Volatility Forecasting

Elements of
Financial Risk Management
Chapter 2

Peter Christoffersen

Overview

- The RiskMetrics Variance Model
- The GARCH Variance Model
- Persistence & Leverage
- QMLE and Diagnostic Checking
- Predictive ability of GARCH models
- Using Intra-day High/Low Information

Simple Variance Forecasting

- Daily return is given by:

$$R_{t+1} = \sigma_{t+1} z_{t+1}, \text{ with}$$

$$z_{t+1} \sim i.i.d.N(0,1)$$

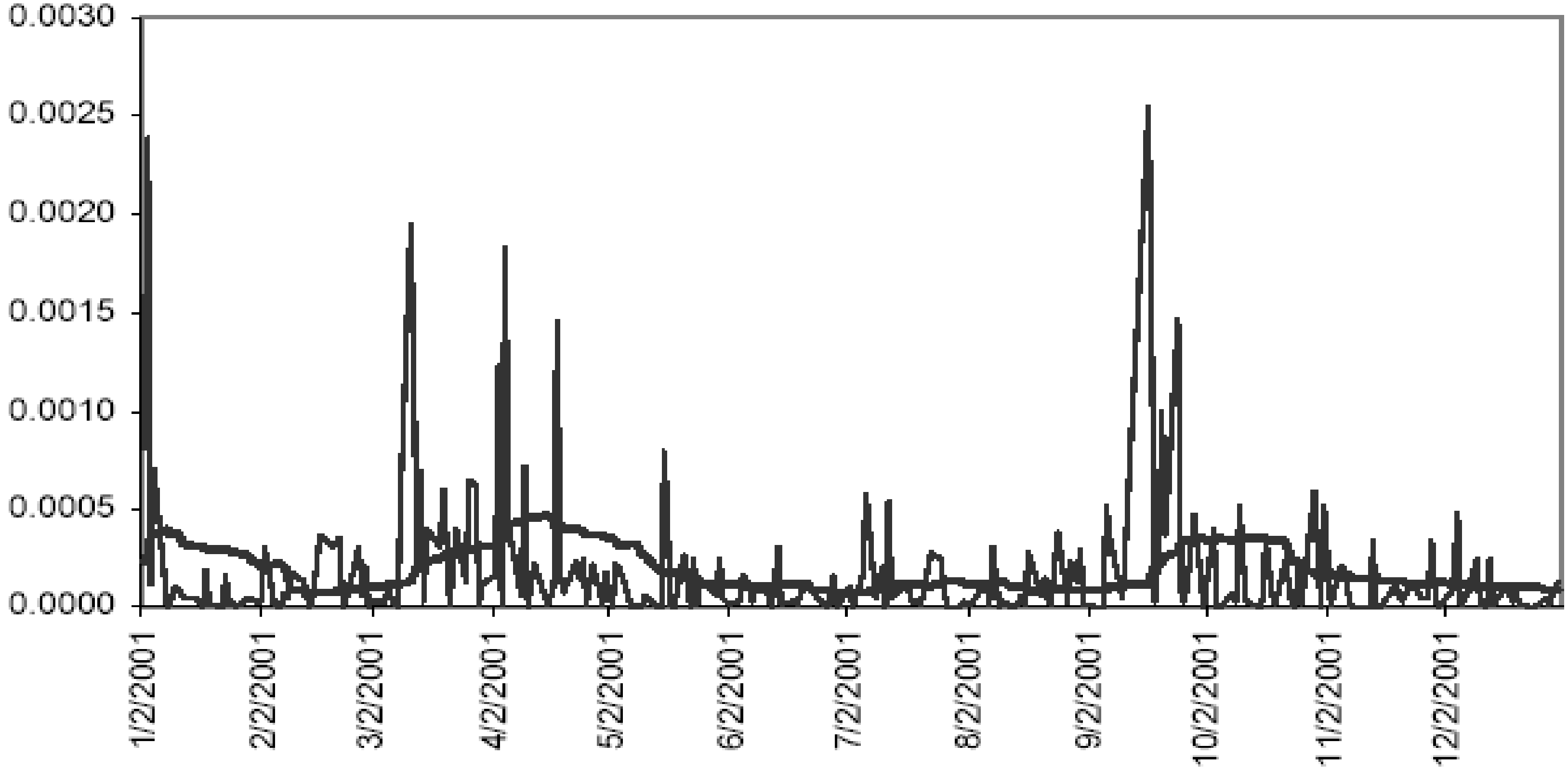
i.i.d. N(0,1) stands for "independently, and identically normally distributed with mean equal to zero and variance equal to 1."

Simple Variance Forecasting

- Tomorrow's variance is given by the simple average of the most recent *m* observations :

$$\sigma^2_{t+1} = \frac{1}{m}\sum_{\tau=1}^{m} R^2_{t+1-\tau}$$
$$= \sum_{\tau=1}^{m} \frac{1}{m} R^2_{t+1-\tau}$$

Squared S&P Returns with Moving Average Variance Estimate (bold) on past 25 observations (m=25) – Fig. 2.1



RiskMetrics

- JP Morgan's RiskMetrics variance model or the exponential smoother is given by:

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

- It can be rewritten as

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

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.
- Estimating 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.

Advantages of RiskMetrics

In this case, no estimation is necessary, which is a huge advantage in large portfolios.

- 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

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

Advantages of RiskMetrics

It is only necessary to store about 100 daily lags of returns in order to calculate tomorrow's variance, σ_{t+1}^2.

- Despite these advantages, we will see shortly that RiskMetrics does have certain shortcomings which will motivate us to consider slightly more elaborate models.
- For example, it does not allow for a leverage effect, which was considered a stylized fact in Chapter 1, and it also provides counterfactual longer-horizon forecasts.

The GARCH Variance Model

- The following set of models 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 GARCH (Generalized AutoRegressive Conditional Heteroskedasticity) 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 $\alpha = 1 - \lambda, \beta = \lambda, \text{ s.t. } \alpha + \beta = 1, \omega = 0$
- However there is an important difference : We can define the unconditional, or long=run average, variance to be

$$\sigma^2 \equiv E[\sigma_{t+1}^2] = \omega + \alpha E[R_t^2] + \beta E[\sigma_t^2]$$

$$= \omega + \alpha\sigma^2 + \beta\sigma^2 \text{ so that}$$

$$\sigma^2 = \omega / (1 - \alpha - \beta)$$

The GARCH Variance Model

- It is now clear that if $\alpha + \beta = 1$ as is the case in the RiskMetrics model, then the long-run variance is infinite or 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 Variance Model

- The GARCH model can also be written

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

- Thus tomorrow's variance is a weighted average of the long-run variance, today's squared return and today's variance.

The GARCH Variance Model

- Our intuition might tell us that ignoring the long-run variance, as the RiskMetrics model does, is more important for longer-horizon forecasting than for forecasting simply one-day ahead. This intuition is correct as we will now see.
- 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}

Longer Horizons

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

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

- The conditional expectation $E_t[*]$ refers to taking the expectation using all the info. Available at the end of day *t,* which includes the squared return on day *t* itself. Persistence is $(\alpha + \beta)$

The GARCH Variance Model

- RiskMetrics model ignores the long-run variance when forecasting.
- If $(\alpha + \beta)$ is close to one as is typically the case, then the two models might yield similar predictions for short horizons, *k*, but their longer horizon implications are very different.
- If today is a low-variance day then the RiskMetrics model predicts that all future days will be low-variance.
- The GARCH model more realistically assumes that eventually in the future variance will revert to the average value.

The GARCH Variance Model

- The forecast of variance of *K*-day cumulative returns,

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

- As we assume that returns have zero autocorrelation, 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:

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

- But in the simple GARCH (1,1) model:

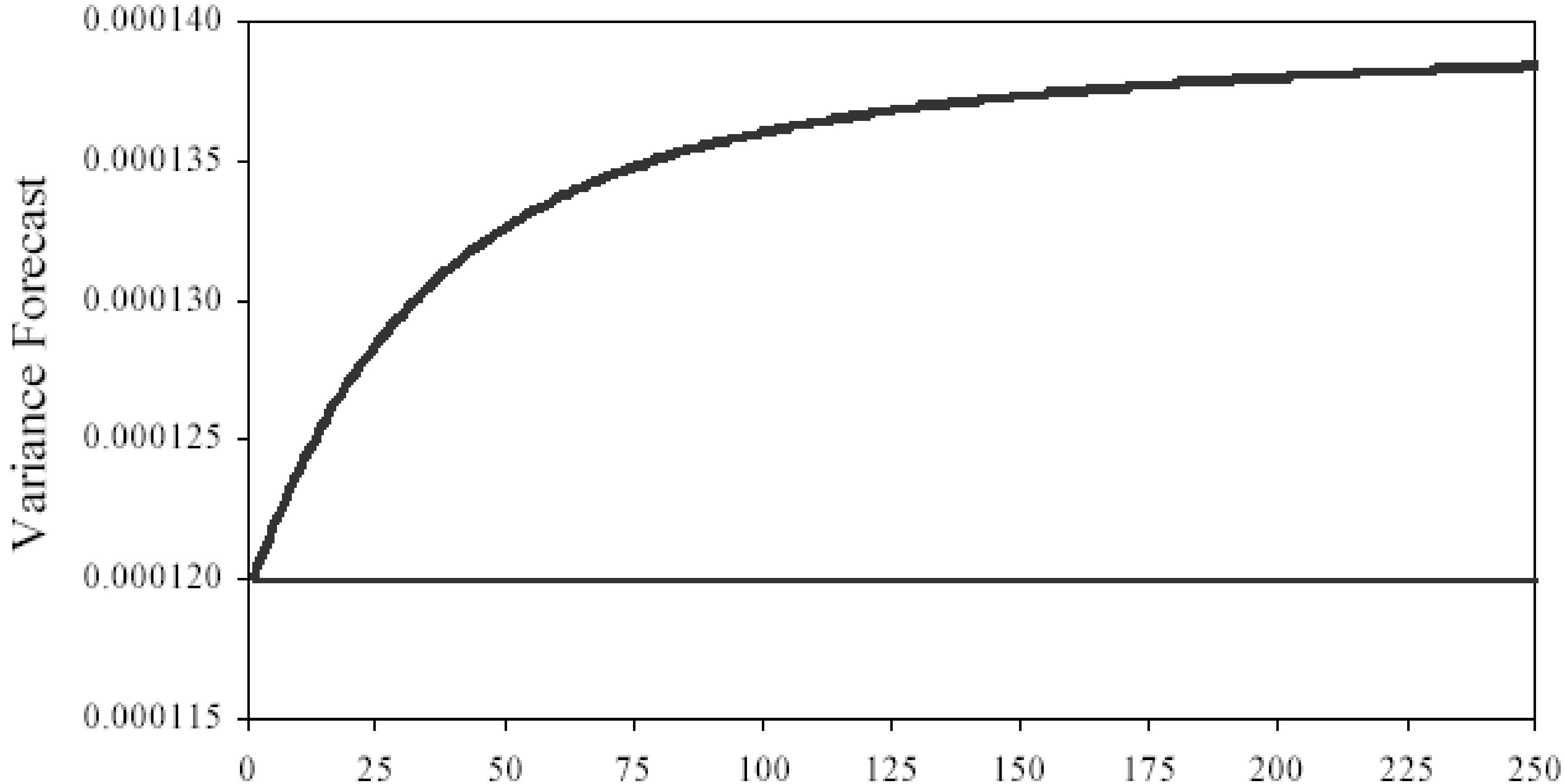
$$\sigma^2_{t+1:t+K} = K\sigma^2 + \sum_{k=1}^{K} (\alpha+\beta)^{k-1}\left(\sigma^2_{t+1} - \sigma^2\right) \neq K\sigma^2_{t+1}$$

- 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

- Thus assuming the Riskmetrics model if the data truly 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.2.2 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_{t+1}^{2} and setting $\alpha = 0.05$ and $\beta = 0.90$. The long run variance in the figure is $\sigma^{2} = 0.000140$

Variance Forecasts for 1-250 days ahead cumulative returns from GARCH(bold) and RiskMetrics(horizontal line) Models



The GARCH Variance Model

- An inconvenience which is 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. 5

Long Memory in Variance

- The component GARCH model, which is a restricted GARCH(2,2), can be seen as allowing the long-term variance to be time varying

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

$$v_{t+1} = \omega + \alpha_v\left(R_t^2 - \sigma_t^2\right) + \beta_v v_t$$

- Implies slowly decaying autocorrelations or long memory.
- FIGARCH. Exponential versus power decay.

The Leverage Effect

- We argued in Chapter 1 that a negative return increases variance by more than a positive return of the same magnitude.
- We can modify the GARCH models so that the weight given to the return depends on whether it is positive or negative, in the following manner

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

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

Leverage Effect

- A different model which 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 $\alpha\phi < 0$

- Advantage : the log.specification ensures a positive variance
- Disadvantage : future expected variance beyond (1) period cannot be calculated analytically.

Explanatory Variables

- We can add explanatory variables on the RHS of the volatility specification

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

Where IT_{t+1} takes on the value 1 if date *t*+1 is a Monday (for example) and otherwise 0.

- Any predetermined variable can be included.

Explanatory Variables

- The volatility index (VIX) from the CBOE.
- Monday effects
- Yesterday’s trading volume
- Prescheduled news announcement dates such as company earnings and FOMC meetings dates.
- Be careful with positivity!

Maximum likelihood Estimation

- We have suggested a range of models, but they contain a number of unknown parameters, which must be estimated.
- The conditional variance is itself an unobserved variable, which must be implicitly estimated along with the parameters of the model.

Standard MLE

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

MLE

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

$$Max \ln L = Max \sum_{t=1}^{T} \ln(l_t) =$$

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

- MLEs have nice asymptotic properties in theory: Consistent, asymptotically normal with minimum variance.
- How large should the sample be in reality?

Quasi MLE Estimation

- The skeptical reader will 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 results establishes what is called quasi maximum likelihood estimation or QMLE, referring to the use of normal-MLE estimation even when the normal distribution assumption is false.

Quasi MLE Estimation

- Notice that QMLE buys us the freedom to worry about the conditional distribution later on (in chap.4) but it does come at a price: The QMLE estimates will in general be less precise than those from MLE.
- Thus we trade off theoretical asymptotic parameter efficiency for practicality.
- The operational aspects of parameter estimation will be discussed in the exercises; here we just point out a trick called variance targeting.

Variance Targeting

- Recall the simple GARCH model can be written as :

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

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

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

- Imposes the long-run variance on the GARCH mode and reduces the number of parameters to be estimated by one.

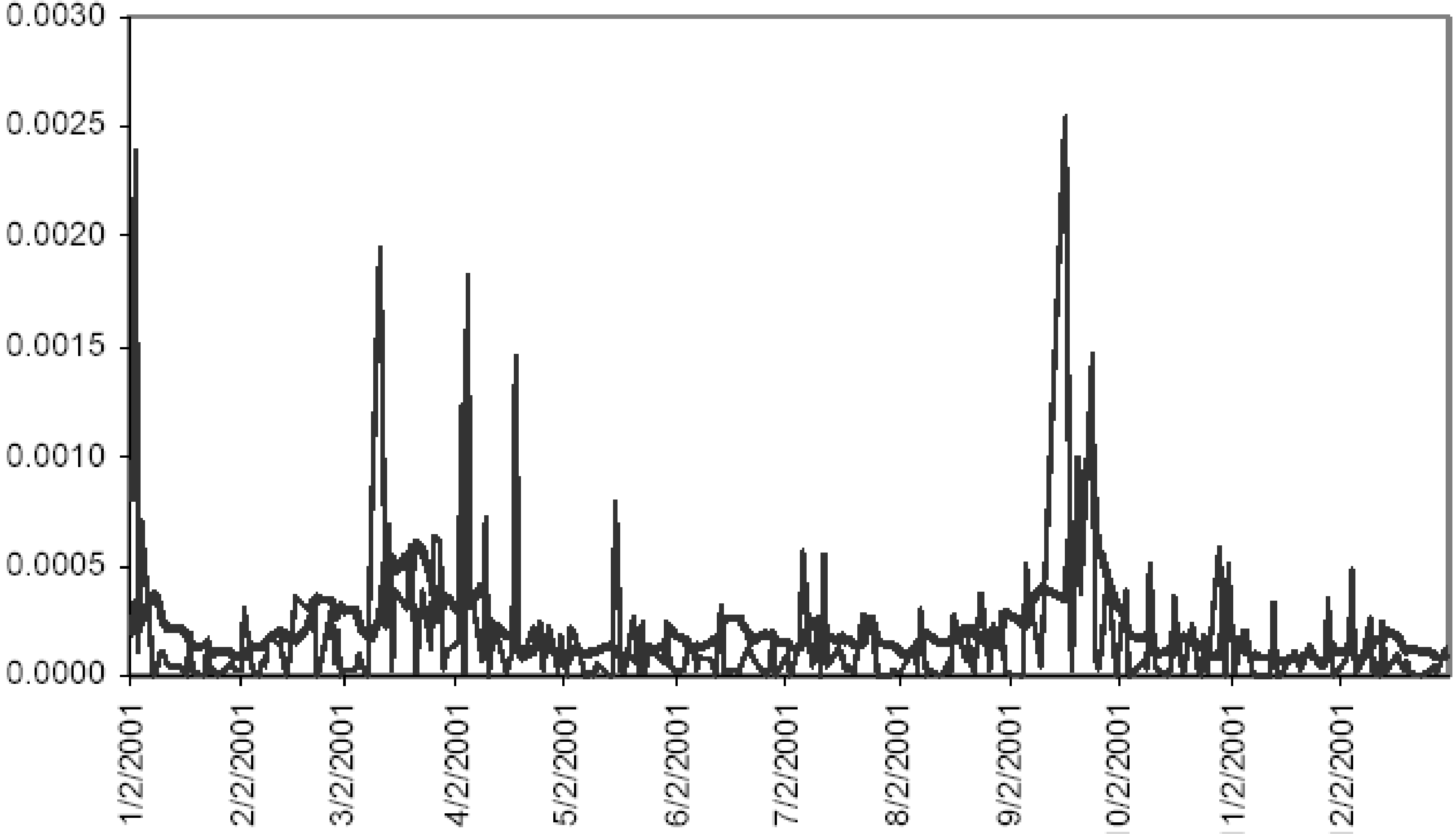
An Example

- Fig. 2.3 shows the S&P500 squared returns from Fig. 2.1 but with an estimated GARCH variance superimposed. The estimated model is the GARCH model with leverage (NGARCH).
- Using numerical optimization of the likelihood function, the optimal parameters imply the following variance dynamics:

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

$$= 0.0000099 + 0.0556\left(R_t - 2.1449\sigma_t\right)^2 + 0.6393\sigma_t^2$$

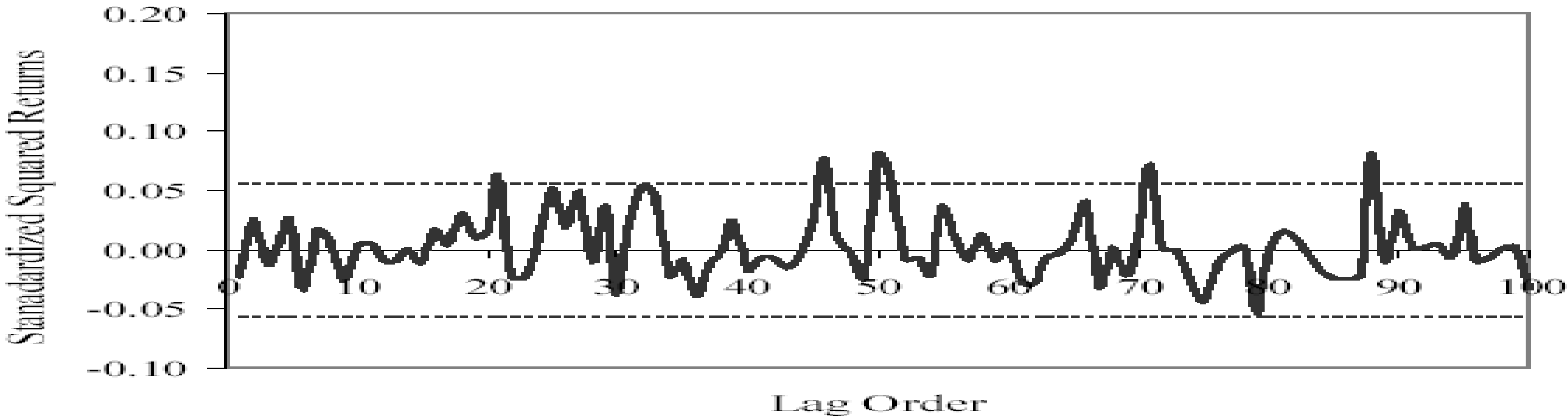
Squared S&P500 Returns with NGARCH Variance Estimate (bold) – Fig. 2.3

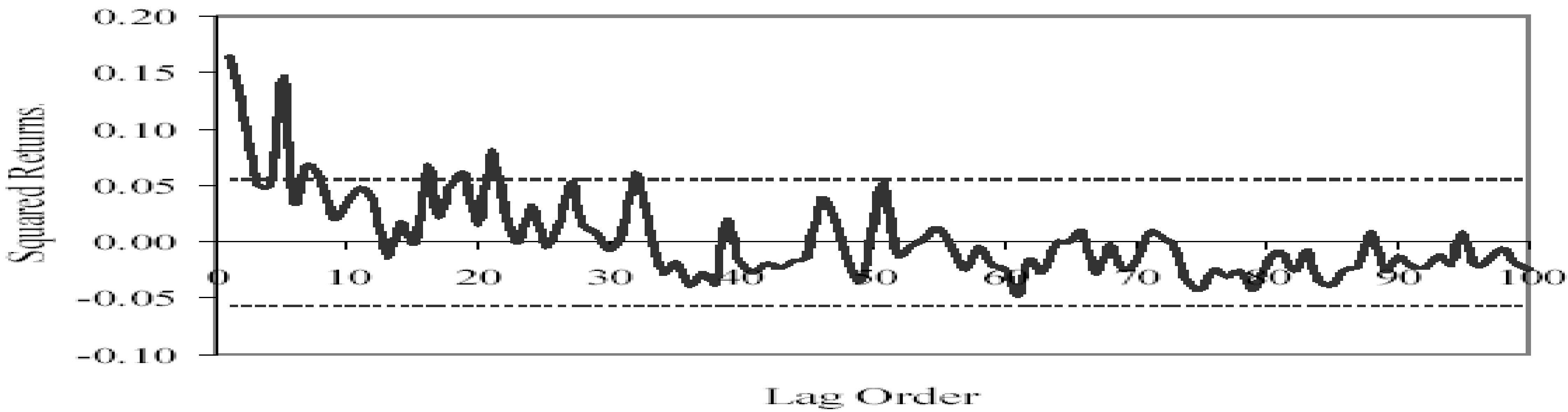


In Sample Check on the Autocorrelations

- The objective of variance modeling is to construct a variance measure, which has the property that the standardized squared returns, R_t^2 / σ_t^2 have no systematic autocorrelation patterns.
- See Figure 2.4. GARCH model with leverage for the S&P500 returns along with their standard error bands.

Autocorrelation of Standardized Squared S&P500 Returns (top) and of Squared Returns (bottom) with Bartlett Standard Errors (dashed)





In Sample Check on the Autocorrelations

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

Out-of-Sample Check using Regression

- Another method of evaluating a variance model is based on simple regressions where squared returns in the forecast period, *t*+1, is regressed on the forecast from the variance model, as in

$$R_{t+1}^{2} = b_{0} + b_{1}\sigma_{t+1|t}^{2} + e_{t+1}$$

- Test b_0=0 and b_1=1.
- In this regression, the squared returns is used as a proxy for the true but unobserved variance in period *t+1.*

The Squared Return Proxy

- The variance of the proxy

$$Var_t[R_{t+1}^2] = E_t[(R_{t+1}^2 - \sigma_{t+1}^2)^2] = E_t[(\sigma_{t+1}^2(z_{t+1}^2 - 1))^2]$$
$$= \sigma_{t+1}^4 E_t[\left(z_{t+1}^2 - 1\right)^2] = \sigma_{t+1}^4(\kappa - 1)$$

where κ is the kurtosis of the innovation which is 3 under conditional normality but higher in reality. Thus the squared return is an unbiased but potentially very noisy proxy for the conditional variance.

- Low regression R^2 (5-10%) is to be expected.

Using Intraday Information

- If the squared return from daily closing prices really is a poor proxy for the true but unobserved daily variance, then we may be able to improve upon variance models which are based purely on squared return by looking for better variance proxies.

Using Intraday High and Low Prices

- One such readily available proxy is the difference between the intraday high and low log-price, which is often referred to as the range.
- The intraday high and low prices are often available along with the daily closing prices in standard financial data bases.
- Range-based variance proxies are therefore easily calculated.
- Let us define the range of the log-prices to be

$$D_t = \ln\left(S_t^{High}\right) - \ln\left(S_t^{Low}\right)$$

Using Intraday High and Low Prices

where S_t^{High} and S_t^{Low} are the highest and lowest prices observed during day *t.*

- One can show that the expected value of the squared range is

$$E[D_t^2] = 4\ln(2)\sigma^2$$

- A natural range-based estimate of volatility is therefore

$$\sigma^2 = \frac{1}{4\ln(2)}\frac{1}{T}\sum_{t=1}^{T} D_t^2$$

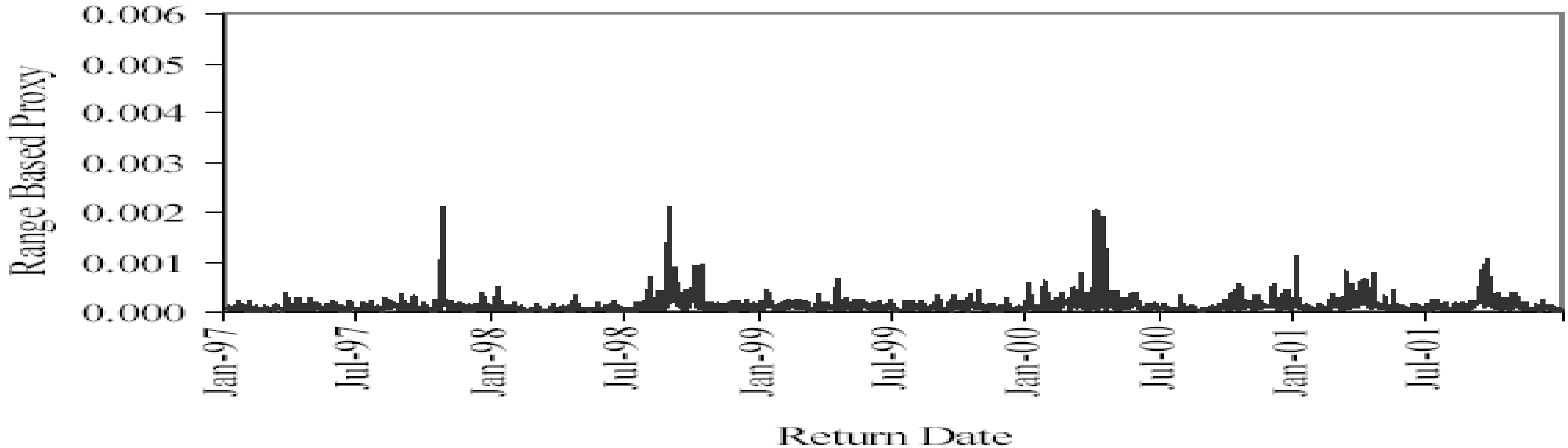
Using Intraday High and Low Prices

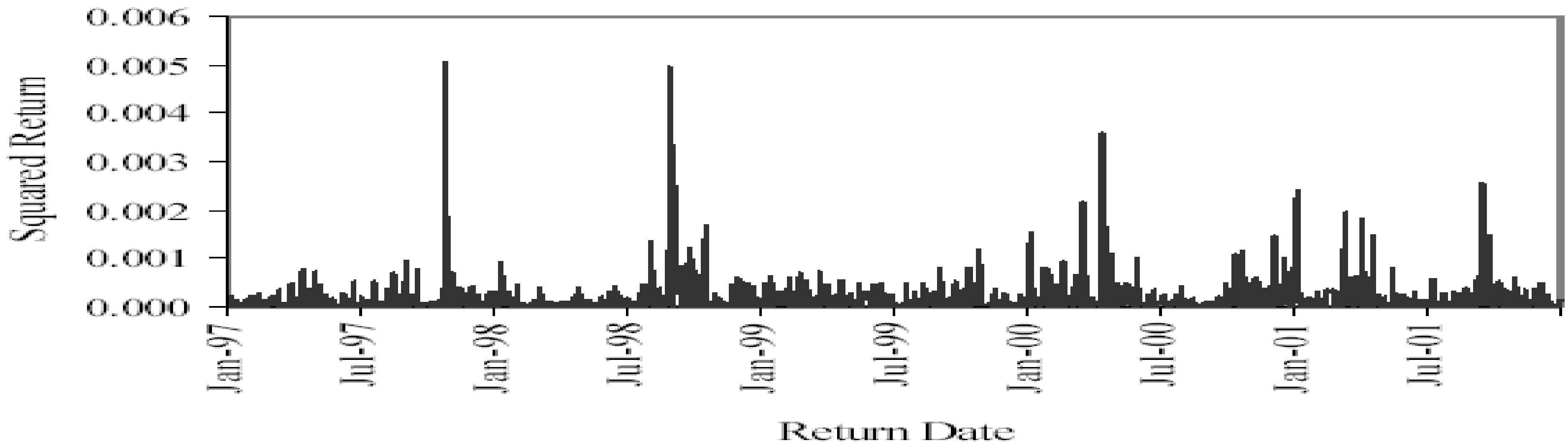
- The range-based estimate of the variance is simply a constant times the average squared range. The constant is $1/(4\ln(2)) \approx 0.361$.
- The range-based estimate of unconditional variance suggests that a proxy for the daily variance can be constructed as

$$\sigma_{r,t}^2 = \frac{1}{4\ln(2)} D_t^2 \approx .361 D_t^2$$

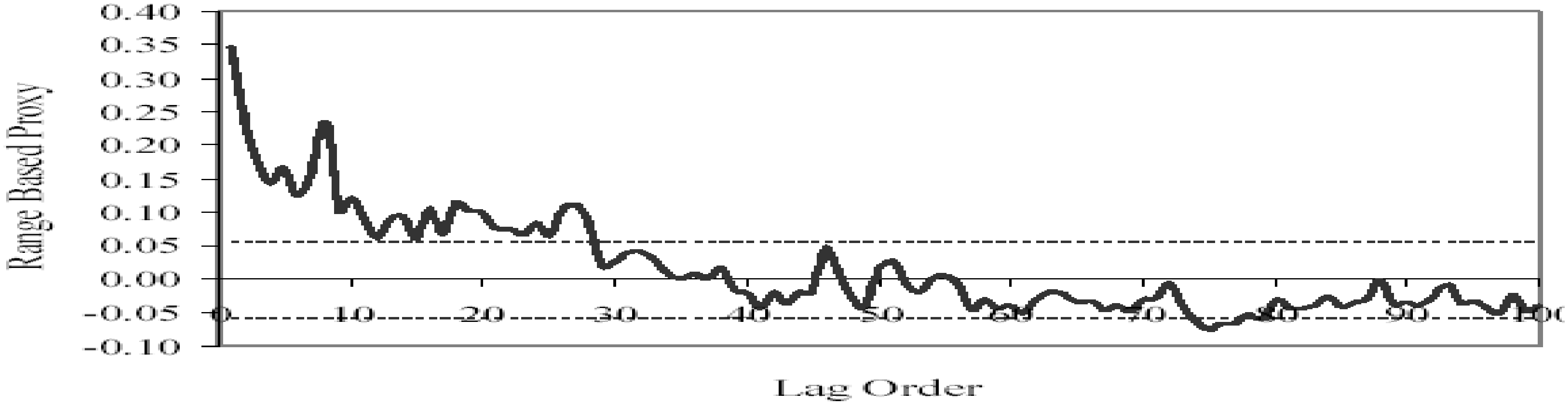
- The top panel of Fig.2.5 plots $\sigma_{r,t}^2$ for the S&P500 data.
- Fig. 2.6 shows the autocorrelation of $\sigma_{r,t}^2$ in the top panel.

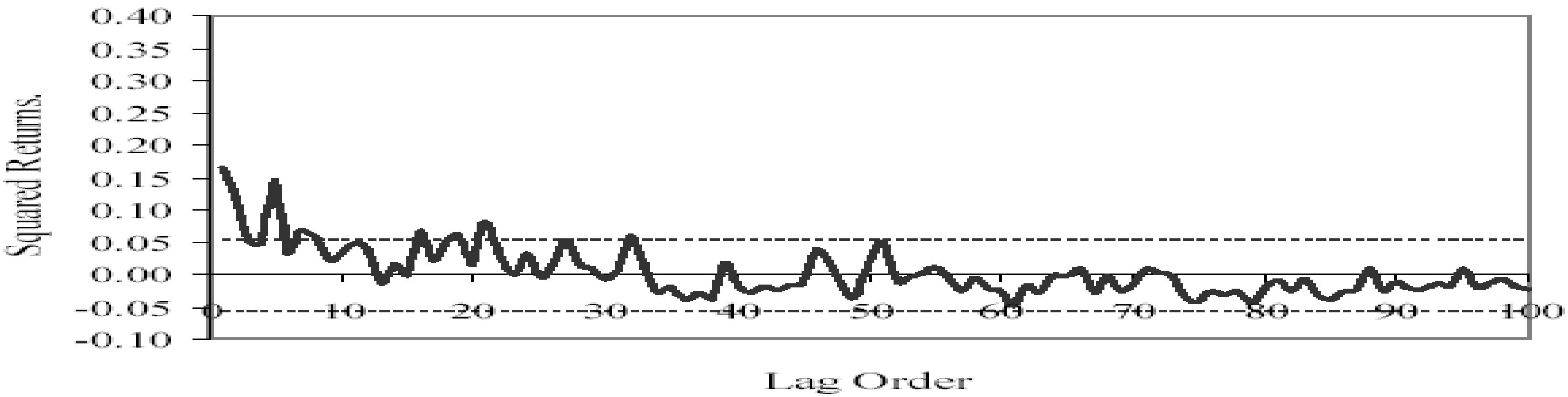
Range-Based Variance Proxy(top) and Squared Returns (bottom) – Fig 2.5





Autocorrelation of Range-Based Variance Proxy (top) and Autocorrelation of Squared Returns (bottom) with Bartlett Standard Errors (dashed)





Using Intraday High and Low Prices

- The $\sigma^2_{r,t}$ variance proxy be used instead of the squared return for evaluating the forecasts from variance models by the regression (done in the exercises)

$$\sigma^2_{r,t+1} = b_0 + b_1 \sigma^2_{t+1|t} + e_{t+1}$$

- The range could also be used as a the driving variable in a variance model; at the least the range could be used as a regressor in the simple GARCH specification as in

$$\sigma^2_{t+1} = \omega + \alpha R^2_t + \beta \sigma^2_t + \gamma D^2_t$$

Using Intraday Returns

- Sometimes, the daily high and low price is not the only intraday information available.
- Liquid assets are traded many times during a day and there is potentially useful information in the intraday prices about daily variance.
- Consider the case where we have observations every (5) minutes on the price of a liquid asset, for example.

Using Intraday Returns

- Let *m* be the number of observations per day.
- If we have 24 hour trading and 5-min observations, then *m*=24*60/5=288.
- Let the *j*th observation on day *t*+1 be denoted $S_{t+j/m}$.
- Then the closing price on day *t*+1 is $S_{t+j/m} = S_{t+1}$ and the *j*th return is

$$R_{t+j/m} = \ln\left(S_{t+j/m}\right) - \ln\left(S_{t+(j-1)/m}\right)$$

Using Intraday Returns

- Having *m* observations within a day, we can calculate an estimate of the daily variance from intraday squared returns as

$$\sigma^2_{m,t+1} = \sum_{j=1}^{m} R^2_{t+j/m}$$

- This variance measure could of course also be used instead of the squared return for evaluating the forecasts from variance models. Thus we could run the regression

$$\sigma^2_{m,t+1} = b_0 + b_1 \sigma^2_{t+1|t} + e_{t+1}$$

Using Intraday Returns

- For liquid securities, the distribution of the logarithm of $\sigma^2_{m,t+1}$ across days appears to be very close to the normal distribution.
- Thus a very practical and sensible forecasting model of variance based on the realized variance measure, would be for example

$$\ln \sigma^2_{m,t+1} = \rho \ln \sigma^2_{m,t} + \varepsilon_{t+1} \quad \text{with} \quad \varepsilon_{t+1} \sim N(0,1)$$

Using Intraday Returns

- From the assumption of normality of the error term we can use the result

$$\varepsilon_{t+1} \sim N\left(0, \sigma_\varepsilon^2\right) \Rightarrow E\left[\exp\left(\varepsilon_{t+1}\right)\right] = \exp\left(\sigma_\varepsilon^2 / 2\right)$$

- Thus in the AR(1) model the forecast for tomorrow is

$$\begin{aligned}
\sigma_{t+1|t}^2 &= E_t[\exp(\rho \ln \sigma_{m,t}^2 + \varepsilon_{t+1})] \\
&= \exp(\rho \ln \sigma_{m,t}^2) E_t[\exp(\varepsilon_{t+1})] \\
&= (\sigma_{m,t}^2)^\rho \exp(\sigma_\varepsilon^2 / 2)
\end{aligned}$$

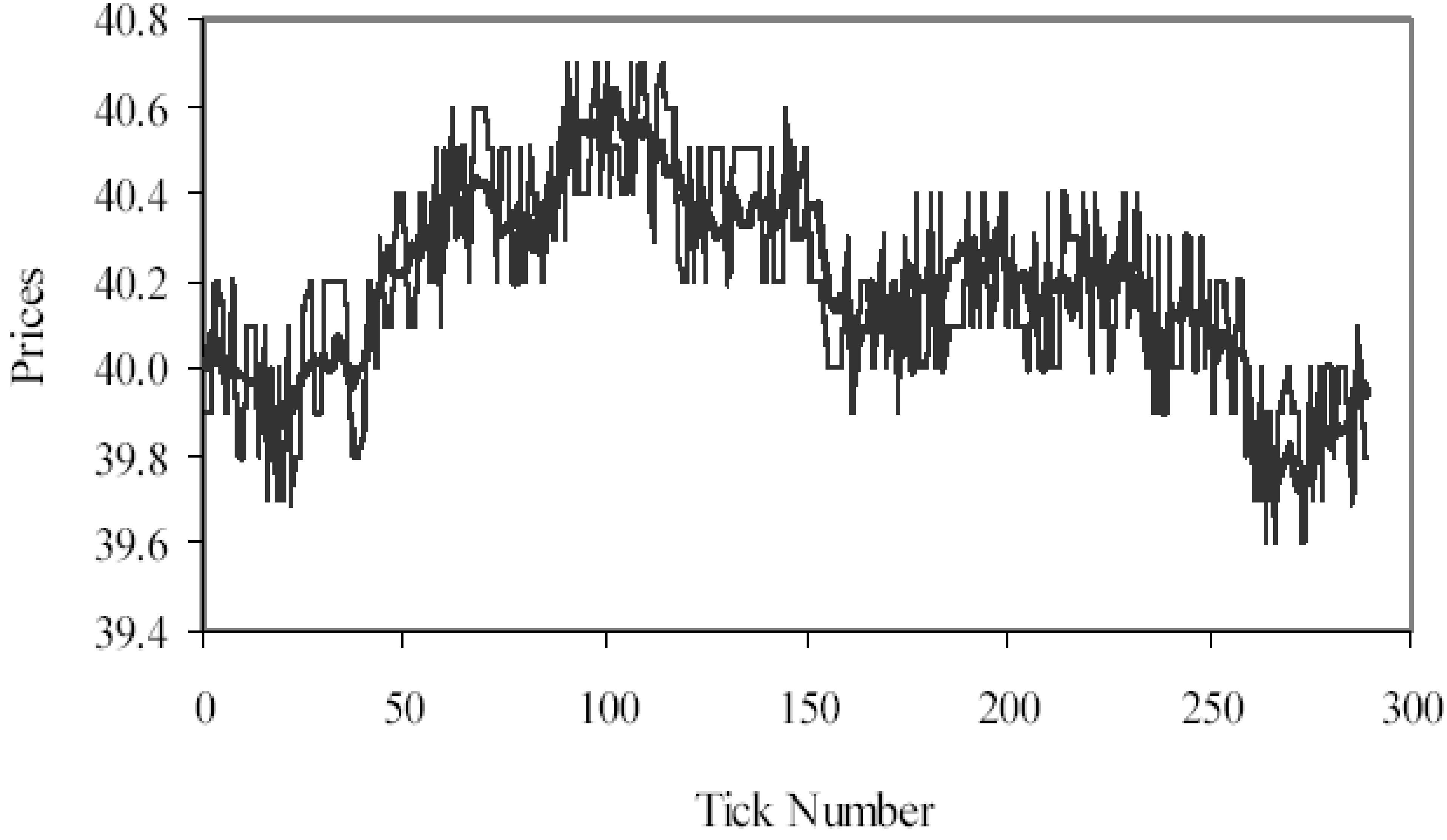
Ranged-Based versus Realized Variance

- There is convincing empirical evidence that for very liquid securities the realized variance modeling approach is useful for risk management purposes.
- The intuition is that using intraday returns gives a reliable estimate of today's variance, which in turn helps forecasting tomorrow's variance.
- In standard GARCH models on the other hand, today's variance is implicitly calculated using exponentially declining weights on many past daily squared returns, where the exact weighting scheme depends on the estimated parameters.

Intraday versus Daily Data

- Thus the GARCH estimate of today's variance is heavily model dependent whereas the realized variance for today is calculated exclusively from today's squared intraday returns.
- When forecasting the future, knowing where you are today is key.
- Market microstructure effects. Fig. 2.7
- Bid-ask easy to correct in range based.

Fundamental Price (bold) and Quoted Price with Bid-Ask Bounces – Fig.2.7



SUMMARY

- This chapter presented a range of variance models which are useful for risk management.
- Simple equally weighted and exponentially weighted models require minimal effort in estimation but do have certain shortcomings.
- Simple, more sophisticated models from the GARCH family are more flexible while taking the following into account : leverage effects, day-of-week effects, announcement effects, etc

SUMMARY

- Quasi Maximum Likelihood estimation technique (QMLE).
- In-sample and out-of-sample model validation techniques.
- Dynamic variance models, which use variance proxies from intraday returns to construct more precise forecasts of future daily variance.