Volatility Modeling Using Daily Data

FNCE 5321

Hang Bai

Overview

- In this Chapter, we will proceed with the univariate models in two steps.
- The first step is to establish a forecasting model for dynamic portfolio variance and to introduce methods for evaluating the performance of these forecasts.
- The second step is to consider ways to model nonnormal aspects of the portfolio return - aspects that are not captured by the dynamic variance.

Overview

- We proceed as follows:
  - We start with the simple variance forecasting and the RiskMetrics variance model.
  - We introduce the GARCH variance model and compare it with the RiskMetrics model.
  - We estimate the GARCH parameters using the quasi-maximum likelihood method.
  - We suggest extensions to the basic model
  - We discuss various methods for evaluating the volatility forecasting models.

Simple Variance Forecasting

- We define the daily asset log-return, R_{t+1}, using the daily closing price, S_{t+1} ,as

$$R_{t+1} \equiv \ln\left(S_{t+1}/S_t\right)$$

- R_{t+1} can refer to an individual asset return or a portfolio return.
- Based on findings of Chapter 1, we assume for short horizons the mean value of R_{t+1} is zero.
- Furthermore, we assume that the innovation to asset return is normally distributed, i.e.

$$R_{t+1} = \sigma_{t+1} z_{t+1}, \text{with } z_{t+1} \sim i.i.d.\ N(0,1)$$

Simple Variance Forecasting

- Where i.i.d. N(0,1) stands for "independently and identically normally distributed with mean equal to zero and variance equal to 1."
- Note that the normality assumption is not realistic.

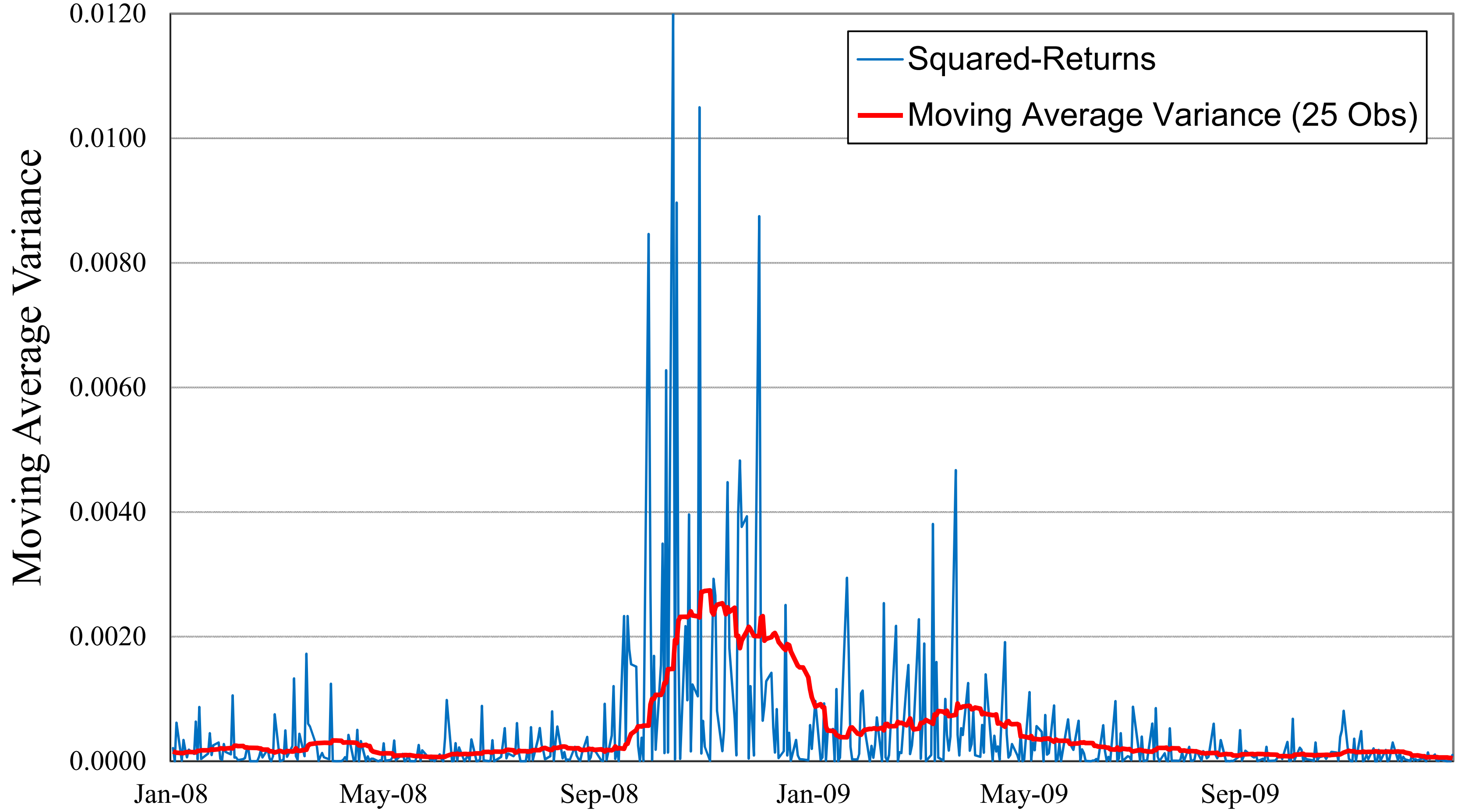
Simple Variance Forecasting

- Variance, as measured by squared returns, exhibits strong autocorrelation
- If the recent period was one of high variance, then tomorrow is likely to be a high-variance day as well.
- Tomorrow's variance is given by the simple average of the most recent *m* observations :

$$\sigma_{t+1}^2 = \frac{1}{m}\sum_{\tau=1}^{m} R_{t+1-\tau}^2 = \sum_{\tau=1}^{m}\frac{1}{m}R_{t+1-\tau}^2$$

- However model puts equal weights on the past *m* observations yielding unwarranted results

Figure 4.1:
Squared S&P 500 Returns with Moving Average Variance Estimated on past 25 observations. 2008-2009



Simple Variance Forecasting

- In RiskMetrics system, the weights on past squared returns decline exponentially as we move backward in time.
- JP Morgan's RiskMetrics variance model or the exponential smoother is given by:

$$\sigma_{t+1}^2 = (1-\lambda)\sum_{\tau=1}^{\infty}\lambda^{\tau-1}R_{t+1-\tau}^2, \text{ for } 0<\lambda<1$$

- Separating from the sum the squared return for $\tau = 1$, where $\lambda^{\tau-1} = \lambda^0 = 1$, we get

$$\sigma_{t+1}^2 = (1-\lambda)\sum_{\tau=2}^{\infty}\lambda^{\tau-1}R_{t+1-\tau}^2 + (1-\lambda)R_t^2$$

Simple Variance Forecasting

- Applying the exponential smoothing definition again we can write today's variance, σ_t^2, as

$$\sigma_t^2 = (1-\lambda)\sum_{\tau=1}^{\infty}\lambda^{\tau-1}R_{t-\tau}^2 = \frac{1}{\lambda}(1-\lambda)\sum_{\tau=2}^{\infty}\lambda^{\tau-1}R_{t+1-\tau}^2$$

- So that tomorrow's variance can be written

$$\sigma_{t+1}^2 = \lambda\sigma_t^2 + (1-\lambda)R_t^2$$

- The RiskMetrics model's forecast for tomorrow's volatility can thus be seen as weighted average of today's volatility and today's squared return.

Advantages of RiskMetrics

- It tracks variance changes in a way which is broadly consistent with observed returns. Recent returns matter more for tomorrow's variance than distant returns.
- It contains only one unknown parameter.
- When estimating λ on a large number of assets, Riskmetrics found that the estimates were quite similar across assets and they therefore simply set $\lambda = 0.94$ for every asset for daily variance forecasting.
- In this case, no estimation is necessary, which is a huge advantage in large portfolios.

Advantages of RiskMetrics

- Relatively little data needs to be stored in order to calculate tomorrow's variance.
- The weight on today's squared returns is $(1-\lambda)=0.06$ and the weight is exponentially decaying to $(1-\lambda)\lambda^{99}=0.000131$ on the 100th lag of squared return. After including 100 lags of squared returns the cumulated weight is

$$(1-\lambda)\sum_{\tau=1}^{100}\lambda^{\tau-1}=0.998$$

- We only need about 100 daily lags of returns in order to calculate tomorrow's variance, σ_{t+1}^2.

Advantages of RiskMetrics

- Despite these advantages, RiskMetrics does have certain shortcomings which motivates us to consider slightly more elaborate models.
- For example, it does not allow for a leverage effect and it also provides counterfactual longer-horizon forecasts.

The GARCH Variance Model

- This model can capture important features of returns data and are flexible enough to accommodate specific aspects of individual assets.
- The downside of the following models is that they require nonlinear parameter estimation
- The simplest generalized autoregressive conditional hetroskedasticity (GARCH) model of dynamic variance can be written as,

$$\sigma_{t+1}^2 = \omega + \alpha R_t^2 + \beta \sigma_t^2, \text{ with } \alpha + \beta < 1$$

The GARCH Variance Model

- The RiskMetrics model can be viewed as a special case of the simple GARCH model where
  $\alpha = 1 - \lambda, \beta = \lambda, \quad \text{s.t.}\ \alpha + \beta = 1, \omega = 0$
- However there is an important difference : We can define the unconditional, or long-run average, variance to be

$$\begin{aligned}
\sigma^2 &\equiv E[\sigma_{t+1}^2] = \omega + \alpha E[R_t^2] + \beta E[\sigma_t^2] \\
&= \omega + \alpha\sigma^2 + \beta\sigma^2, \text{ so that} \\
\sigma^2 &= \omega / (1 - \alpha - \beta)
\end{aligned}$$

The GARCH Variance Model

- If $\alpha + \beta = 1$ as is the case in RiskMetrics, then the long-run variance is not well-defined in that model.
- Thus an important quirk of the RiskMetrics model is that it ignores the fact that the long-run average variance tends to be relative stable over time.
- The GARCH model implicitly relies on σ^2
- By solving for ω in the long-run variance equation and substitute it into the dynamic variance equation, we get :

$$\sigma_{t+1}^2 = (1 - \alpha - \beta)\sigma^2 + \alpha R_t^2 + \beta\sigma_t^2 = \sigma^2 + \alpha(R_t^2 - \sigma^2) + \beta(\sigma_t^2 - \sigma^2)$$

The GARCH Variance Model

- Thus tomorrow's variance is a weighted average of the long-run variance, today's squared return and today's variance.
- Ignoring the long-run variance is more important for longer-horizon forecasting than for forecasting simply one-day ahead
- A key advantage of GARCH models for risk management is that the one-day forecast of variance , $\sigma^2_{t+1|t}$ is given directly by the model as σ^2_{t+1}

The GARCH Variance Model

- Consider forecasting the variance of the daily return *k* days ahead; the expected value of future variance at horizon *k* is

$$\begin{aligned} E_t\left[\sigma_{t+k}^2\right]-\sigma^2 &= \alpha E_t\left[R_{t+k-1}^2-\sigma^2\right]+\beta E_t\left[\sigma_{t+k-1}^2-\sigma^2\right] \\ &= \alpha E_t\left[\sigma_{t+k-1}^2 z_{t+k-1}^2-\sigma^2\right]+\beta E_t\left[\sigma_{t+k-1}^2-\sigma^2\right] \\ &= (\alpha+\beta)\left(E_t\left[\sigma_{t+k-1}^2\right]-\sigma^2\right), \text{ so that} \end{aligned}$$

$$E_t\left[\sigma_{t+k}^2\right]-\sigma^2 = (\alpha+\beta)^{k-1}\left(E_t\left[\sigma_{t+1}^2\right]-\sigma^2\right)$$

$$= (\alpha+\beta)^{k-1}\left(\sigma_{t+1}^2-\sigma^2\right)$$

The GARCH Variance Model

- The conditional expectation $E_t[\bullet]$, refers to taking the expectation using all the information available at the end of day *t*, which includes the squared return on day *t* itself.
- $(\alpha + \beta)$ is the persistence.
- A high persistence -$(\alpha + \beta)$ close to 1- implies that shocks that push variance away from its longrun average will persist for a long time
- Similar calculations for RiskMetrics model reveal

$$E_t\left[\sigma_{t+k}^2\right] = \sigma_{t+1}^2, \forall k$$

- as $\alpha + \beta = 1$ and σ^2 is undefined

The GARCH Variance Model

- Thus, persistence in this model is 1, which implies that a shock to variance persists forever
- An increase in variance will push up the variance forecast by an identical amount for all future forecast horizons.
- RiskMetrics model ignores the long-run variance when forecasting.
- If $\alpha + \beta$ is close to one, then the two models might yield similar predictions for short horizons, k, but their longer horizon implications are very different.

The GARCH Variance Model

- If today is a low-variance day then RiskMetrics model predicts that all future days will be low variance.
- The GARCH model assumes that eventually in the future variance will revert to the average value
- The forecast of variance of *K*-day cumulative returns

$$R_{t+1:t+K} \equiv \sum_{k=1}^{K} R_{t+k}$$

- We assume that returns have zero autocorrelation, then the variance is simply

$$\sigma^2_{t+1:t+K} \equiv E_t \left(\sum_{k=1}^{K} R_{t+k} \right)^2 = \sum_{k=1}^{K} E_t \left[\sigma^2_{t+k} \right]$$

The GARCH Variance Model

- So, in the RiskMetrics model we get

$$\sigma^2_{t+1:t+K} = K\sigma^2_{t+1}$$

- But in GARCH model, we get

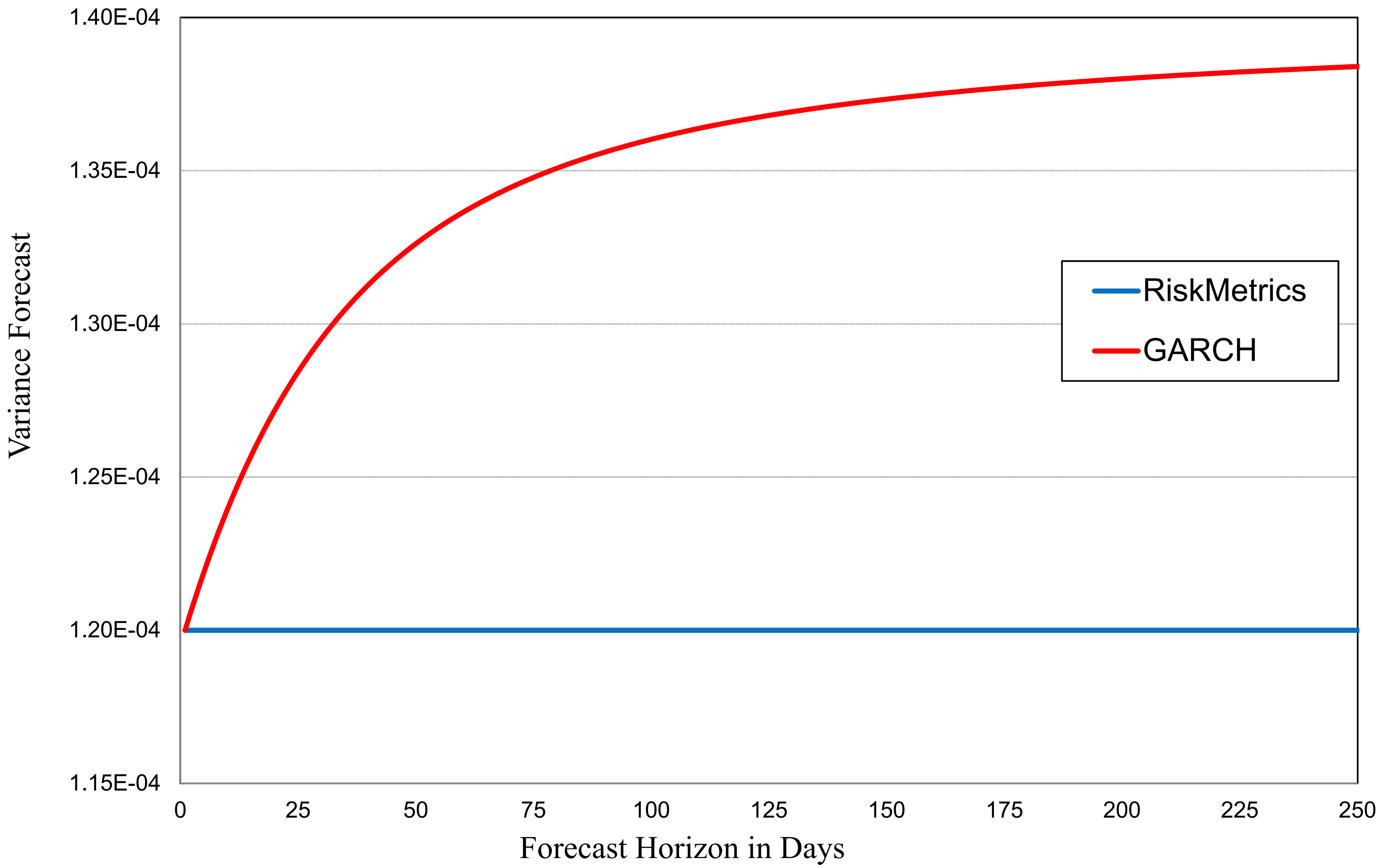
$$\sigma^2_{t+1:t+K} = K\sigma^2 + \sum_{k=1}^{K} (\alpha+\beta)^{k-1}\left(\sigma^2_{t+1} - \sigma^2\right) \neq K\sigma^2_{t+1}$$

- If the RiskMetrics and GARCH model have identical σ^2_{t+1}, and if, $\sigma^2_{t+1} < \sigma^2$, then the GARCH variance forecast will be higher than the RiskMetrics forecast.

The GARCH Variance Model

- Assuming Riskmetrics model, if the data looks more like GARCH will give risk managers a false sense of the calmness of the market in the future, when the market is calm today and $\sigma_{t+1}^2 < \sigma^2$.
- Fig.4.2 illustrates this crucial point.
- We plot $\sigma_{t+1:t+K}^2 / K$ for $K = 1,2,...,250$ for both the RiskMetrics and the GARCH model starting from a low σ^2_{t+1} and setting $\alpha = 0.05$ and $\beta = 0.90$.
- The long run variance in the figure is $\sigma^2 = 0.000140$

Figure 4.2:Variance Forecast for 1-250 Days Cumulative Returns



The GARCH Variance Model

- An inconvenience shared by the two models is that the multi-period distribution is unknown even if the one-day ahead distribution is assumed to be normal

- Thus while it is easy to forecast longer-horizon variance in these models, it is not as easy to forecast the entire conditional distribution.

- This issue will be further analyzed in Chap. 8

Maximum likelihood Estimation

- GARCH model contain a number of unknown parameters that must be estimated.
- The conditional variance is an unobserved variable, which must be implicitly estimated along with the parameters of the model.

Standard Maximum Likelihood Estimation

- MLE can be used to find parameter values
- Recall the assumption that

$$R_t = \sigma_t z_t, \text{ with } z_t \sim i.i.d.N(0,1)$$

- The assumption of i.i.d. normality implies that the probability, or the likelihood, l_t, of R_t is

$$l_t = \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{R_t^2}{2\sigma_t^2}\right)$$

- Thus the joint likelihood of our entire sample is

$$L = \prod_{t=1}^{T} l_t = \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{R_t^2}{2\sigma_t^2}\right)$$

Standard Maximum Likelihood Estimation

- Choose parameters $(\alpha, \beta, \ldots)$ to maximize the joint log likelihood of our observed sample

$$Max \ln L = Max \sum_{t=1}^{T} \ln(l_t) = Max \sum_{t=1}^{T} \left[-\frac{1}{2} \ln(2\pi) - \frac{1}{2} \ln\left(\sigma_t^2\right) - \frac{1}{2} \frac{R_t^2}{\sigma_t^2} \right]$$

- First term in the likelihood function is a constant and so independent of the parameters of the models.
- We can therefore equally well optimize

$$Max \sum_{t=1}^{T} \left[-\frac{1}{2} \ln\left(\sigma_t^2\right) - \frac{1}{2} \frac{R_t^2}{\sigma_t^2} \right] = Max \left[-\frac{1}{2} \left(\sum_{t=1}^{T} \ln\left(\sigma_t^2\right) + \frac{R_t^2}{\sigma_t^2} \right) \right]$$

Standard Maximum Likelihood Estimation

- The MLE approach has the desirable property that as the sample size, T, goes to infinity the parameter estimates converge to their true values.
- MLE gives the smallest variance for the estimates.
- In reality we don’t have infinite history of past data
- We may also have structural breaks.
- A good general rule of thumb is to use 1,000 daily observations when estimating GARCH.

Quasi Maximum Likelihood Estimation

- One may argue that the MLEs rely on the conditional normal distribution assumption which we argued in chapter 1 is false.
- A key result in econometrics says that even if the conditional distribution is not normal, MLE will yield estimates of the mean and variance parameters which converge to the true parameters as the sample gets infinitely large, as long as mean and variance functions are properly specified.
- This establishes the quasi maximum likelihood estimation or QMLE, referring to the use of normal-MLE estimation even when the normal distribution assumption is false.

Quasi Maximum Likelihood Estimation

- The QMLE estimates will in general be less precise than those from MLE.
- Thus we trade off theoretical asymptotic parameter efficiency for practicality.
- A simple trick than can be used in estimations is variance targeting.
- Recall the simple GARCH model

$$\sigma_{t+1}^2 = \omega + \alpha R_t^2 + \beta \sigma_t^2 = (1 - \alpha - \beta)\sigma^2 + \alpha R_t^2 + \beta \sigma_t^2$$

Quasi Maximum Likelihood Estimation

- Thus instead of estimating ω by MLE, we simply set the long-run variance, σ^2 ,equal to the sample variance

$$\sigma^2 = \frac{1}{T}\sum_{t=1}^{T} R_t^2$$

- Variance targeting imposes the long-run variance on the GARCH mode and reduces the number of parameters to be estimated in the model

An Example

- Fig. 4.3 shows the S&P500 squared returns from Fig. 4.1 but with an estimated GARCH variance superimposed

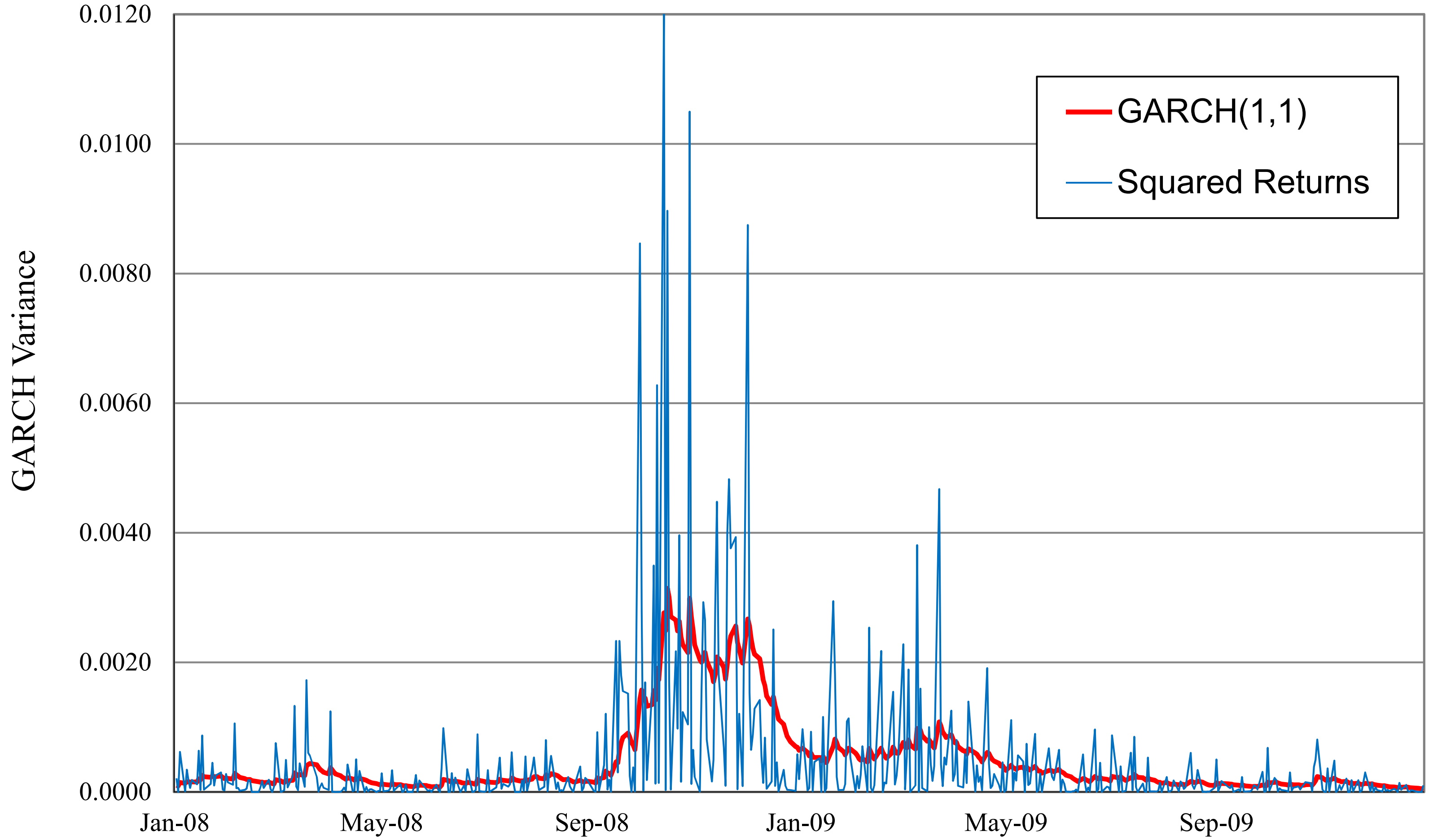
- Using numerical optimization of the likelihood function, the optimal parameters imply the following variance dynamics:

$$
\begin{aligned}
\sigma_{t+1}^2 &= \omega + \alpha R_t^2 + \beta \sigma_t^2 \\
&= 0.0000011 + 0.100 \cdot R_t^2 + 0.899 \cdot \sigma_t^2
\end{aligned}
$$

An Example

- The persistence of variance in this model is $\alpha + \beta = 0.999$, which is only slightly lower than in RiskMetrics where it is 1
- However, even if small, this difference will have consequences for the variance forecasts for horizons beyond one day
- Furthermore, this very simple GARCH model may be misspecified driving the persistence close to one
- So we consider more flexible models next

Figure 4.3: Squared S&P 500 Returns with GARCH Variance Parameters Are Estimated Using QMLE



The Leverage Effect

- A negative return increases variance by more than a positive return of the same magnitude
- This is referred to as the leverage effect
- We modify the GARCH models so that the weight given to the return depends on whether it is positive or negative, as follows:

$$\sigma_{t+1}^2 = \omega + \alpha\left(R_t - \theta\sigma_t\right)^2 + \beta\sigma_t^2 = \omega + \alpha\sigma_t^2\left(z_t - \theta\right)^2 + \beta\sigma_t^2$$

- which is sometimes referred to as the NGARCH (Nonlinear GARCH) model

The Leverage Effect

- The persistence of variance in this model is $\alpha(1+\theta^2)+\beta$ and the long-run variance is: $\sigma^2 = \omega/(1-\alpha(1+\theta^2)-\beta)$

- Another way of capturing the leverage effect is to define an indicator variable, I_t, to take on the value 1 if day t's return is negative and zero otherwise

$$I_t = \begin{cases} 1, \text{if } R_t < 0 \\ 0, \text{ if } R_t \geq 0 \end{cases}$$

- The variance dynamics can now be specified as

$$\sigma_{t+1}^2 = \omega + \alpha R_t^2 + \alpha\theta I_t R_t^2 + \beta\sigma_t^2$$

The Leverage Effect

- Thus, $\theta > 0$ will capture the leverage effect.
- This is referred to as the GJR-GARCH model.
- A different model that also captures the leverage is the exponential GARCH model or EGARCH

$$\ln \sigma_{t+1}^2 = \omega + \alpha\left(\phi R_t + \gamma\left[\left|R_t\right| - E\left|R_t\right|\right]\right) + \beta \ln \sigma_t^2$$

which displays the usual leverage effect if $\alpha\phi < 0$

- EGARCH model - Advantage : the log.specification ensures a positive variance
- Disadvantage : future expected variance beyond one period cannot be calculated analytically.

More General News Impact Functions

- Variance news impact function, *NIF is the* relationship in which today's shock to return, z_t, impacts tomorrow's variance σ^2_{t+1}
- In general we can write

$$\sigma^2_{t+1} = \omega + \alpha\sigma^2_t NIF\left(z_t\right) + \beta\sigma^2_t$$

- In the simple GARCH model we have

$$NIF\,(z_t) = z^2_t$$

- so that the NIF is a symmetric parabola that takes the minimum value 0 when z_t is zero
- In the NGARCH model with leverage we have

$$NIF\left(z_t\right) = \left(z_t - \theta\right)^2$$

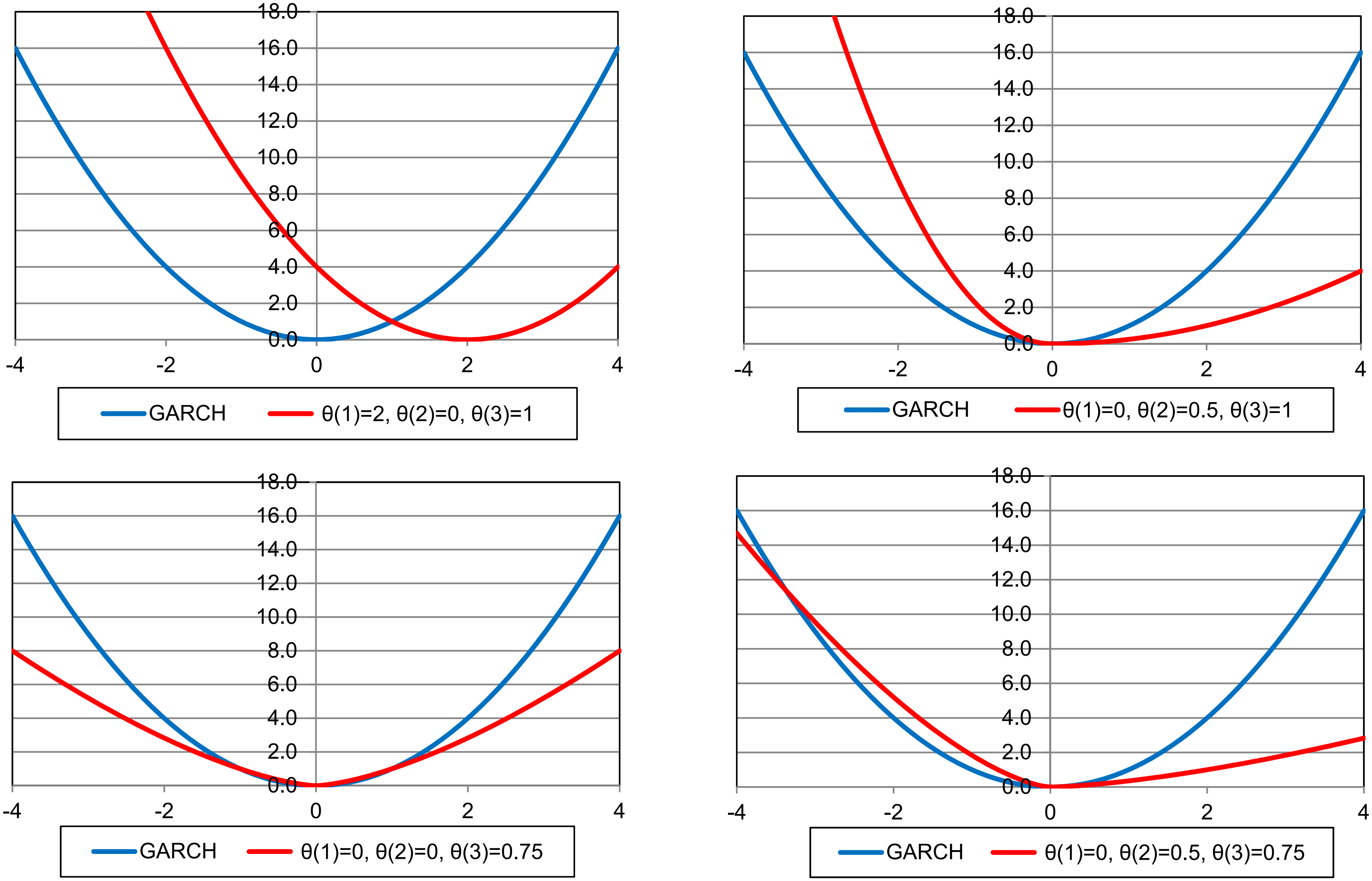
More General News Impact Functions

- so that the NIF is still a parabola but now with the minimum value zero when $z_t = \theta$
- A very general NIF can be defined by

$$NIF(z_t) = (|z_t - \theta_1| - \theta_2 (z_t - \theta_1))^{2\theta_3}$$

- The simple GARCH model is nested when $\theta_1 = \theta_2 = 0$, and $\theta_3 = 1$
- The NGARCH model with leverage is nested when $\theta_2 = 0$ and $\theta_3 = 1$.

Figure 4.4: News Impact Function



More General Dynamics

- A simple GARCH model GARCH(1,1) relies on only one lag of returns squared and one lag of variance.
- Higher order dynamics is made possible through GARCH(p,q) which allows for longer lags as follows:

$$\sigma_{t+1}^2 = \omega + \sum_{i=1}^{p} \alpha_i R_{t+1-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t+1-j}^2$$

- The disadvantage of this more generalized models is that the parameters are not easily interpretable.

More General Dynamics

- The component GARCH structure helps to interpret the parameters easily
- Using $\sigma^2 = \omega / (1 - \alpha - \beta)$ we can rewrite the GARCH(1,1) model as

$$\sigma_{t+1}^2 = \sigma^2 + \alpha \left(R_t^2 - \sigma^2\right) + \beta \left(\sigma_t^2 - \sigma^2\right)$$

- In the component GARCH model the long-run variance, s^2, is allowed to be time varying and captured by the long-run variance factor v_{t+1}:

$$\begin{aligned} \sigma_{t+1}^2 &= v_{t+1} + \alpha_\sigma \left(R_t^2 - v_t\right) + \beta_\sigma \left(\sigma_t^2 - v_t\right) \\ v_{t+1} &= \sigma^2 + \alpha_v \left(R_t^2 - \sigma_t^2\right) + \beta_v \left(v_t - \sigma^2\right) \end{aligned}$$

More General Dynamics

- Note that the dynamic long-term variance, v_{t+1}, itself has a GARCH(1,1) structure.
- Thus, a component GARCH model is a GARCH(1,1) model around another GARCH(1,1) model.
- The component model can potentially capture autocorrelation patterns in variance
- The component model can be rewritten as a GARCH(2,2) model as

$$\sigma_{t+1}^2 = \omega + \alpha_1 R_t^2 + \alpha_2 R_{t-1}^2 + \beta_1 \sigma_t^2 + \beta_2 \sigma_{t-1}^2$$

where the parameters in the GARCH(2,2) are functions of the parameters in the component GARCH model

More General Dynamics

- The component GARCH structure has the advantage that it is easier to interpret its parameters and therefore easier to come up with good starting values for the parameters than in the GARCH(2,2) model
- In the component model $\alpha_\sigma + \beta_\sigma$ capture the persistence of the short-run variance component and $\alpha_v + \beta_v$ capture the persistence in the long-run variance component.
- The GARCH(2,2) dynamic parameters α_1, α_2, β_1, β_2 have no such straightforward interpretation

Explanatory Variables

- In dynamic models of daily variance, we need to account for days with no trading activity
- Days that follow a weekend or a holiday have higher variance than average days
- As these days are perfectly predictable, we need to include them in the variance model
- So, we can model this by:

$$\sigma_{t+1}^2 = \omega + \beta\sigma_t^2 + \alpha\sigma_t^2 z_t^2 + \gamma IT_{t+1}$$

- where IT_{t+1} takes on the value 1 if date t+1 is a Monday, for example

Explanatory Variables

- In general, we can write the GARCH variance forecasting model as follows:

$$\sigma_{t+1}^2 = \omega + h(X_t) + \alpha\sigma_t^2 z_t^2 + \beta\sigma_t^2$$

- where X_t denotes variables known at the end of day t
- As the variance is always a positive number, the GARCH model should always generates a positive variance forecast
- In the above model, positivity of $h(X_t)$ along with positive ω, α and β will ensure positivity of σ^2_{t+1}
- We can write

$$\sigma_{t+1}^2 = \omega + \exp(b' X_t) + \alpha\sigma_t^2 z_t^2 + \beta\sigma_t^2$$

Generalizing the Low Frequency Dynamics

- Volatility often spikes up for a few days and then quickly reverts back down to normal levels.
- Such quickly reverting spikes make volatility appear noisy and thus difficult to capture by explanatory variables.
- Explanatory variables are important for capturing longer-term trends in variance, which need to be modeled separately so as to not be contaminated by the daily spikes.

Generalizing the Low Frequency Dynamics

- In order to capture low-frequency changes in volatility we generalize the simple GARCH(1,1) model to the following multiplicative structure

$$
\begin{aligned}
\sigma_{t+1}^2 &= \tau_{t+1} g_{t+1}, \text{ where} \\
g_{t+1} &= (1-\alpha-\beta) + \alpha g_t z_t^2 + \beta g_t, \text{ and} \\
\tau_{t+1} &= \omega_0 \exp\left(\omega_1 t + \omega_2 \max\left(t - t_0, 0\right)^2 + \gamma X_t\right)
\end{aligned}
$$

  - The Spline-GARCH model captures low frequency dynamics in variance via the τ_{t+1} process, and higher-frequency dynamics in variance via the g_{t+1} process.

Generalizing the Low Frequency Dynamics

- Low-frequency variance is kept positive via the exponential function

- The low frequency variance has a log linear time-trend captured by ω_1 and a quadratic time-trend starting at time t_0 and captured by ω_2

- The low-frequency variance is also driven by the explanatory variables in the vector X_t.

Generalizing the Low Frequency Dynamics

- The long-run variance in the Spline-GARCH model is captured by the low-frequency process

$$E\left[\sigma_{t+1}^2\right] = E\left[\tau_{t+1} g_{t+1}\right] = \tau_{t+1} E\left[g_{t+1}\right] = \tau_{t+1}$$

- We can generalize the quadratic trend by allowing for many, say *l*, quadratic pieces, each starting at different time points and each with different slope parameters:

$$\tau_{t+1} = \omega_0 \exp\left(\omega_1 t + \sum_{i=1}^{l} \omega_{1+i} \max\left(t - t_{i-1}, 0\right)^2 + \gamma X_t\right)$$

Estimation of Extended Models

- GARCH family of models can all be estimated using the same quasi MLE technique used for the simple GARCH(1,1) model.
- The model parameters can be estimated by maximizing the nontrivial part of the log likelihood

$$Max\left[-\frac{1}{2}\left(\sum_{t=1}^{T}\ln\left(\sigma_t^2\right)+\frac{R_t^2}{\sigma_t^2}\right)\right]$$

- The variance path, $\sigma^2{}_t$, is a function of the parameters to be estimated.

Model Comparison using LR Tests

- Basic GARCH model can be extended by adding parameters and explanatory variables
- The likelihood ratio test provides a simple way to judge if the added parameter(s) are significant in the statistical sense.
- Consider two different models with likelihood values L_0 and L_1, respectively.
- Assume that model 0 is a special case of model 1
- In this case we can compare the two models via the likelihood ratio statistic

$$LR = 2\left(\ln\left(L_1\right) - \ln\left(L_0\right)\right)$$

Model Comparison using LR Tests

- The *LR* statistic will be a positive number because model 1 contains model 0 as a special case and so model 1 will always fit the data better

- The LR statistic tells us if the improvement offered by model 1 over model 0 is statistically significant

- It can be shown that the *LR* statistic will have a chi-squared distribution under the null hypothesis that the added parameters in model 1 are insignificant.

Model Comparison using LR Tests

- If only one parameter is added then the degree of freedom in the chi-squared distribution will be 1

- A good rule of thumb is that if the log-likelihood of model 1 is 3 to 4 points higher than that of model 0 then the added parameter in model 1 is significant

- The degrees of freedom in the chi-squared test is equal to the number of parameters added in model 1

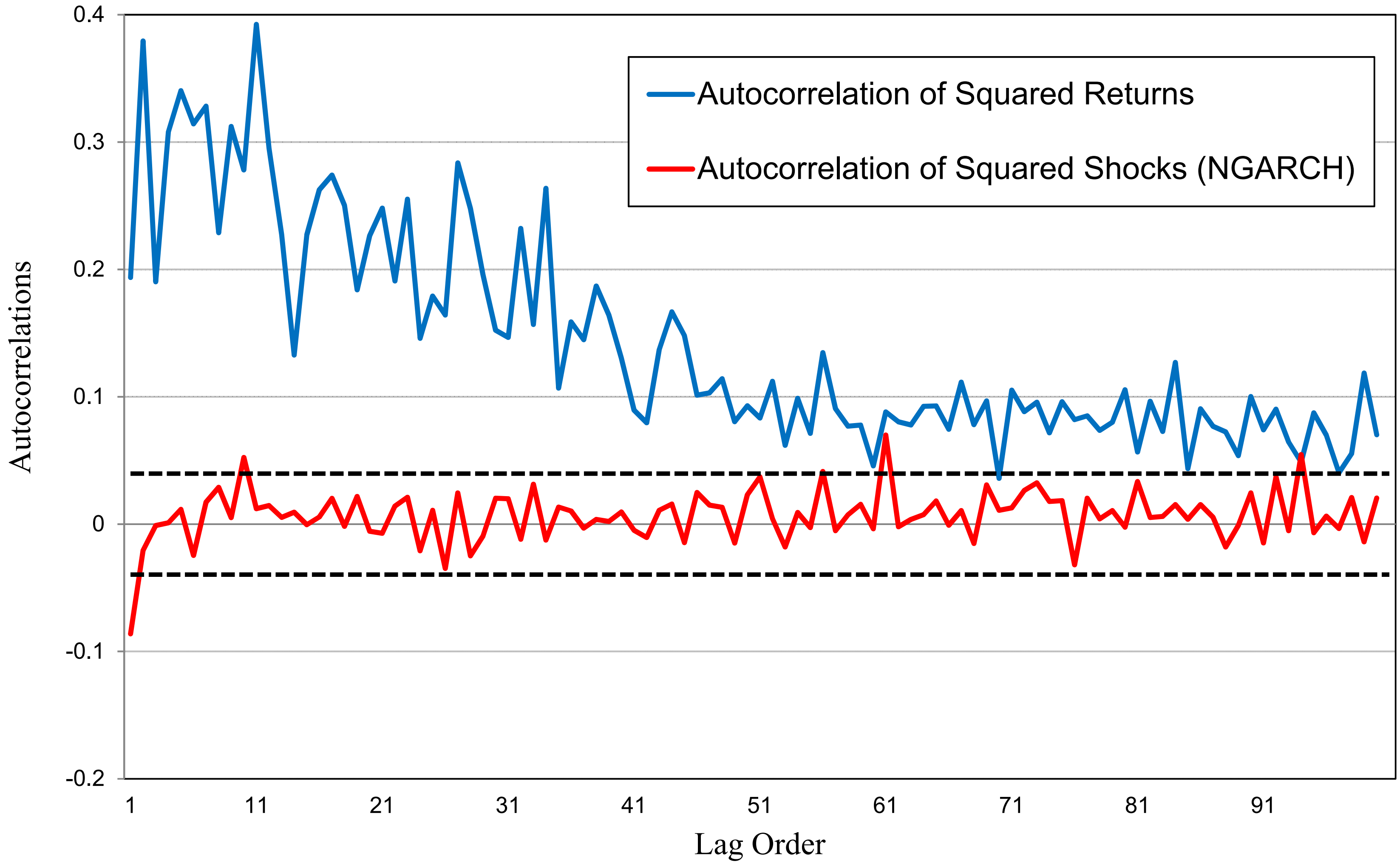
Diagnostic Check on Autocorrelations

- In Chapter 1 we saw that the raw return auto-correlations didn't display any systematic patterns
- The squared return autocorrelations is positive for short lags and decreases as the lag order increases
- We use variance modelling to construct σ^2_t which has the property that standardized squared returns, R^2_t / σ^2_t have no systematic autocorrelation patterns
- The red line in Figure 4.5, show the autocorrelation of R^2_t / σ^2_t from the GARCH model with leverage for the S&P 500 returns along with their standard error bands.

Diagnostic Check on Autocorrelations

- The standard errors are calculated simply as, $1/\sqrt{T}$ where T is the number of observations in the sample.
- Autocorrelation is shown along with plus/minus two standard error bands around zero, which mean horizontal lines at $-2/\sqrt{T}$ and $2/\sqrt{T}$
- These Bartlett standard error bands give the range in which the autocorrelations would fall roughly 95% of the time if the true but unknown autocorrelations of R^2_t/σ^2_t were all zero.

Figure 4.5: Autocorrelation: Squared Returns and Squared Returns over Variance



Volatility Forecast Evaluating Using Regression

- A variance model can be evaluated based on simple regressions where squared returns in the forecast period, t+1, are regressed on the forecast from the variance model, as in

$$R_{t+1}^2 = b_0 + b_1 \sigma_{t+1}^2 + e_{t+1}$$

- A good variance forecast should be unbiased, that is, have an intercept $b_0 = 0$, and be efficient, that is, have a slope, $b_1 = 1$.
- Note that $E_t[R_{t+1}^2] = \sigma_{t+1}^2$, so that the squared return is an unbiased proxy for true variance.

Volatility Forecast Evaluating Using Regression

- But the variance of the proxy is

$$\begin{aligned} Var_t[R^2_{t+1}] &= E_t[\left(R^2_{t+1} - \sigma^2_{t+1}\right)^2] = E_t[\left(\sigma^2_{t+1}(z^2_{t+1} - 1)\right)^2] \\ &= \sigma^4_{t+1} E_t[(z^2_{t+1} - 1)^2] = \sigma^4_{t+1}(\kappa - 1) \end{aligned}$$

- where κ is the kurtosis of the innovation
- Due to the high degree of noise in the squared returns, the regression R^2 will be very low, typically around 5% to 10%
- The conclusion is that the proxy for true but unobserved variance is simply very inaccurate.

The Volatility Forecast Loss Function

- The ordinary least squares estimation of a linear regression chooses the parameter values that minimize the mean squared error in the regression
- The regression-based approach to volatility forecast evaluation therefore implies a quadratic volatility forecast loss function
- A correct volatility forecasting model should have $b_0 = 0$ and $b_1 = 1$ as discussed earlier
- Loss function to compare volatility models is therefore

$$MSE = \left(R_{t+1}^2 - \sigma_{t+1}^2\right)^2$$

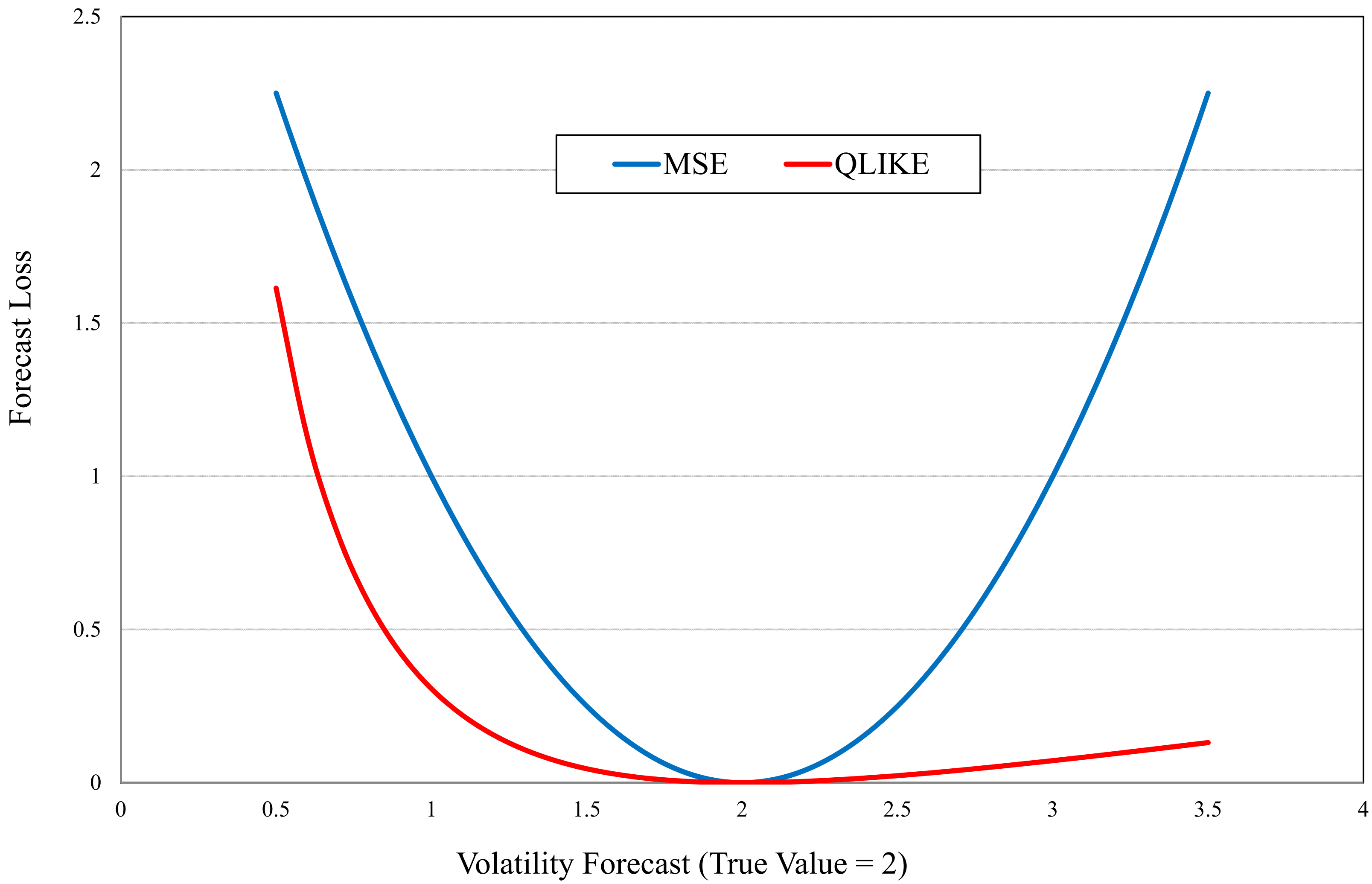
The Volatility Forecast Loss Function

- In order to evaluate volatility forecasts allowing for asymmetric loss, the following function can be used instead of MSE

$$QLIKE = \frac{R_{t+1}^2}{\sigma_{t+1}^2} - \ln\left(\frac{R_{t+1}^2}{\sigma_{t+1}^2}\right) - 1$$

- QLIKE loss function depends on the relative volatility forecast error, R_{t+1}^2/σ_{t+1}^2, rather than on the absolute error, $|R_{t+1}^2 - \sigma_{t+1}^2|$; which is the key ingredient in MSE
- The QLIKE loss function will always penalize more heavily volatility forecasts that underestimate volatility

Figure 4.6: Volatility Loss Function



Summary

- In this Chapter we have
  – Discussed the simple variance forecasting and the RiskMetrics variance model.
  – Introduced the GARCH variance model and compare it with the RiskMetrics model.
  – Estimated the GARCH parameters using the quasi-maximum likelihood method.
  – Suggested extensions to the basic model
  – Discussed various methods for evaluating the volatility forecasting models.