is a regression of the *generalized error* from a normal likelihood. If one were to use the realized variance as the proxy, the GMZ and GMZ-GLS regressions become

$$RV_{t+h} - \hat{\sigma}^2_{t+h|t} = \gamma_0 + \gamma_1 \hat{\sigma}^2_{t+h|t} + \gamma_2 z_{1t} + \ldots + \gamma_{K+1} z_{Kt} + \eta_t \tag{7.95}$$

and

$$\frac{RV_{t+h} - \hat{\sigma}^2_{t+h|t}}{\hat{\sigma}^2_{t+h|t}} = \gamma_0 \frac{1}{\hat{\sigma}^2_{t+h|t}} + \gamma_1 \frac{\hat{\sigma}^2_{t+h|t}}{\hat{\sigma}^2_{t+h|t}} + \gamma_2 \frac{z_{1t}}{\hat{\sigma}^2_{t+h|t}} + \ldots + \gamma_{K+1} \frac{z_{Kt}}{\hat{\sigma}^2_{t+h|t}} + \frac{\eta_t}{\hat{\sigma}^2_{t+h|t}} \tag{7.96}$$

Diebold-Mariano tests can also be used to assess the relative performance of two models. To perform a DM test, a loss function must be specified. Two obvious choices for the loss function are MSE,

$$\left(r^2_{t+h} - \hat{\sigma}^2_{t+h|t}\right)^2 \tag{7.97}$$

and QML-loss (which is simply the kernel of the normal log-likelihood),

$$\left(\ln(\hat{\sigma}^2_{t+h|t}) + \frac{r^2_{t+h}}{\hat{\sigma}^2_{t+h|t}}\right). \tag{7.98}$$

The DM statistic is a t-test of the null $H_0 : \mathrm{E}[\delta_t] = 0$ where

$$\delta_t = \left(r^2_{t+h} - \hat{\sigma}^2_{A,t+h|t}\right)^2 - \left(r^2_{t+h} - \hat{\sigma}^2_{B,t+h|t}\right)^2 \tag{7.99}$$

in the case of the MSE loss or

$$\delta_t = \left(\ln(\hat{\sigma}^2_{A,t+h|t}) + \frac{r^2_{t+h}}{\hat{\sigma}^2_{A,t+h|t}}\right) - \left(\ln(\hat{\sigma}^2_{B,t+h|t}) + \frac{r^2_{t+h}}{\hat{\sigma}^2_{B,t+h|t}}\right) \tag{7.100}$$

in the case of the QML-loss. Statistically significant positive values of $\bar{\delta} = R^{-1}\sum_{r=1}^{R} \delta_r$ indicate that B is a better model than A while negative values indicate the opposite (recall R is used to denote the number of out-of-sample observations used to compute the DM statistic). The QML-loss should be generally preferred as it is a "heteroskedasticity corrected" version of the MSE. For more on evaluation of volatility forecasts using MZ regressions see Patton and Sheppard (2009).

7.4 Realized Variance

Realized variance is a relatively new tool for measuring and modeling the conditional variance of asset returns that is novel for not requiring a model to measure the volatility, unlike ARCH models. Realized variance is a nonparametric estimator of the variance that is computed *using ultra high-frequency data.*[17]

Consider a log-price process, p_t, driven by a standard Wiener process with a constant mean and variance,

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

where the coefficients have been normalized such that the return during one day is the difference between p at two consecutive integers (i.e. $p_1 - p_0$ is the first day's return). For the S&P 500 index, $\mu \approx .00031$ and $\sigma \approx .0125$, which correspond to 8% and 20% for the annualized mean and volatility, respectively.

Realized variance is constructed by frequently sampling p_t throughout the trading day. Suppose that prices on day t were sampled on a regular grid of $m+1$ points, $0, 1, \ldots, m$ and let $p_{i,t}$ denote the i$^{\text{th}}$ observation of the log price. The m-sample realized variance is defined

$$RV_t^{(m)} = \sum_{i=1}^{m} \left(p_{i,t} - p_{i-1,t}\right)^2 = \sum_{i=1}^{m} r_{i,t}^2. \tag{7.101}$$

Since the price process is a standard Brownian motion, each return is an i.i.d. normal with mean μ/m, variance σ^2/m (and so a volatility of $\sigma/\sqrt{m}$). First, consider the expectation of $RV_t^{(m)}$,

$$\mathrm{E}\left[RV_t^{(m)}\right] = \mathrm{E}\left[\sum_{i=1}^{m} r_{i,t}^2\right] = \mathrm{E}\left[\sum_{i=1}^{m}\left(\frac{\mu}{m} + \frac{\sigma}{\sqrt{m}}\epsilon_{i,t}\right)^2\right] \tag{7.102}$$

where $\epsilon_{i,t}$ is a standard normal.

$$\begin{aligned}
\mathrm{E}\left[RV_t^{(m)}\right] &= \mathrm{E}\left[\sum_{i=1}^{m}\left(\frac{\mu}{m} + \frac{\sigma}{\sqrt{m}}\epsilon_{i,t}\right)^2\right] \qquad (7.103)\\
&= \mathrm{E}\left[\sum_{i=1}^{m}\frac{\mu^2}{m^2} + 2\frac{\mu\sigma}{m^{\frac{3}{2}}}\epsilon_{i,t} + \frac{\sigma^2}{m}\epsilon_{i,t}^2\right]\\
&= \mathrm{E}\left[\sum_{i=1}^{m}\frac{\mu^2}{m^2}\right] + \mathrm{E}\left[\sum_{i=1}^{m}2\frac{\mu\sigma}{m^{\frac{3}{2}}}\epsilon_{i,t}\right] + \mathrm{E}\left[\sum_{i=1}^{m}\frac{\sigma^2}{m}\epsilon_{i,t}^2\right]\\
&= \frac{\mu^2}{m} + \sum_{i=1}^{m}2\frac{\mu\sigma}{m^{\frac{3}{2}}}\mathrm{E}\left[\epsilon_{i,t}\right] + \sum_{i=1}^{m}\frac{\sigma^2}{m}\mathrm{E}\left[\epsilon_{i,t}^2\right]
\end{aligned}$$

[17] RV was invented somewhere between 1972 and 1997. However, its introduction to modern econometrics clearly dates to the late 1990s (Andersen and Bollerslev, 1998; Andersen, Bollerslev, Diebold, and Labys, 2003; Barndorff-Nielsen and Shephard, 2004).

$$
\begin{aligned}
&= \frac{\mu^2}{m} + 2\frac{\mu\sigma}{m^{\frac{3}{2}}}\sum_{i=1}^{m}\mathrm{E}\left[\epsilon_{i,t}\right] + \frac{\sigma^2}{m}\sum_{i=1}^{m}\mathrm{E}\left[\epsilon_{i,t}^2\right] \\
&= \frac{\mu^2}{m} + 2\frac{\mu\sigma}{m^{\frac{3}{2}}}\sum_{i=1}^{m} 0 + \frac{\sigma^2}{m}\sum_{i=1}^{m} 1 \\
&= \frac{\mu^2}{m} + \frac{\sigma^2}{m} m \\
&= \frac{\mu^2}{m} + \sigma^2
\end{aligned}
$$

The expected value is nearly σ^2, the variance, and, as $m \to \infty$, $\lim_{m\to\infty} \mathrm{E}\left[RV_t^{(m)}\right] = \sigma^2$. The variance of $RV_t^{(m)}$ can be similarly derived,

$$
\begin{aligned}
\mathrm{V}\left[RV_t^{(m)}\right] &= \mathrm{V}\left[\sum_{i=1}^{m}\frac{\mu^2}{m^2} + 2\frac{\mu\sigma}{m^{\frac{3}{2}}}\epsilon_{i,t} + \frac{\sigma^2}{m}\epsilon_{i,t}^2\right] \qquad (7.104)\\
&= \mathrm{V}\left[\sum_{i=1}^{m}\frac{\mu^2}{m^2}\right] + \mathrm{V}\left[\sum_{i=1}^{m}2\frac{\mu\sigma}{m^{\frac{3}{2}}}\epsilon_{i,t}\right] + \mathrm{V}\left[\sum_{i=1}^{m}\frac{\sigma^2}{m}\epsilon_{i,t}^2\right] + 2\mathrm{Cov}\left[\sum_{i=1}^{m}\frac{\mu^2}{m^2}, \sum_{i=1}^{m}2\frac{\mu\sigma}{m^{\frac{3}{2}}}\epsilon_{i,t}\right] \\
&\quad + 2\mathrm{Cov}\left[\sum_{i=1}^{m}\frac{\mu^2}{m^2}, \sum_{i=1}^{m}\frac{\sigma^2}{m}\epsilon_{i,t}^2\right] + 2\mathrm{Cov}\left[\sum_{i=1}^{m}2\frac{\mu\sigma}{m^{\frac{3}{2}}}\epsilon_{i,t}, \sum_{i=1}^{m}\frac{\sigma^2}{m}\epsilon_{i,t}^2\right].
\end{aligned}
$$

Working through these 6 terms, it can be determined that

$$
\mathrm{V}\left[\sum_{i=1}^{m}\frac{\mu^2}{m^2}\right] = \mathrm{Cov}\left[\sum_{i=1}^{m}\frac{\mu^2}{m^2}, \sum_{i=1}^{m}2\frac{\mu\sigma}{m^{\frac{3}{2}}}\epsilon_{i,t}\right] = \mathrm{Cov}\left[\sum_{i=1}^{m}\frac{\mu^2}{m^2}, \sum_{i=1}^{m}\frac{\sigma^2}{m}\epsilon_{i,t}^2\right] = 0
$$

since $\frac{\mu^2}{m^2}$ is a constant, and that the remaining covariance term also has expectation 0 since $\epsilon_{i,t}$ are i.i.d. standard normal and so have a skewness of 0,

$$
\mathrm{Cov}\left[\sum_{i=1}^{m}2\frac{\mu\sigma}{m^{\frac{3}{2}}}\epsilon_{i,t}, \sum_{i=1}^{m}\frac{\sigma^2}{m}\epsilon_{i,t}^2\right] = 0
$$

The other two terms can be shown to be (also left as exercises)

$$
\begin{aligned}
\mathrm{V}\left[\sum_{i=1}^{m}2\frac{\mu\sigma}{m^{\frac{3}{2}}}\epsilon_{i,t}\right] &= 4\frac{\mu^2\sigma^2}{m^2} \\
\mathrm{V}\left[\sum_{i=1}^{m}\frac{\sigma^2}{m}\epsilon_{i,t}^2\right] &= 2\frac{\sigma^4}{m}
\end{aligned}
$$

and so

$$\mathrm{V}\left[RV_t^{(m)}\right] = 4\frac{\mu^2\sigma^2}{m^2} + 2\frac{\sigma^4}{m}. \tag{7.105}$$

The variance is decreasing as $m \to \infty$ and $RV_t^{(m)}$ is a consistent estimator for σ^2.

In the more realistic case of a price process with time-varying drift and stochastic volatility,

$$d p_t = \mu_t \mathrm{d}t + \sigma_t \mathrm{d}W_t,$$

$RV_t^{(m)}$ is a consistent estimator of the *integrated variance*,

$$\lim_{m\to\infty} RV_t^{(m)} \xrightarrow{p} \int_t^{t+1} \sigma_s^2 \mathrm{d}s. \tag{7.106}$$

If the price process contains jumps, $RV_t^{(m)}$ is still a consistent estimator although its limit is the *quadratic variation* rather then the integrated variance, and so

$$\lim_{m\to\infty} RV_t^{(m)} \xrightarrow{p} \int_t^{t+1} \sigma_s^2 \mathrm{d}s + \sum_{t\le 1} \Delta J_s^2. \tag{7.107}$$

where ΔJ_s^2 measures the contribution of jumps to the total variation. Similar result hold if the price process exhibits leverage (instantaneous correlation between the price and the variance). The conditions for $RV_t^{(m)}$ to be a reasonable method to estimate the integrated variance on day t are essentially that the price process, p_t, is arbitrage-free and that the efficient price is observable. Data seem to support the first condition but the second is clearly violated.

There are always practical limits to m. This said, the maximum m is always much higher than 1 – a close-to-close return – and is typically somewhere between 13 (30-minute returns on the NYSE) and 390 (1-minute return on the NYSE), and so even when $RV_t^{(m)}$ is not consistent, it is still a better proxy, often substantially, for the latent variance on day t than r_t^2 (the "1-sample realized variance"). The signal-to-noise ratio (which measures the ratio of useful information to pure noise) is approximately 1 for RV but is between .05 and .1 for r_t^2. In other words, RV is 10-20 times more precise than squared daily returns. In terms of a linear regression, the difference between RV and daily squared returns is similar to fitting a regression with an R^2 of about 50% and one with an R^2 of 5% (Andersen and Bollerslev, 1998).

7.4.1 Modeling RV

If RV can reasonably be treated as observable, then it can be modeled using standard time series tools such as ARMA models. This has been the approach pursued thus far although there are issues in treating the RV as error free. The observability of the variance through RV is questionable and if there are errors in RV (which there almost certainly are), these will lead to an errors-in-variables problem in ARMA models and bias in estimated coefficients (see chapter 4).

In particular, Corsi (2009) have advocated the use of a *heterogeneous autoregressions* (HAR) where the current realized volatility depends on the realized volatility in the previous day, the average realized variance over the previous week, and the average realized variance of the previous month (22 days). HARs have been used in both levels

$$RV_t = \phi_0 + \phi_1 RV_{t-1} + \phi_5 \overline{RV}_{t-5} + \phi_2 2\overline{RV}_{t-22} + \epsilon_t \tag{7.108}$$

where $\overline{RV}_{t-5} = \frac{1}{5}\sum_{i=1}^{5} RV_{t-i}$ and $\overline{RV}_{t-22} = \frac{1}{22}\sum_{i=1}^{2} 2RV_{t-i}$ (suppressing the (m) terms), or in logs,

$$\ln RV_t = \phi_0 + \phi_1 \ln RV_{t-1} + \phi_5 \ln \overline{RV}_{t-5} + \phi_2 2 \ln \overline{RV}_{t-22} + \epsilon_t. \tag{7.109}$$

HARs allow for highly persistent volatility (through he 22-day moving average) and short term dynamics (through the 1-day and 5-day terms).

An alternative strategy is to apply ARCH-family models to the realized variance. ARCH-family models can be interpreted as a multiplicative error model for *any* non-negative process, not just squared returns (R. F. Engle, 2002a).[18] To use RV in an ARCH model, define $\tilde{r}_t = \text{sign}(r_t)\sqrt{RV_t}$ where $\text{sign}(r_t)$ takes the value 1 if the return was positive in period t and -1 if it was negative, and so $\tilde{r}_t$ is the signed square root of the realized variance on day t. Any ARCH-family model can be applied to these transformed realized variances to estimate a model for forecasting the quadratic variation. For example, consider the variance evolution in a GJR-GARCH(1,1,1) using the transformed data,

$$\sigma_t^2 = \omega + \alpha_1 \tilde{r}_{t-1}^2 + \gamma_1 \tilde{r}_{t-1}^2 I_{[\tilde{r}_{t-1}<0]} + \beta_1 \sigma_{t-1}^2 \tag{7.110}$$

which, in terms of realized variance, is equivalently expressed

$$\sigma_t^2 = \omega + \alpha_1 RV_{t-1} + \gamma_1 RV_{t-1} I_{[r_{t-1}<0]} + \beta_1 \sigma_{t-1}^2. \tag{7.111}$$

Maximum likelihood estimation, assuming normally distributed errors, can be used to estimate the parameters of this model. This procedure solves the errors-in-variables problem present when RV is treated as observable and makes modeling RV simple using standard estimators. Inference and model building is identical to that of ARCH processes.

7.4.2 Problems with RV

Realized variance suffers from a number of problems. The most pronounced is that observed prices are contaminated by noise since they are only observed at the bid and the ask. Consider a simple model of bid-ask bounce where returns are computed as the log difference in observed prices composed of the true (unobserved) efficient prices, $p_{i,t}^*$, contaminated by an independent mean zero shock, $\nu_{i,t}$.

[18] ARCH-family models have, for example, been successfully applied to both durations (time between trades) and hazards (number of trades in an interval of time), two non-negative processes.

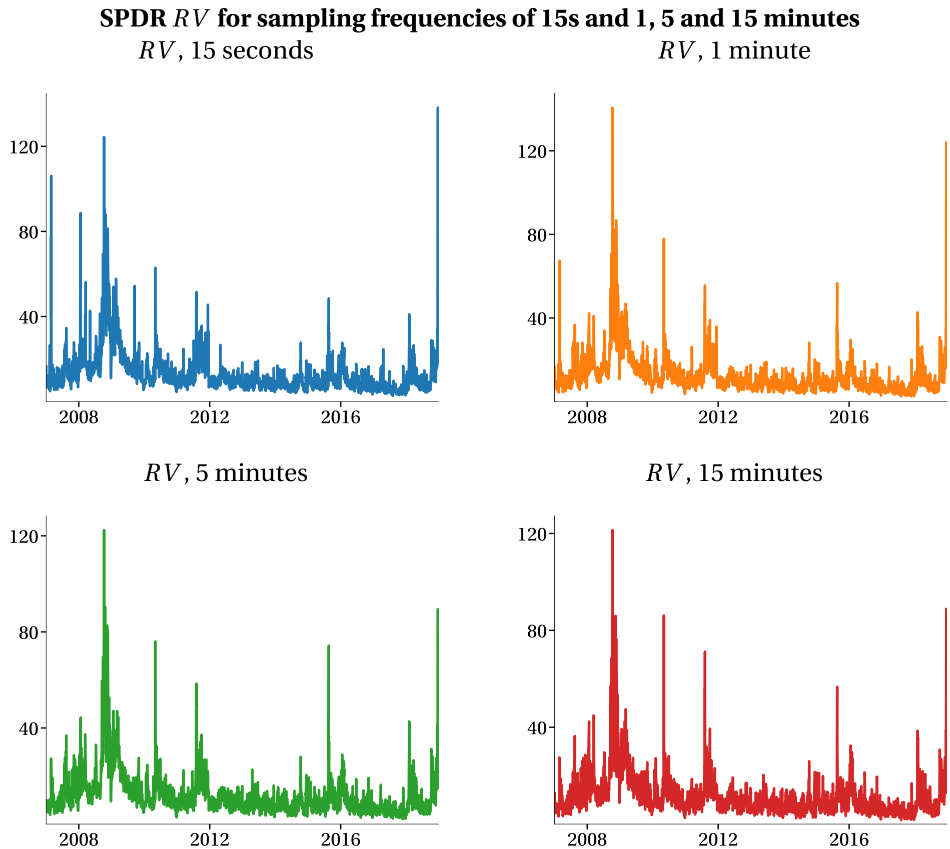

Figure 7.9: The four panels of this figure contain the Realized Variance for every day the market was open from January 2007 until December 2017. The estimated RV have been transformed into annualized volatility ($\sqrt{252 \cdot RV_t^{(m)}}$). While these plots appear superficially similar, the 1- and 5-minute RV are the most precise and the 15-second RV is biased upward.

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

The i$^{\text{th}}$ observed return, $r_{i,t}$ can be decomposed into the actual (unobserved) return $r^*_{i,t}$ and an independent noise term $\eta_{i,t}$.

$$\begin{aligned} p_{i,t} - p_{i-1,t} &= \left(p^*_{i,t} + \nu_{i,t}\right) - \left(p^*_{i-1,t} + \nu_{i-1,t}\right) \\ p_{i,t} - p_{i-1,t} &= \left(p^*_{i,t} - p^*_{i-1,t}\right) + \left(\nu_{i,t} - \nu_{i-1,t}\right) \end{aligned} \tag{7.112}$$

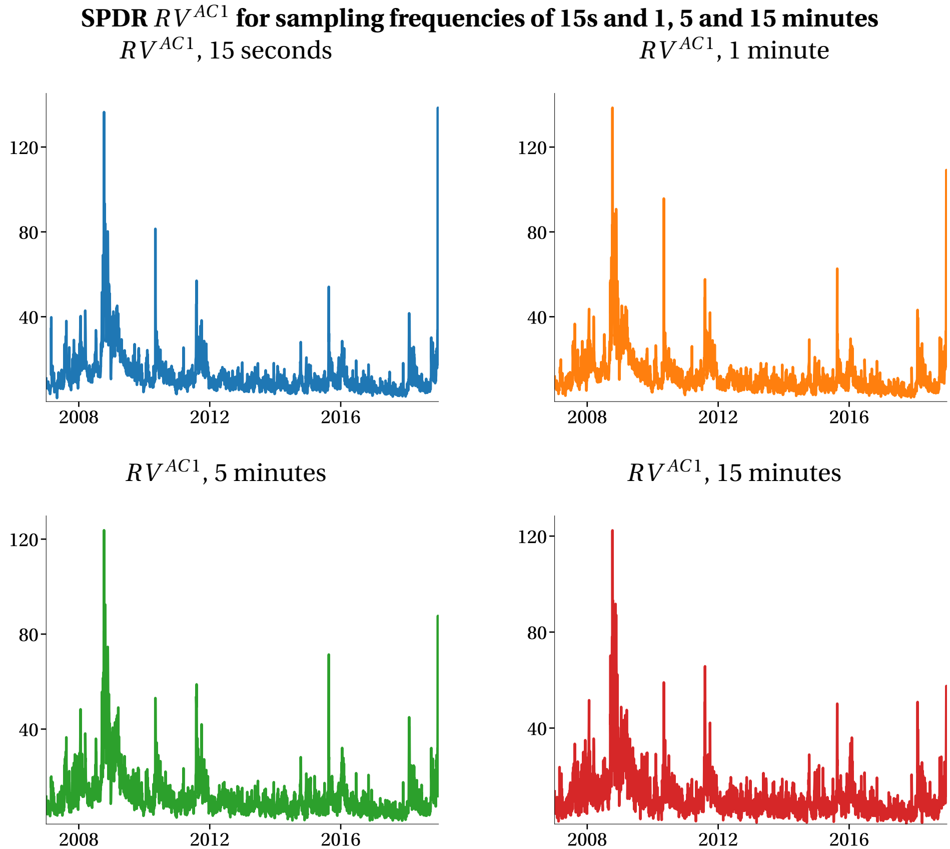

Figure 7.10: The four panels of this figure contain a noise robust version Realized Variance, RV^{AC1}, for every day the market was open from January 2007 until December 2017 transformed into annualized volatility. The 15-second RV^{AC1} is much better behaved than the 15-second RV although it still exhibits some strange behavior. In particular the negative spikes are likely due to errors in the data, a common occurrence in the TAQ database.

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

where $\eta_{i,t} = \nu_{i,t} - \nu_{i-1,t}$ is a MA(1) shock.

Computing the RV from returns contaminated by noise has an unambiguous effect on realized variance; RV is biased upward.

$$RV_t^{(m)} = \sum_{i=1}^{m} r^2_{i,t} \tag{7.113}$$

$$
\begin{aligned}
&= \sum_{i=1}^{m} (r_{i,t}^{*} + \eta_{i,t})^2 \\
&= \sum_{i=1}^{m} r_{i,t}^{*\,2} + 2 r_{i,t}^{*} \eta_{i,t} + \eta_{i,t}^2 \\
&\approx \widehat{RV}_t + m\tau^2
\end{aligned}
$$

where τ^2 is the variance of $\eta_{i,t}$ and $\widehat{RV}_t$ is the realized variance computed using the efficient returns. The bias is increasing in the number of samples (m), and may be large for stocks with large bid-ask spreads. There are a number of solutions to this problem. The simplest "solution" is to avoid the issue by not sampling prices relatively infrequently, which usually means limiting samples to somewhere between 1 to 30 minutes (see Bandi and Russell (2008)). Another simple, but again not ideal, method is to filter the data using an MA(1). Transaction data contain a strong negative MA due to bid-ask bounce, and so RV computed using the errors ($\hat{\epsilon}_{i,t}$) from a model,

$$
r_{i,t} = \theta \epsilon_{i-1,t} + \epsilon_{i,t} \tag{7.114}
$$

will eliminate much of the bias. A better method to remove the bias is to use an estimator known as RV^{AC1} which is similar to a Newey-West estimator.

$$
RV_t^{AC1(m)} = \sum_{i=1}^{m} r_{i,t}^2 + 2 \sum_{i=2}^{m} r_{i,t} r_{i-1,t} \tag{7.115}
$$

In the case of a constant drift, constant volatility Brownian motion subject to bid-ask bounce, this estimator can be shown to be unbiased, although it is not (quite) consistent. A more general class of estimators that use a kernel structure and which are consistent has been studied in Barndorff-Nielsen, P. R. Hansen, Lunde, and Shephard (2008).[19]

Another problem for realized variance is that prices are not available at regular intervals. This fortunately has a simple solution: *last price interpolation*. Last price interpolation sets the price at time t to the last observed price p_τ where τ is the largest time index less where $\tau \leq t$. Linearly interpolated prices set the time-t price to $p_t = w p_{\tau_1} + (1-w) p_{\tau_2}$ where τ_1 is the time subscript of the last observed price before t and τ_2 is the time subscript of the first price after time t, and the interpolation weight is $w = (\tau_2 - t)/(\tau_2 - \tau_1)$. When a transaction occurs at t, $\tau_1 = \tau_2$ and no interpolation is needed. Using a linearly interpolated price, which averages the prices around using the two closest observations weighted by the relative time to t (theoretically) produces a downward bias in RV.

Finally, most markets do not operate 24 hours a day and so the RV cannot be computed when markets are closed. The standard procedure is to use high-frequency returns when available and then use the close-to-open return squared to supplement this estimate.

[19]The Newey-West estimator is a special case of a broad class of estimators known as kernel estimators. They all share the property that they are weighted sums where the weights are determined by a kernel function.

Volatility Signature Plots

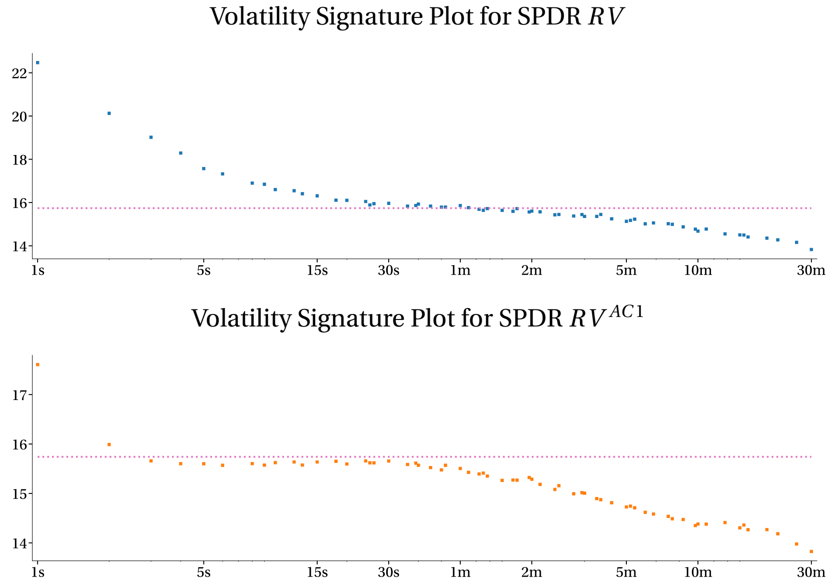

Figure 7.11: The volatility signature plot for the RV shows a clear trend. Based on visual inspection, it would be difficult to justify sampling more frequently than 1 minute. Unlike the volatility signature plot of the RV, the signature plot of RV^{AC1} does not monotonically increase with the sampling frequency except when sampling every second, and the range of the values is considerably smaller than in the RV signature plot. The decreases for the highest frequencies may be due to errors in the data which are more likely to show up when prices are sampled frequently.

$$RV_t^{(m)} = r_{\text{CtO},t}^2 + \sum_{i=1}^{m} \left(p_{i,t} - p_{i-1,t}\right)^2 = \sum_{i=1}^{m} r_{i,t}^2. \tag{7.116}$$

where $r_{\text{CtO},t}^2$ is the return between the close on day $t-1$ and the market open on day t. Since the overnight return is not measured frequently, the resulting RV must be treated as a random variable (and not an observable). An alternative method to handle the overnight return has been proposed in P. R. Hansen and Lunde (2005) and P. R. Hansen and Lunde (2006) which weighs the overnight squared return by λ_1 and the daily realized variance by λ_2 to produce an estimator with a lower mean-square error.

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

7.4.3 Realized Variance of the S&P 500

Returns on S&P 500 Depository Receipts, known as SPiDeRs (AMEX:SPY) will be used to illustrate the gains and pitfalls of RV. Price data was taken from TAQ and includes every transaction between January 2007 until December 2017, a total of 2,756 days. SPDRs track the S&P 500 and are among the most liquid assets in the U.S. market with an average volume of 60 million shares per day. During the 2.5 year sample there were over 100,000 trades on an average day – more than 4 per second when the market is open. TAQ data contain errors and so the data was filtered by matching the daily high and low from an audited database. Any prices out of these high-low bands or outside of the usual trading hours of 9:30 – 16:00 were discarded.

The primary tool of examining different Realized Variance estimators is the volatility signature plot.

Definition 7.10 (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.

Figures 7.9 and 7.10 contain plots of the *annualized volatility* constructed from the RV and RV^{AC1} for each day in the sample where estimates have been transformed into annualized volatility to facilitate comparison. Figures 7.9 shows that the 15-second RV is somewhat larger than the RV sampled at 1, 5 or 15 minutes and that the 1 and 5 minute RV are less noisy than the 15-minute RV. These plots provide some evidence that sampling more frequently than 15 minutes may be desirable. Comparing figure 7.10 to figure 7.9, there is a reduction in the scale of the 15-second RV^{AC1} relative to the 15-second RV. The 15-second RV is heavily influenced by the noise in the data (bid-ask bounce) while the RV^{AC1} is less affected. The negative values in the 15-second RV^{AC1} are likely due by errors in the recorded data which may contain a typo or a trade reported at the wrong time. This is a common occurrence in the TAQ database and frequent sampling increases the chances that one of the sampled prices is erroneous.

Figures 7.11 and 7.11 contain the volatility signature plot for RV and RV^{AC1}, respectively. These are presented in the original levels of RV to avoid averaging nonlinear transformations of random variables. The dashed horizontal line depicts the usual variance estimator computed from daily returns. There is a striking difference between the two figures. The RV volatility signature plot diverges when sampling more frequently than 3 minutes while the RV^{AC1} plot is non-monotone, probably due to errors in the price data. RV^{AC1} appears to allow sampling every 5 seconds, or nearly 10 times more frequently than RV. This is a common finding when comparing RV^{AC1} to RV across a range of equity data.

7.5 Implied Volatility and VIX

Implied volatility differs from other measures in that it is both market-based and forward-looking. Implied volatility was originally conceived as the "solution" to the Black-Scholes options pricing formula when everything except the volatility is known. Recall that the Black-Scholes formula is derived by assuming that stock prices follow a geometric Brownian motion plus drift,

$$dS_t = \mu S_t dt + \sigma S_t dW_t \tag{7.117}$$

where S_t is the time t stock prices, μ is the drift, σ is the (constant) volatility and dW_t is a Wiener process. Under some additional assumptions sufficient to ensure no arbitrage, the price of a call option can be shown to be

$$C(T,K) = S\Phi(d_1) + Ke^{-rT}\Phi(d_2) \tag{7.118}$$

where

$$d_1 = \frac{\ln(S/K) + \left(r + \sigma^2/2\right)T}{\sigma\sqrt{T}} \tag{7.119}$$

$$d_2 = \frac{\ln(S/K) + \left(r - \sigma^2/2\right)T}{\sigma\sqrt{T}}. \tag{7.120}$$

and where K is the strike price, T is the time to maturity – usually reported in years – r is the risk-free interest rate and $\Phi(\cdot)$ is the normal CDF. Given the stock price S_t, time to maturity T, the strike price K, the risk free rate, r, and the volatility, σ, the price of call options can be directly and calculated. Moreover, since the price of a call option is monotonic in the volatility, the formula can be inverted to express the volatility required to match a known call price. In this case, the implied volatility is

$$\sigma_t^{\text{Implied}} = g\left(C_t(T,K), S_t, K, T, r\right) \tag{7.121}$$

which is the expected volatility between t and T under the risk neutral measure (which is the same as under the physical when volatility is constant).[20]

7.5.1 The smile

Under normal conditions, the Black-Scholes implied volatility exhibits a "smile" (higher IV for out of the money than in the money) or "smirk" (higher IV for out of the money puts). The "smile" or "smirk" arises since actual returns are heavy tailed ("smile") and skewed ("smirk") although BSIV is derived under an assumption the prices follow a *geometric Brownian motion*, and so log returns are assumed to be normal. These patterns are reflecting the misspecification in the BSIV

[20] The value of the function can be determined by numerically inverting the B-S pricing formula.

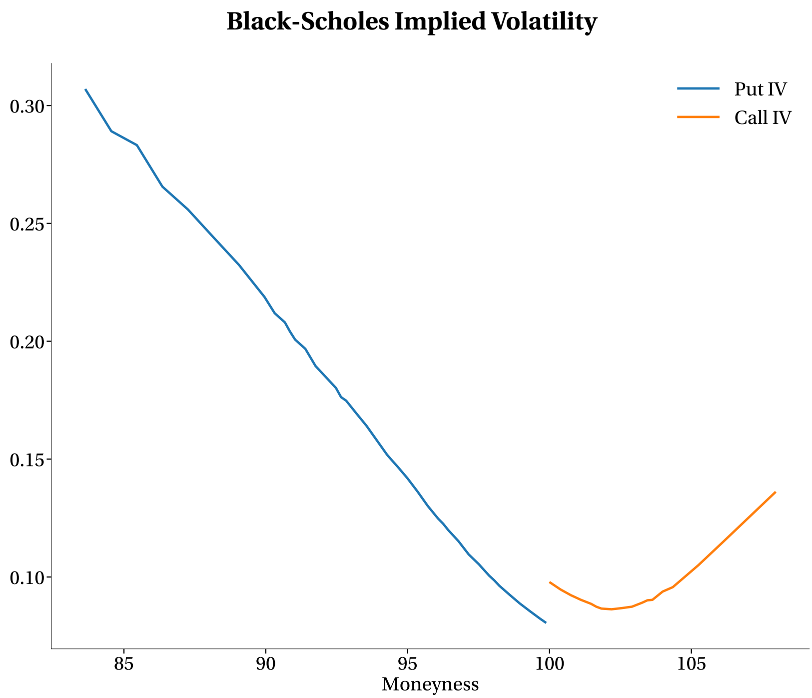

Figure 7.12: Plot of the Black-Scholes implied volatility "smile" on January 15, 2019 based on options on SPY expiring on February 2, 2018.

---

assumptions. Figure 7.12 contains the BSIV for SPY options on January 15, 2017 which exhibits the typical smile patters. Only out-of-the money options were used. The x-axis is expressed in terms of *moneyness* and so 100 indicates the current price, smaller values indicate strikes below the current price (out-of-the-money puts) and positive values the opposite (out-of-the-money calls).

7.5.2 Model-Free Volatility

B-S implied volatility suffers from a number of issues:

- Derived under constant volatility: The returns on most asset prices exhibit conditional heteroskedasticity, and time-variation in the volatility of returns will generate heavy tails which increases the chance of seeing a large stock price move.
- Leverage effects are ruled out: Leverage, or negative correlation between the price and volatility of a stock, can generate negative skewness. This would produce a larger probability of seeing a large downward movement in a stock price than the B-S model allows.

- No jumps: Jumps are also an empirical fact of most asset prices. Jumps, like time-varying volatility, increase the chance of seeing a large return.

The consequences of these limits are that, contrary to what the model underlying the B-S implies, B-S implied volatilities will not be constant across strike prices for a fixed maturity. Normally B-S implied volatilities follow a "smile", where options at-the-money (strike close to current price) will imply a lower volatility that options deep out of the money (strike larger then current price) and options deep in the money (strike lower then current price). In other times, the pattern of implied volatilities can take the shape of a "frown" – at the money implying a higher volatility than in- or out-of-the-money – or a "smirk", where the implied volatility of out-of-the-money options is higher than at- or in-the-money options. The various shapes the implied volatility curve can produce are all signs of misspecification of the B-S model. In particular they point to the three lacking features described above.

Model-free implied volatility has been developed as an alternative to B-S implied volatility by Britten-Jones and Neuberger (2000) with an important extension to jump processes and empirical implementation details provided by Jiang and Tian (2005). An important result highlighting the usefulness of options prices for computing expectations under the risk-neutral measures can be found in Breeden and Litzenberger (1978). Suppose that the risk-neutral measure $\mathbb{Q}$ exists and is unique. Then, under the risk neutral measure, it must be the case that

$$\frac{\partial S_t}{S_t} = \sigma(t,\cdot)\mathrm{d}W_t \tag{7.122}$$

is a martingale where $\sigma(t,\cdot)$ is a (possibly) time-varying volatility process that may depend on the stock price or other state variables. From the relationship, the price of a call option can be computed as

$$C(t,K) = \mathrm{E}_{\mathbb{Q}}\left[(S_t - K)^+\right] \tag{7.123}$$

for $t > 0$, $K > 0$ where the function $(x)^+ = \max(x,0)$. Thus

$$C(t,K) = \int_K^\infty (S_t - K)\,\phi_t(S_t)\mathrm{d}S_t \tag{7.124}$$

where $\phi_t(\cdot)$ is the risk-neutral measure. Differentiating with respect to K,

$$\frac{\partial C(t,K)}{\partial K} = -\int_K^\infty \phi_t(S_t)\mathrm{d}S_t \tag{7.125}$$

and differentiating this expression again with respect to K (note K in the lower integral bound)

$$\frac{\partial^2 C(t,K)}{\partial K^2} = \phi_t(K) \tag{7.126}$$

the risk neutral measure can be recovered from an options pricing function. This result provides a basis for non-parametrically estimating the risk-neutral density from observed options prices (see, for example Aït-Sahalia and Lo (1998)). Another consequence of this result is that the expected (under $\mathbb{Q}$) variation in a stock price over the interval $[t_1, t_2]$ measure can be recovered from

$$\mathrm{E}_{\mathbb{Q}}\left[\int_{t_1}^{t_2}\left(\frac{\partial S_t}{S_t}\right)^2\right] = 2\int_0^{\infty}\frac{C(t_2,K)-C(t_1,K)}{K^2}\mathrm{d}K. \tag{7.127}$$

This expression cannot be directly implemented to recover the expected volatility since it would require a continuum of strike prices.

Equation 7.127 assumes that the risk free rate is 0. When it is not, a similar result can be derived using the forward price

$$\mathrm{E}_{\mathbb{F}}\left[\int_{t_1}^{t_2}\left(\frac{\partial F_t}{F_t}\right)^2\right] = 2\int_0^{\infty}\frac{C^F(t_2,K)-C^F(t_1,K)}{K^2}\mathrm{d}K \tag{7.128}$$

where $\mathbb{F}$ is the forward probability measure – that is, the probability measure where the forward price is a martingale and $C^F(\cdot,\cdot)$ is used to denote that this option is defined on the forward price. Additionally, when t_1 is 0, as is usually the case, the expression simplifies to

$$\mathrm{E}_{\mathbb{F}}\left[\int_{0}^{t}\left(\frac{\partial F_t}{F_t}\right)^2\right] = 2\int_0^{\infty}\frac{C^F(t,K)-(F_0-K)^+}{K^2}\mathrm{d}K. \tag{7.129}$$

A number of important caveats are needed for employing this relationship to compute MFIV from option prices:

- Spot rather than forward prices. Because spot prices are usually used rather than forwards, the dependent variable needs to be redefined. If interest rates are non-stochastic, then define $B(0,T)$ to be the price of a bond today that pays \$1 time T. Thus, $F_0 = S_0/B(0,T)$, is the forward price and $C^F(T,K) = C(T,K)/B(0,T)$ is the forward option price. With the assumption of non-stochastic interest rates, the model-free implied volatility can be expressed

$$\mathrm{E}_{\mathbb{F}}\left[\int_{0}^{t}\left(\frac{\partial S_t}{S_t}\right)^2\right] = 2\int_0^{\infty}\frac{C(t,K)/B(0,T)-(S_0/B(0,T)-K)^+}{K^2}\mathrm{d}K \tag{7.130}$$

  or equivalently using a change of variables as

$$\mathrm{E}_{\mathbb{F}}\left[\int_{0}^{t}\left(\frac{\partial S_t}{S_t}\right)^2\right] = 2\int_0^{\infty}\frac{C(t,K/B(0,T))-(S_0-K)^+}{K^2}\mathrm{d}K. \tag{7.131}$$

- Discretization. Because only finitely many options prices are available, the integral must be approximated using a discrete grid. Thus the approximation

$$\mathrm{E}_{\mathbb{F}}\left[\int_0^t\left(\frac{\partial S_t}{S_t}\right)^2\right]=2\int_0^{\infty}\frac{C(t,K/B(0,T))-(S_0-K)^+}{K^2}\mathrm{d}K \tag{7.132}$$

$$\approx\sum_{m=1}^{M}[g(T,K_m)+g(T,K_{m-1})](K_m-K_{m-1}) \tag{7.133}$$

is used where

$$g(T,K)=\frac{C(t,K/B(0,T))-(S_0-K)^+}{K^2} \tag{7.134}$$

If the option tree is rich, this should not pose a significant issue. For option trees on individual firms, more study (for example using a Monte Carlo) may be needed to ascertain whether the MFIV is a good estimate of the volatility under the forward measure.

- Maximum and minimum strike prices. That the integral cannot be implemented from 0 to ∞ produces a downward bias in the implied volatility where the bias captures the missed volatility in the upper and lower tails. For rich options trees such as the S&P 500, this should not be a major issue.

7.5.3 VIX

The VIX – Volatility Index – is a volatility measure produced by the Chicago Board Options Exchange (CBOE). It is computed using a "model-free" like estimator which uses both call and put prices.[21] The VIX is computed according to

$$\sigma^2=\frac{2}{T}e^{rT}\sum_{i=1}^{N}\frac{\Delta K_i}{K_i^2}Q(K_i)-\frac{1}{T}\left(\frac{F_0}{K_0}-1\right)^2 \tag{7.135}$$

where T is the time to expiration of the options used, F_0 is the forward price which is computed from index option prices, K_i is the strike of the i$^{\text{th}}$ out-of-the-money option, $\Delta K_i=(K_{i+1}-K_{i-1})/2$ is half of the distance of the interval surrounding the option with a strike price of K_i, K_0 is the strike of the option immediately below the forward level, F_0, r is the risk free rate and $Q(K_i)$ is the mid-point of the bid and ask for the call or put used at strike K_i. The forward index price is extracted using put-call parity as $F_0=K_0+e^{rT}(C_0-P_0)$ where K_0 is the strike price where the price difference between put and call is smallest, and C_0 and P_0 are the respective call and put prices at this node.

The VIX uses options at two time horizons that attempt to bracket the 30-days (for example 15- and 45-days) with the caveat that options with 8 or fewer days to expiration are not used. This is to avoid pricing anomalies since the very short term options market is not liquid. Immediately subsequent to moving the time horizon, the time to expiration will not bracket 30-days (for

[21] The VIX is based exclusively on out-of-the-money prices, so calls are used for strikes above the current price and puts are used for strikes below the current price.

VIX and alternative measures of volatility

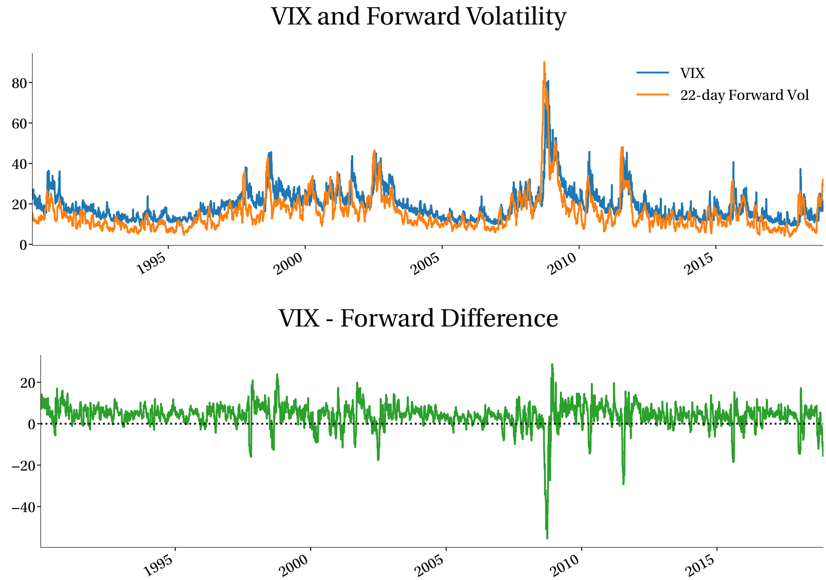

Figure 7.13: Plots of the VIX against a TARCH-based estimate of the volatility (top panel) and a 22-day forward moving average (bottom panel). The VIX is consistently above both measures reflecting the presence of a risk premium that compensates for time-varying volatility and/or jumps in the market return.

---

example, after a move the times to expiration may be 37 and 65 days). More details on the implementation of the VIX can be found in the CBOE whitepaper (CBOE, 2003).

7.5.4 Empirical Relationships

The daily VIX series from January 1990 until December 2017 is plotted in figure 7.13 against a 22-day *forward* moving average computed as

$$\sigma_t^{MA} = \sqrt{\frac{252}{22}\sum_{i=0^{2}1} r_{t+i}^2}.$$

The second panel shows the difference between these two series. The VIX is consistently, but not uniformly, higher than the forward volatility. This highlights both a feature and a drawback of using a measure of the volatility computed under the risk-neutral measure: it will contain a (possibly) time-varying risk premium. This risk premium will capture the propensity of the volatility to change (volatility of volatility) and any compensated jump risk.

7.A Kurtosis of an ARCH(1)

The necessary steps to derive the kurtosis of an ARCH(1) process are

$$
\begin{aligned}
\mathrm{E}[\epsilon_t^4] &= \mathrm{E}[\mathrm{E}_{t-1}[\epsilon_t^4]] \\
&= \mathrm{E}[3(\omega + \alpha_1 \epsilon_{t-1}^2)^2] \\
&= 3\mathrm{E}[(\omega + \alpha_1 \epsilon_{t-1}^2)^2] \\
&= 3\mathrm{E}[\omega^2 + 2\omega\alpha_1 \epsilon_{t-1}^2 + \alpha_1^2 \epsilon_{t-1}^4] \\
&= 3\left(\omega^2 + \omega\alpha_1 \mathrm{E}[\epsilon_{t-1}^2] + \alpha_1^2 \mathrm{E}[\epsilon_{t-1}^4]\right) \\
&= 3\omega^2 + 6\omega\alpha_1 \mathrm{E}[\epsilon_{t-1}^2] + 3\alpha_1^2 \mathrm{E}[\epsilon_{t-1}^4].
\end{aligned}
\tag{7.136}
$$

Using μ_4 to represent the expectation of the fourth power of ϵ_t ($\mu_4 = \mathrm{E}[\epsilon_t^4]$),

$$
\begin{aligned}
\mathrm{E}[\epsilon_t^4] - 3\alpha_1^2 \mathrm{E}[\epsilon_{t-1}^4] &= 3\omega^2 + 6\omega\alpha_1 \mathrm{E}[\epsilon_{t-1}^2] \\
\mu_4 - 3\alpha_1^2 \mu_4 &= 3\omega^2 + 6\omega\alpha_1 \bar{\sigma}^2 \\
\mu_4(1 - 3\alpha_1^2) &= 3\omega^2 + 6\omega\alpha_1 \bar{\sigma}^2 \\
\mu_4 &= \frac{3\omega^2 + 6\omega\alpha_1 \bar{\sigma}^2}{1 - 3\alpha_1^2} \\
\mu_4 &= \frac{3\omega^2 + 6\omega\alpha_1 \frac{\omega}{1-\alpha_1}}{1 - 3\alpha_1^2} \\
\mu_4 &= \frac{3\omega^2 (1 + 2\frac{\alpha_1}{1-\alpha_1})}{1 - 3\alpha_1^2} \\
\mu_4 &= \frac{3\omega^2(1 + \alpha_1)}{(1 - 3\alpha_1^2)(1 - \alpha_1)}.
\end{aligned}
\tag{7.137}
$$

This derivation makes use of the same principals as the intuitive proof and the identity that $\bar{\sigma}^2 = \omega/(1-\alpha_1)$. The final form highlights two important issues: first, μ_4 (and thus the kurtosis) is only finite if $1 - 3\alpha_1^2 > 0$ which requires that $\alpha_1 < \sqrt{\frac{1}{3}} \approx .577$, and second, the kurtosis, $\kappa = \frac{\mathrm{E}[\epsilon_t^4]}{\mathrm{E}[\epsilon_t^2]^2} = \frac{\mu_4}{\bar{\sigma}^2}$, is always greater than 3 since

$$
\begin{aligned}
\kappa &= \frac{\mathrm{E}[\epsilon_t^4]}{\mathrm{E}[\epsilon_t^2]^2} \\
&= \frac{\frac{3\omega^2(1+\alpha_1)}{(1-3\alpha_1^2)(1-\alpha_1)}}{\frac{\omega^2}{(1-\alpha_1)^2}}
\end{aligned}
\tag{7.138}
$$

$$
\begin{aligned}
&= \frac{3(1-\alpha_1)(1+\alpha_1)}{(1-3\alpha_1^2)} \\
&= \frac{3(1-\alpha_1^2)}{(1-3\alpha_1^2)} > 3.
\end{aligned}
$$

Finally, the variance of ϵ_t^2 can be computed noting that for any variable y, $\mathrm{V}[y] = \mathrm{E}[y^2] - \mathrm{E}[y]^2$, and so

$$
\begin{aligned}
\mathrm{V}[\epsilon_t^2] &= \mathrm{E}[\epsilon_t^4] - \mathrm{E}[\epsilon_t^2]^2 \\
&= \frac{3\omega^2(1+\alpha_1)}{(1-3\alpha_1^2)(1-\alpha_1)} - \frac{\omega^2}{(1-\alpha_1)^2} \\
&= \frac{3\omega^2(1+\alpha_1)(1-\alpha_1)^2}{(1-3\alpha_1^2)(1-\alpha_1)(1-\alpha_1)^2} - \frac{\omega^2(1-3\alpha_1^2)(1-\alpha_1)}{(1-3\alpha_1^2)(1-\alpha_1)(1-\alpha_1)^2} \\
&= \frac{3\omega^2(1+\alpha_1)(1-\alpha_1)^2 - \omega^2(1-3\alpha_1^2)(1-\alpha_1)}{(1-3\alpha_1^2)(1-\alpha_1)(1-\alpha_1)^2} \\
&= \frac{3\omega^2(1+\alpha_1)(1-\alpha_1) - \omega^2(1-3\alpha_1^2)}{(1-3\alpha_1^2)(1-\alpha_1)^2} \\
&= \frac{3\omega^2(1-\alpha_1^2) - \omega^2(1-3\alpha_1^2)}{(1-3\alpha_1^2)(1-\alpha_1)^2} \\
&= \frac{3\omega^2(1-\alpha_1^2) - 3\omega^2(\frac{1}{3}-\alpha_1^2)}{(1-3\alpha_1^2)(1-\alpha_1)^2} \\
&= \frac{3\omega^2[(1-\alpha_1^2) - (\frac{1}{3}-\alpha_1^2)]}{(1-3\alpha_1^2)(1-\alpha_1)^2} \\
&= \frac{2\omega^2}{(1-3\alpha_1^2)(1-\alpha_1)^2} \\
&= \left(\frac{\omega}{1-\alpha_1}\right)^2 \frac{2}{(1-3\alpha_1^2)} \\
&= \frac{2\bar{\sigma}^4}{(1-3\alpha_1^2)}
\end{aligned}
\tag{7.139}
$$

The variance of the squared returns depends on the unconditional level of the variance, $\bar{\sigma}^2$ and the innovation term (α_1) squared.

7.B Kurtosis of a GARCH(1,1)

First note that $\mathrm{E}[\sigma_t^2 - \epsilon_t^2] = 0$ so $\mathrm{V}[\sigma_t^2 - \epsilon_t^2]$ is just $\mathrm{E}[(\sigma_t^2 - \epsilon_t^2)^2]$. This can be expanded to $\mathrm{E}[\epsilon_t^4] - 2\mathrm{E}[\epsilon_t^2\sigma_t^2] + \mathrm{E}[\sigma_t^4]$ which can be shown to be $2\mathrm{E}[\sigma_t^4]$ since

$$\begin{aligned}
\mathrm{E}[\epsilon_t^4] &= \mathrm{E}[\mathrm{E}_{t-1}[e_t^4\sigma_t^4]] \\
&= \mathrm{E}[\mathrm{E}_{t-1}[e_t^4]\sigma_t^4] \\
&= \mathrm{E}[3\sigma_t^4] \\
&= 3\mathrm{E}[\sigma_t^4]
\end{aligned} \tag{7.140}$$

and

$$\begin{aligned}
\mathrm{E}[\epsilon_t^2\sigma_t^2] &= \mathrm{E}[\mathrm{E}_{t-1}[e_t^2\sigma_t^2]\sigma_t^2] \\
&= \mathrm{E}[\sigma_t^2\sigma_t^2] \\
&= \mathrm{E}[\sigma_t^4]
\end{aligned} \tag{7.141}$$

so

$$\begin{aligned}
\mathrm{E}[\epsilon_t^4] - 2\mathrm{E}[\epsilon_t^2\sigma_t^2] + \mathrm{E}[\sigma_t^4] &= 3\mathrm{E}[\sigma_t^4] - 2\mathrm{E}[\sigma_t^4] + \mathrm{E}[\sigma_t^4] \\
&= 2\mathrm{E}[\sigma_t^4]
\end{aligned} \tag{7.142}$$

The only remaining step is to complete the tedious derivation of the expectation of this fourth power,

$$\begin{aligned}
\mathrm{E}[\sigma_t^4] &= \mathrm{E}[(\sigma_t^2)^2] \\
&= \mathrm{E}[(\omega + \alpha_1\epsilon_{t-1}^2 + \beta_1\sigma_{t-1}^2)^2] \\
&= \mathrm{E}[\omega^2 + 2\omega\alpha_1\epsilon_{t-1}^2 + 2\omega\beta_1\sigma_{t-1}^2 + 2\alpha_1\beta_1\epsilon_{t-1}^2\sigma_{t-1}^2 + \alpha_1^2\epsilon_{t-1}^4 + \beta_1^2\sigma_{t-1}^4] \\
&= \omega^2 + 2\omega\alpha_1\mathrm{E}[\epsilon_{t-1}^2] + 2\omega\beta_1\mathrm{E}[\sigma_{t-1}^2] + 2\alpha_1\beta_1\mathrm{E}[\epsilon_{t-1}^2\sigma_{t-1}^2] + \alpha_1^2\mathrm{E}[\epsilon_{t-1}^4] + \beta_1^2\mathrm{E}[\sigma_{t-1}^4]
\end{aligned} \tag{7.143}$$

Noting that

- $\mathrm{E}[\epsilon_{t-1}^2] = \mathrm{E}[\mathrm{E}_{t-2}[\epsilon_{t-1}^2]] = \mathrm{E}[\mathrm{E}_{t-2}[e_{t-1}^2\sigma_{t-1}^2]] = \mathrm{E}[\sigma_{t-1}^2\mathrm{E}_{t-2}[e_{t-1}^2]] = \mathrm{E}[\sigma_{t-1}^2] = \bar{\sigma}^2$

- $\mathrm{E}[\epsilon_{t-1}^2\sigma_{t-1}^2] = \mathrm{E}[\mathrm{E}_{t-2}[\epsilon_{t-1}^2]\sigma_{t-1}^2] = \mathrm{E}[\mathrm{E}_{t-2}[e_{t-1}^2\sigma_{t-1}^2]\sigma_{t-1}^2] = \mathrm{E}[\mathrm{E}_{t-2}[e_{t-1}^2]\sigma_{t-1}^2\sigma_{t-1}^2] = \mathrm{E}[\sigma_t^4]$

- $E[\epsilon_{t-1}^4] = \mathrm{E}[\mathrm{E}_{t-2}[\epsilon_{t-1}^4]] = \mathrm{E}[\mathrm{E}_{t-2}[e_{t-1}^4\sigma_{t-1}^4]] = 3\mathrm{E}[\sigma_{t-1}^4]$

the final expression for $\mathrm{E}[\sigma_t^4]$ can be arrived at

$$\mathrm{E}[\sigma_t^4] = \omega^2 + 2\omega\alpha_1\mathrm{E}[\epsilon_{t-1}^2] + 2\omega\beta_1\mathrm{E}[\sigma_{t-1}^2] + 2\alpha_1\beta_1\mathrm{E}[\epsilon_{t-1}^2\sigma_{t-1}^2] + \alpha_1^2\mathrm{E}[\epsilon_{t-1}^4] + \beta_1^2\mathrm{E}[\sigma_{t-1}^4] \tag{7.144}$$

$$= \omega^2 + 2\omega\alpha_1\bar{\sigma}^2 + 2\omega\beta_1\bar{\sigma}^2 + 2\alpha_1\beta_1\mathrm{E}[\sigma_{t-1}^4] + 3\alpha_1^2\mathrm{E}[\sigma_{t-1}^4] + \beta_1^2\mathrm{E}[\sigma_{t-1}^4].$$

$\mathrm{E}[\sigma_t^4]$ can be solved for (replacing $\mathrm{E}[\sigma_t^4]$ with μ_4),

$$\mu_4 = \omega^2 + 2\omega\alpha_1\bar{\sigma}^2 + 2\omega\beta_1\bar{\sigma}^2 + 2\alpha_1\beta_1\mu_4 + 3\alpha_1^2\mu_4 + \beta_1^2\mu_4 \tag{7.145}$$

$$\mu_4 - 2\alpha_1\beta_1\mu_4 - 3\alpha_1^2\mu_4 - \beta_1^2\mu_4 = \omega^2 + 2\omega\alpha_1\bar{\sigma}^2 + 2\omega\beta_1\bar{\sigma}^2$$

$$\mu_4(1 - 2\alpha_1\beta_1 - 3\alpha_1^2 - \beta_1^2) = \omega^2 + 2\omega\alpha_1\bar{\sigma}^2 + 2\omega\beta_1\bar{\sigma}^2$$

$$\mu_4 = \frac{\omega^2 + 2\omega\alpha_1\bar{\sigma}^2 + 2\omega\beta_1\bar{\sigma}^2}{1 - 2\alpha_1\beta_1 - 3\alpha_1^2 - \beta_1^2}$$

finally substituting $\bar{\sigma}^2 = \omega/(1-\alpha_1-\beta_1)$ and returning to the original derivation,

$$\mathrm{E}[\epsilon_t^4] = \frac{3(1+\alpha_1+\beta_1)}{(1-\alpha_1-\beta_1)(1-2\alpha_1\beta_1-3\alpha_1^2-\beta_1^2)} \tag{7.146}$$

and the kurtosis, $\kappa = \frac{\mathrm{E}[\epsilon_t^4]}{\mathrm{E}[\epsilon_t^2]^2} = \frac{\mu_4}{\bar{\sigma}^2}$, which simplifies to

$$\kappa = \frac{3(1+\alpha_1+\beta_1)(1-\alpha_1-\beta_1)}{1-2\alpha_1\beta_1-3\alpha_1^2-\beta_1^2} > 3. \tag{7.147}$$

Short Problems

Problem 7.1. What is Realized Variance and why is it useful?

Problem 7.2. Suppose $r_t = \sigma_t\epsilon_t$ where $\sigma_t^2 = \omega + \alpha r_{t-1}^2 + \beta\sigma_{t-1}^2$, and $\epsilon_t \overset{\text{i.i.d.}}{\sim} N(0,1)$. What conditions are required on the parameters ω, α and β for r_t to be covariance stationary?

Problem 7.3. What is realized variance?

Problem 7.4. Discuss any properties the generalized error should have when evaluating volatility models.

Exercises

Exercise 7.1. Suppose we model log-prices at time t, written p_t, is an ARCH(1) process

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

where $\mathcal{F}_t$ denotes the information up to and including time t and

$$\sigma_t^2 = \alpha + \beta\left(p_{t-1} - p_{t-2}\right)^2.$$

1. Is p_t a martingale?
2. What is
$$\mathrm{E}\left[\sigma_t^2\right]?$$
3. Calculate
$$\mathrm{Cov}\left[\left(p_t - p_{t-1}\right)^2, \left(p_{t-s} - p_{t-1-s}\right)^2\right]$$
for $s > 0$.
4. Comment on the importance of this result from a practical perspective.
5. How do you use a likelihood function to estimate an ARCH model?
6. How can the ARCH(1) model be generalized to be more empirically realistic models of the innovations in price processes in financial economics?
7. In the ARCH(1) case, what can you find out about the properties of
$$p_{t+s}|\mathcal{F}_{t-1},$$
for $s > 0$, i.e. the multistep ahead forecast of prices?
8. Why are Bollerslev-Wooldridge standard errors important when testing coefficients in ARCH models?

Exercise 7.2. Derive explicit relationships between the parameters of an APARCH(1,1,1) and

1. ARCH(1)
2. GARCH(1,1)
3. AVGARCH(1,1)
4. TARCH(1,1,1)
5. GJR-GARCH(1,1,1)

Exercise 7.3. Consider the following GJR-GARCH process,

$$\begin{aligned}
r_t &= \rho r_{t-1} + \epsilon_t \\
\epsilon_t &= \sigma_t e_t \\
\sigma_t^2 &= \omega + \alpha\epsilon_{t-1}^2 + \gamma\epsilon_{t-1}^2 I_{[\epsilon_{t-1}<0]} + \beta\sigma_{t-1}^2 \\
e_t &\overset{\text{i.i.d.}}{\sim} N(0,1)
\end{aligned}$$

where $\mathrm{E}_t[\cdot] = \mathrm{E}[\cdot|\mathcal{F}_t]$ is the time t conditional expectation and $\mathrm{V}_t[\cdot] = \mathrm{V}[\cdot|\mathcal{F}_t]$ is the time t conditional variance.

1. What conditions are necessary for this process to be covariance stationary?

   Assume these conditions hold in the remaining questions. *Note*: If you cannot answer one or more of these questions for an arbitrary γ, you can assume that $\gamma = 0$ and receive partial credit.

2. What is $\mathrm{E}[r_{t+1}]$?

3. What is $\mathrm{E}_t[r_{t+1}]$?

4. What is $\mathrm{V}[r_{t+1}]$?

5. What is $\mathrm{V}_t[r_{t+1}]$?

6. What is $\mathrm{V}_t[r_{t+2}]$?

Exercise 7.4. Let r_t follow a GARCH process

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

1. What are the values of the following quantities?

   (a) $\mathrm{E}[r_{t+1}]$

   (b) $\mathrm{E}_t[r_{t+1}]$

   (c) $\mathrm{V}[r_{t+1}]$

   (d) $\mathrm{V}_t[r_{t+1}]$

   (e) ρ_1

2. What is $\mathrm{E}[(r_t^2 - \bar{\sigma}^2)(r_{t-1}^2 - \bar{\sigma}^2)]$

3. Describe the h-step ahead forecast from this model.

Exercise 7.5. Let r_t follow a ARCH process

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

1. What are the values of the following quantities?

(a) $\mathrm{E}[r_{t+1}]$

(b) $\mathrm{E}_t[r_{t+1}]$

(c) $\mathrm{V}[r_{t+1}]$

(d) $\mathrm{V}_t[r_{t+1}]$

(e) ρ_1

2. What is $\mathrm{E}[(r_t^2 - \bar{\sigma}^2)(r_{t-1}^2 - \bar{\sigma}^2)]$ Hint: Think about the AR duality.

3. Describe the h-step ahead forecast from this model.

Exercise 7.6. Consider an EGARCH(1,1,1) model:

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

where $e_t \overset{\text{i.i.d.}}{\sim} N(0,1)$.

1. What are the important conditions for stationarity in this process?

2. What is the one-step-ahead forecast of σ_t^2 ($\mathrm{E}_t\left[\sigma_{t+1}^2\right]$)?

3. What is the most you can say about the two-step-ahead forecast of σ_t^2 ($\mathrm{E}_t\left[\sigma_{t+2}^2\right]$)?

Exercise 7.7. Answer the following questions:

1. Describe three fundamentally different procedures to estimate the volatility over some interval. What the strengths and weaknesses of each?

2. Why is Realized Variance useful when evaluating a volatility model?

3. What considerations are important when computing Realized Variance?

4. Why is does the Black-Scholes implied volatility vary across strikes?

Exercise 7.8. Consider a general volatility specification for an asset return r_t :

$$\begin{aligned} r_t|\mathcal{F}_{t-1} &\sim N\left(0, \sigma_t^2\right) \\ \text{and let } e_t &\equiv \frac{r_t}{\sigma_t} \\ \text{so } e_t|\mathcal{F}_{t-1} &\sim iid\ N(0,1) \end{aligned}$$

1. Find the conditional kurtosis of the returns:

$$\mathrm{Kurt}_{t-1}[r_t] \equiv \frac{\mathrm{E}_{t-1}\left[(r_t - \mathrm{E}_{t-1}[r_t])^4\right]}{(\mathrm{V}_{t-1}[r_t])^2}$$

2. Show that if $V\left[\sigma_t^2\right] > 0$, then the *unconditional* kurtosis of the returns,

$$\text{Kurt}\left[r_t\right] \equiv \frac{\text{E}\left[\left(r_t - \text{E}\left[r_t\right]\right)^4\right]}{\left(\text{V}\left[r_t\right]\right)^2}$$

is greater than 3.

3. Find the conditional skewness of the returns:

$$\text{Skew}_{t-1}\left[r_t\right] \equiv \frac{\text{E}_{t-1}\left[\left(r_t - \text{E}_{t-1}\left[r_t\right]\right)^3\right]}{\left(\text{V}_{t-1}\left[r_t\right]\right)^{3/2}}$$

4. Find the *unconditional* skewness of the returns:

$$\text{Skew}\left[r_t\right] \equiv \frac{\text{E}\left[\left(r_t - \text{E}\left[r_t\right]\right)^3\right]}{\left(\text{V}\left[r_t\right]\right)^{3/2}}$$

Exercise 7.9. Answer the following questions:

1. Describe three fundamentally different procedures to estimate the volatility over some interval. What are the strengths and weaknesses of each?

2. Why does the Black-Scholes implied volatility vary across strikes?

3. Consider the following GJR-GARCH process,

$$\begin{aligned}
r_t &= \mu + \rho r_{t-1} + \epsilon_t \\
\epsilon_t &= \sigma_t e_t \\
\sigma_t^2 &= \omega + \alpha\epsilon_{t-1}^2 + \gamma\epsilon_{t-1}^2 I_{[\epsilon_{t-1}<0]} + \beta\sigma_{t-1}^2 \\
e_t &\overset{\text{i.i.d.}}{\sim} N(0,1)
\end{aligned}$$

where $\text{E}_t[\cdot] = \text{E}[\cdot|\mathcal{F}_t]$ is the time t conditional expectation and $\text{V}_t[\cdot] = \text{V}[\cdot|\mathcal{F}_t]$ is the time t conditional variance.

   (a) What conditions are necessary for this process to be covariance stationary?

   Assume these conditions hold in the remaining questions.

   (b) What is $\text{E}[r_{t+1}]$?

   (c) What is $\text{E}_t[r_{t+1}]$?

   (d) What is $\text{E}_t[r_{t+2}]$?

   (e) What is $\text{V}[r_{t+1}]$?

   (f) What is $\text{V}_t[r_{t+1}]$?

(g) What is $V_t[r_{t+2}]$?

Exercise 7.10. Answer the following questions about variance estimation.

1. What is Realized Variance?

2. How is Realized Variance estimated?

3. Describe two models which are appropriate for modeling Realized Variance.

4. What is an Exponential Weighted Moving Average (EWMA)?

5. Suppose an ARCH model for the conditional variance of daily returns was fit

$$\begin{aligned} r_{t+1} &= \mu + \sigma_{t+1} e_{t+1} \\ \sigma_{t+1}^2 &= \omega + \alpha_1 \epsilon_t^2 + \alpha_2 \epsilon_{t-1}^2 \\ e_t &\stackrel{\text{i.i.d.}}{\sim} N(0,1) \end{aligned}$$

   What are the forecasts for $t+1$, $t+2$ and $t+3$ given the current (time t) information set?

6. Suppose an EWMA was used instead for the model of conditional variance with smoothing parameter = .94. What are the forecasts for $t+1$, $t+2$ and $t+3$ given the current (time t) information set?

7. Compare the ARCH(2) and EWMA forecasts when the forecast horizon is large (e.g. $E_t\left[\sigma_{t+h}^2\right]$ for large h).

8. What is VIX?

Exercise 7.11. Suppose $\{y_t\}$ is covariance stationary and can be described by the following process:

$$\begin{aligned} y_t &= \phi_1 y_{t-1} + \theta_1 \epsilon_{t-1} + \epsilon_t \\ \epsilon_t &= \sigma_t e_t \\ \sigma_t^2 &= \omega + \alpha \epsilon_{t-1}^2 \\ e_t &\stackrel{\text{i.i.d.}}{\sim} N(0,1) \end{aligned}$$

what are the values of the following quantities:

1. $E_t[y_{t+1}]$

2. $E_t[y_{t+2}]$

3. $\lim_{h\to\infty} E_t[y_{t+h}]$

4. $V_t[\epsilon_{t+1}]$

5. $\mathrm{V}_t\left[y_{t+1}\right]$

6. $\mathrm{V}_t\left[y_{t+2}\right]$

7. $\lim_{h\to\infty} \mathrm{V}_t\left[\epsilon_{t+h}\right]$

Exercise 7.12. Answer the following questions:

Suppose $\{y_t\}$ is covariance stationary and can be described by the following process:

$$
\begin{aligned}
y_t &= \phi_0 + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \epsilon_t \\
\epsilon_t &= \sigma_t e_t \\
\sigma_t^2 &= \omega + \alpha_1 \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2 \\
e_t &\overset{\text{i.i.d.}}{\sim} N(0,1)
\end{aligned}
$$

what are the values of the following quantities:

1. $\mathrm{E}_t\left[y_{t+1}\right]$

2. $\mathrm{E}_t\left[y_{t+2}\right]$

3. $\lim_{h\to\infty} \mathrm{E}_t\left[y_{t+h}\right]$

4. $\mathrm{V}_t\left[\epsilon_{t+1}\right]$

5. $\mathrm{V}_t\left[y_{t+2}\right]$

6. $\lim_{h\to\infty} \mathrm{V}_t\left[\epsilon_{t+h}\right]$

Exercise 7.13. Consider the AR(2)-ARCH(2) model

$$
\begin{aligned}
y_t &= \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t \\
\epsilon_t &= \sigma_t e_t \\
\sigma_t^2 &= \omega + \alpha_1 \epsilon_{t-1}^2 + \alpha_2 \epsilon_{t-2}^2 \\
e_t &\overset{\text{i.i.d.}}{\sim} N(0,1)
\end{aligned}
$$

1. What conditions are required for ϕ_0, ϕ_1 and ϕ_2 for the model to be covariance stationary?

2. What conditions are required for ω,α_1, α_2 for the model to be covariance stationary?

3. Show that $\{\epsilon_t\}$ is a white noise process.

4. Are ϵ_t and ϵ_{t-s} independent for $s \neq 0$?

5. What are the values of the following quantities:

   (a) $\mathrm{E}\left[y_t\right]$

(b) $\mathrm{E}_t\left[y_{t+1}\right]$

(c) $\mathrm{E}_t\left[y_{t+2}\right]$

(d) $\mathrm{V}_t\left[y_{t+1}\right]$

(e) $\mathrm{V}_t\left[y_{t+2}\right]$

Exercise 7.14. Suppose $\{y_t\}$ is covariance stationary and can be described by the following process:

$$
\begin{aligned}
y_t &= \phi_1 y_{t-1} + \theta_1 \epsilon_{t-1} + \epsilon_t \\
\epsilon_t &= \sigma_t e_t \\
\sigma_t^2 &= \omega + \alpha \epsilon_{t-1}^2 \\
e_t &\stackrel{\text{i.i.d.}}{\sim} N(0,1)
\end{aligned}
$$

1. What are the values of the following quantities:

   (a) $\mathrm{E}_t\left[y_{t+1}\right]$

   (b) $\mathrm{E}_t\left[y_{t+2}\right]$

   (c) $\lim_{h\rightarrow\infty} \mathrm{E}_t\left[y_{t+h}\right]$

   (d) $\mathrm{V}_t\left[\epsilon_{t+1}\right]$

   (e) $\mathrm{V}_t\left[y_{t+1}\right]$

   (f) $\mathrm{V}_t\left[y_{t+2}\right]$

   (g) $\mathrm{V}\left[y_{t+1}\right]$