

Chapter 8

Value-at-Risk, Expected Shortfall and Density Forecasting

Note: The primary reference for these notes is [Gourieroux & Jasiak \(2009\)](#), although it is fairly technical. An alternative and less technical textbook treatment can be found in [Christoffersen \(2003\)](#) while a comprehensive and technical treatment can be found in [McNeil, Frey & Embrechts \(2005\)](#).

The American Heritage dictionary, Fourth Edition, defines risk as “the possibility of suffering harm or loss; danger”. In finance, harm or loss has a specific meaning: decreases in the value of a portfolio. This chapter provides an overview of three methods used to assess the riskiness of a portfolio: Value-at-Risk (VaR), Expected Shortfall, and modeling the entire density of a return.

8.1 Defining Risk

Portfolios are exposed to many classes of risk. These six categories represent an overview of the risk factors which may affect a portfolio.

8.1.1 The Types of risk

Market Risk contains all uncertainty about the future price of an asset. For example, changes in the share price of IBM due to earnings news represent market risk.

Liquidity Risk complements market risk by measuring the extra loss involved if a position must be rapidly changed. For example, if a fund wished to sell 20,000,000 shares of IBM on a single day (typical daily volume 10,000,000), this sale would be expected to have a substantial effect on the price. Liquidity risk is distinct from market risk since it represents a transitory distortion due to transaction pressure.

Credit Risk, also known as default risk, covers cases where a 2nd party is unable to pay per previously agreed to terms. Holders of corporate bonds are exposed to credit risk since the bond issuer may not be able to make some or all of the scheduled coupon payments.

Counterparty Risk generalizes credit risk to instruments other than bonds and represents the chance they the counterparty to a transaction, for example the underwriter of an option contract, will be unable to complete the transaction at expiration. Counterparty risk was a major factor in the financial crisis of 2008 where the protection offered in Credit Default Swaps (CDS) was not available when the underlying assets defaulted.

Model Risk represents an econometric form of risk which captures uncertainty over the correct form of the model used to assess the price of an asset and/or the asset's riskiness. Model risk is particularly important when prices of assets are primarily determined by a model rather than in a liquid market, as was the case in the Mortgage Backed Securities (MBS) market in 2007.

Estimation Risk captures an aspect of risk that is present whenever econometric models are used to manage risk since all model contain estimated parameters. Moreover, estimation risk is distinct from model risk since it is present even if a model is correctly specified. In many practical applications, parameter estimation error can result in a substantial misstatement of risk. Model and estimation risk are always present and are generally substitutes – parsimonious models are increasingly likely to be misspecified but have less parameter estimation uncertainty.

This chapter deals exclusively with market risk. Liquidity, credit risk and counterparty risk all require special treatment beyond the scope of this course.

8.2 Value-at-Risk (VaR)

The most common reported measure of risk is Value-at-Risk (VaR). The VaR of a portfolio is the amount risked over some period of time with a fixed probability. VaR provides a more sensible measure of the risk of the portfolio than variance since it focuses on losses, although VaR is not without its own issues. These will be discussed in more detail in the context of coherent risk measures (section 8.5).

8.2.1 Defined

The VaR of a portfolio measures the value (in £, \$, €, ¥, etc.) which an investor would lose with some small probability, usually between 1 and 10%, over a selected period of time. Because the VaR represents a hypothetical *loss*, it is usually a positive number.

A graphical representation of Value-at-Risk

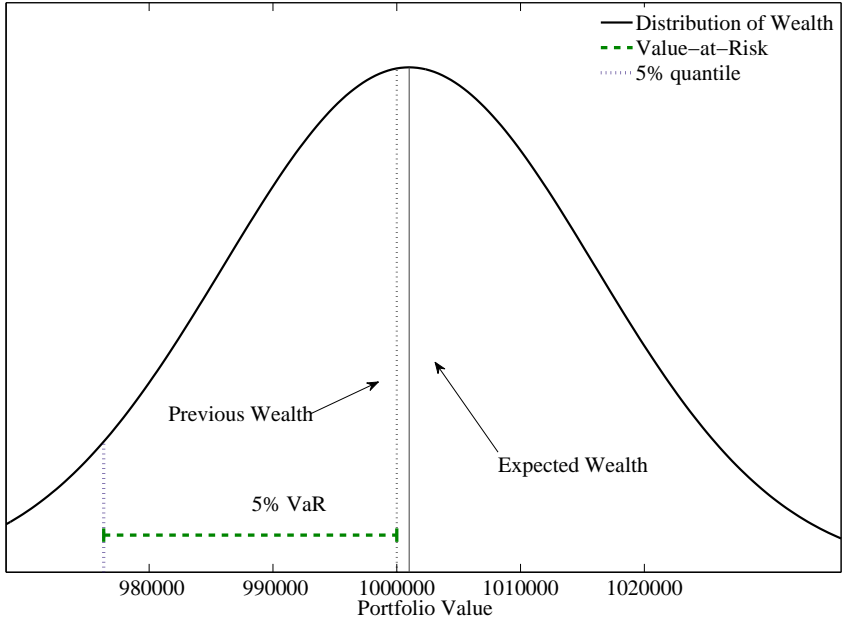


Figure 8.1: A graphical representation of Value-at-Risk. The VaR is represented by the magnitude of the horizontal bar and measures the distance between the value of the portfolio in the current period and its α -quantile. In this example, $\alpha = 5\%$ and returns are $N(.001, .015^2)$.

Definition 8.1 (Value-at-Risk). The α Value-at-Risk (VaR) of a portfolio is defined as the largest number such that the probability that the loss in portfolio value over some period of time is greater than the VaR is α ,

$$Pr(R_t < -VaR) = \alpha \tag{8.1}$$

where $R_t = W_t - W_{t-1}$ is the change in the value of the portfolio, W_t and the time span depends on the application (e.g. one day or two weeks).

For example, if an investor had a portfolio value of £10,000,000 and had a daily portfolio return which was $N(.001, .015^2)$ (annualized mean of 25%, volatility of 23.8%), the daily α Value-at-Risk of this portfolio would

$$£10,000,000(-.001 - .015\Phi^{-1}(\alpha)) = £236,728.04$$

where $\Phi(\cdot)$ is the CDF of a standard normal (and so $\Phi^{-1}(\cdot)$ is the inverse CDF). This expression may appear backward; it is not. The negative sign on the mean indicates that increases in the mean *decrease* the VaR and the negative sign on the standard deviation term indicates that increases in the volatility raise the VaR since for $\alpha < .5$, $\Phi^{-1}(\alpha) < 0$. It is often more useful to express

Value-at-Risk as a percentage of the portfolio value – e.g. 1.5% – rather than in units of currency since it removes the size of the portfolio from the reported number.

Definition 8.2 (Percentage Value-at-Risk). The α percentage Value-at-Risk ($\%VaR$) of a portfolio is defined as the largest return such that the probability that the return on the portfolio over some period of time is less than $-\%VaR$ is α ,

$$Pr(r_t < -\%VaR) = \alpha \quad (8.2)$$

where r_t is the percentage return on the portfolio. $\%VaR$ can be equivalently defined as $\%VaR = VaR/W_{t-1}$.

Since percentage VaR and VaR only differ by the current value of the portfolio, the remainder of the chapter will focus on percentage VaR in place of VaR.

8.2.1.1 The relationship between VaR and quantiles

Understanding that VaR and quantiles are fundamentally related provides a key insight into computing VaR. If r represents the return on a portfolio, the α -VaR is $-1 \times q_\alpha(r)$ where $q_\alpha(r)$ is the α -quantile of the portfolio's return. In most cases α is chosen to be some small quantile – 1, 5 or 10% – and so $q_\alpha(r)$ is a negative number, and VaR should generally be positive.¹

8.2.2 Conditional Value-at-Risk

Most applications of VaR are used to control for risk over short horizons and require a conditional Value-at-Risk estimate that employs information up to time t to produce a VaR for some time period $t + h$.

Definition 8.3 (Conditional Value-at-Risk). The conditional α Value-at-Risk is defined as

$$Pr(r_{t+1} < -VaR_{t+1|t} | \mathcal{F}_t) = \alpha \quad (8.3)$$

where $r_{t+1} = \frac{W_{t+1} - W_t}{W_t}$ is the time $t + 1$ return on a portfolio. Since t is an arbitrary measure of time, $t + 1$ also refers to an arbitrary unit of time (day, two-weeks, 5 years, etc.)

Most conditional models for VaR forecast the density directly, although some only attempt to estimate the required quantile of the time $t + 1$ return distribution. Five standard methods will be presented in the order of the restrictiveness of the assumptions needed to justify the method, from strongest to weakest.

¹If the VaR is negative, either the portfolio has no risk, the portfolio manager has unbelievable skill or most likely the model used to compute the VaR is badly misspecified.

8.2.2.1 RiskMetrics®

The RiskMetrics group has produced a surprisingly simple yet robust method for producing conditional VaR. The basic structure of the RiskMetrics model is a restricted GARCH(1,1), where $\alpha + \beta = 1$ and $\omega = 0$, is used to model the conditional variance,

$$\sigma_{t+1}^2 = (1 - \lambda)r_t^2 + \lambda\sigma_t^2. \quad (8.4)$$

where r_t is the (percentage) return on the portfolio in period t . In the RiskMetrics specification σ_{t+1}^2 follows an *exponentially weighted moving average* which places weight $\lambda^j(1 - \lambda)$ on r_{t-j}^2 .² This model includes no explicit mean model for returns and is only applicable to assets with returns that are close to zero or when the time horizon is short (e.g. one day to one month). The VaR is derived from the α -quantile of a normal distribution,

$$VaR_{t+1} = -\sigma_{t+1}\Phi^{-1}(\alpha) \quad (8.5)$$

where $\Phi^{-1}(\cdot)$ is the inverse normal CDF. The attractiveness of the RiskMetrics model is that there are no parameters to estimate; λ is fixed at .94 for daily data (.97 for monthly).³ Additionally, this model can be trivially extended to portfolios using a vector-matrix switch by replacing the squared return with the outer product of a vector of returns, $\mathbf{r}_t\mathbf{r}'_t$, and σ_{t+1}^2 with a matrix, Σ_{t+1} . The disadvantages of the procedure are that the parameters aren't estimated (which was also an advantage), it cannot be modified to incorporate a leverage effect, and the VaR follows a random walk since $\lambda + (1 - \lambda) = 1$.

8.2.2.2 Parametric ARCH Models

Fully parametric ARCH-family models provide a natural method to compute VaR. For simplicity, only a constant mean GARCH(1,1) will be described, although the mean could be described using other time-series models and the variance evolution could be specified as any ARCH-family member.⁴

$$\begin{aligned} r_{t+1} &= \mu + \epsilon_{t+1} \\ \sigma_{t+1}^2 &= \omega + \gamma_1\epsilon_t^2 + \beta_1\sigma_t^2 \\ \epsilon_{t+1} &= \sigma_{t+1}e_{t+1} \\ e_{t+1} &\stackrel{\text{i.i.d.}}{\sim} f(0, 1) \end{aligned}$$

²An EWMA is similar to a traditional moving average although the EWMA places relatively more weight on recent observations than on observation in the distant past.

³The suggested coefficient for λ are based on a large study of the RiskMetrics model across different asset classes.

⁴The use of α_1 in ARCH models has been avoided to avoid confusion with the α in the VaR.

where $f(0, 1)$ is used to indicate that the distribution of innovations need not be normal but must have mean 0 and variance 1. For example, f could be a standardized Student's t with ν degrees of freedom or Hansen's skewed t with degree of freedom parameter ν and asymmetry parameter λ . The parameters of the model are estimated using maximum likelihood and the time t conditional VaR is

$$VaR_{t+1} = -\hat{\mu} - \hat{\sigma}_{t+1} F_{\alpha}^{-1}$$

where F_{α}^{-1} is the α -quantile of the distribution of e_{t+1} . Fully parametric ARCH models provide substantial flexibility for modeling the conditional mean and variance as well as specifying a distribution for the standardized errors. The limitations of this procedure are that implementations require knowledge of a density family which includes f – if the distribution is misspecified then the quantile used will be wrong – and that the residuals must come from a location-scale family. The second limitation imposes that all of the dynamics of returns can be summarized by a time-varying mean and variance, and so higher order moments must be time invariant.

8.2.2.3 Semiparametric ARCH Models/Filtered Historical Simulation

Semiparametric estimation mixes parametric ARCH models with nonparametric estimators of the distribution.⁵ Again, consider a constant mean GARCH(1,1) model

$$\begin{aligned} r_{t+1} &= \mu + \epsilon_{t+1} \\ \sigma_{t+1}^2 &= \omega + \gamma_1 \epsilon_t^2 + \beta_1 \sigma_t^2 \\ \epsilon_{t+1} &= \sigma_{t+1} e_{t+1} \\ e_{t+1} &\stackrel{\text{i.i.d.}}{\sim} g(0, 1) \end{aligned}$$

where $g(0, 1)$ is an unknown distribution with mean zero and variance 1.

When $g(\cdot)$ is unknown standard maximum likelihood estimation is not available. Recall that assuming a normal distribution for the standardized residuals, even if misspecified, produces estimates which are *strongly consistent*, and so ω , γ_1 and β_1 will converge to their true values for most any $g(\cdot)$. The model can be estimated using QMLE by assuming that the errors are normally distributed and then the Value-at-Risk for the α -quantile can be computed

$$VaR_{t+1}(\alpha) = -\hat{\mu} - \hat{\sigma}_{t+1} \hat{G}_{\alpha}^{-1} \quad (8.6)$$

where \hat{G}_{α}^{-1} is the empirical α -quantile of $e_{t+1} = \left\{ \frac{\epsilon_{t+1}}{\sigma_{t+1}} \right\}$. To estimate this quantile, define $\hat{e}_{t+1} = \frac{\epsilon_{t+1}}{\hat{\sigma}_{t+1}}$, and order the errors such that

⁵This is only one example of a semiparametric estimator. Any semiparametric estimator has elements of both a parametric estimator and a nonparametric estimator.

$$\hat{e}_1 < \hat{e}_2 < \dots < \hat{e}_{n-1} < \hat{e}_n.$$

where n replaces T to indicate the residuals are no longer time ordered. $\hat{G}_\alpha^{-1} = \hat{e}_{\lfloor \alpha n \rfloor}$ or $\hat{G}_\alpha^{-1} = \hat{e}_{\lceil \alpha n \rceil}$ where $\lfloor x \rfloor$ and $\lceil x \rceil$ denote the floor (largest integer smaller than) and ceiling (smallest integer larger than) of x .⁶ In other words, the estimate of G^{-1} is the α -quantile of the empirical distribution of \hat{e}_{t+1} which corresponds to the αn ordered \hat{e}_n .

Semiparametric ARCH models provide one clear advantage over their parametric ARCH cousins; the quantile, and hence the VaR, will be consistent under weaker conditions since the density of the standardized residuals does not have to be assumed. The primary disadvantage of the semiparametric approach is that \hat{G}_α^{-1} may be poorly estimated – especially if α is very small (e.g. 1%). Semiparametric ARCH models also share the limitation that their use is only justified if returns are generated by some location-scale distribution.

8.2.2.4 Cornish-Fisher Approximation

The Cornish-Fisher approximation splits the difference between a fully parametric model and a semi parametric model. The setup is identical to that of the semiparametric model

$$\begin{aligned} r_{t+1} &= \mu + \epsilon_{t+1} \\ \sigma_{t+1}^2 &= \omega + \gamma \epsilon_t^2 + \beta \sigma_t^2 \\ \epsilon_{t+1} &= \sigma_{t+1} e_{t+1} \\ e_{t+1} &\stackrel{\text{i.i.d.}}{\sim} g(0, 1) \end{aligned}$$

where $g(\cdot)$ is again an unknown distribution. The unknown parameters are estimated by quasi-maximum likelihood assuming conditional normality to produce standardized residuals, $\hat{e}_{t+1} = \frac{\hat{\epsilon}_{t+1}}{\hat{\sigma}_{t+1}}$. The Cornish-Fisher approximation is a Taylor-series like expansion of the α -VaR around the α -VaR of a normal and is given by

$$\text{VaR}_{t+1} = -\mu - \sigma_{t+1} F_{CF}^{-1}(\alpha) \quad (8.7)$$

$$\begin{aligned} F_{CF}^{-1}(\alpha) &\equiv \Phi^{-1}(\alpha) + \frac{\zeta}{6} \left([\Phi^{-1}(\alpha)]^2 - 1 \right) + \\ &\frac{\kappa - 3}{24} \left([\Phi^{-1}(\alpha)]^3 - 3\Phi^{-1}(\alpha) \right) - \frac{\zeta^2}{36} \left(2 [\Phi^{-1}(\alpha)]^3 - 5\Phi^{-1}(\alpha) \right) \end{aligned} \quad (8.8)$$

where ζ and κ are the skewness and kurtosis of \hat{e}_{t+1} , respectively. From the expression for $F_{CF}^{-1}(\alpha)$, negative skewness and excess kurtosis ($\kappa > 3$, the kurtosis of a normal) decrease the

⁶When estimating a quantile from discrete data and not smoothing, the is quantile “set valued” and defined as any point between $\hat{e}_{\lfloor \alpha n \rfloor}$ and $\hat{e}_{\lceil \alpha n \rceil}$, inclusive.

estimated quantile and increases the VaR. The Cornish-Fisher approximation shares the strength of the semiparametric distribution in that it can be accurate without a parametric assumption. However, unlike the semi-parametric estimator, Cornish-Fisher estimators are not necessarily consistent which may be a drawback. Additionally, estimates of higher order moments of standardized residuals may be problematic or, in extreme cases, the moments may not even exist.

8.2.2.5 Conditional Autoregressive Value-at-Risk (CaViaR)

Engle & Manganelli (2004) developed ARCH-like models to directly estimate the conditional Value-at-Risk using quantile regression. Like the variance in a GARCH model, the α -quantile of the return distribution, $F_{\alpha,t+1}^{-1}$, is modeled as a weighted average of the previous quantile, a constant, and a 'shock'. The shock can take many forms although a "HIT", defined as an exceedance of the previous Value-at-Risk, is the most natural.

$$HIT_{t+1} = I_{[r_{t+1} < F_{t+1}^{-1}]} - \alpha \quad (8.9)$$

where r_{t+1} the (percentage) return and F_{t+1}^{-1} is the time $t + 1$ α -quantile of this distribution.

Defining q_{t+1} as the time $t + 1$ α -quantile of returns, the evolution in a standard CaViaR model is defined by

$$q_{t+1} = \omega + \gamma HIT_t + \beta q_t. \quad (8.10)$$

Other forms which have been examined are the symmetric absolute value,

$$q_{t+1} = \omega + \gamma |r_t| + \beta q_t. \quad (8.11)$$

the asymmetric absolute value,

$$q_{t+1} = \omega + \gamma_1 |r_t| + \gamma_2 |r_t| I_{[r_t < 0]} + \beta q_t \quad (8.12)$$

the indirect GARCH,

$$q_{t+1} = (\omega + \gamma r_t^2 + \beta q_t^2)^{\frac{1}{2}} \quad (8.13)$$

or combinations of these. The parameters can be estimated by minimizing the "tick" loss function

$$\arg \min_{\theta} T^{-1} \sum_{t=1}^T \alpha (r_t - q_t)(1 - I_{[r_t < q_t]}) + (1 - \alpha)(q_t - r_t)I_{[r_t < q_t]} = \quad (8.14)$$

$$\arg \min_{\theta} T^{-1} \sum_{t=1}^T \alpha (r_t - q_t) + (q_t - r_t)I_{[r_t < q_t]}$$

where $I_{[r_t < q_t]}$ is an indicator variable which is 1 if $r_t < q_t$. Estimation of the parameters in this problem is tricky since the objective function may have many flat spots and is non-differentiable. Derivative free methods, such as simplex methods or genetic algorithms, can be used to overcome these issues. The VaR in a CaViaR framework is then

$$VaR_{t+1} = -q_{t+1} = -\hat{F}_{t+1}^{-1} \quad (8.15)$$

Because a CaViaR model does not specify a distribution of returns or any moments, its use is justified under much weaker assumptions than other VaR estimators. Additionally, its parametric form provides reasonable convergence of the unknown parameters. The main drawbacks of the CaViaR modeling strategy are that it may produce out-of order quantiles (i.e. 5% VaR is less than 10% VaR) and that estimation of the model parameters is challenging.

8.2.2.6 Weighted Historical Simulation

Weighted historical simulation applies weights to returns where the weight given to recent data is larger than the weight given to returns further in the past. The estimator is non-parametric in that no assumptions about either distribution or the dynamics of returns is made.

Weights are assigned using an exponentially declining function. Assuming returns were available from $i = 1, \dots, t$. The weight given to data point i is

$$w_i = \lambda^{t-i} (1 - \lambda) / (1 - \lambda^t), \quad i = 1, 2, \dots, t$$

Typical values for λ range from .99 to .995. When $\lambda = .99$, 99% of the weight occurs in the most recent 450 data points – .995 changes this to the most recent 900 data points. Smaller values of lambda will make the value-at-risk estimator more “local” while larger weights result in a weighting that approaches an equal weighting.

The weighted cumulative CDF is then

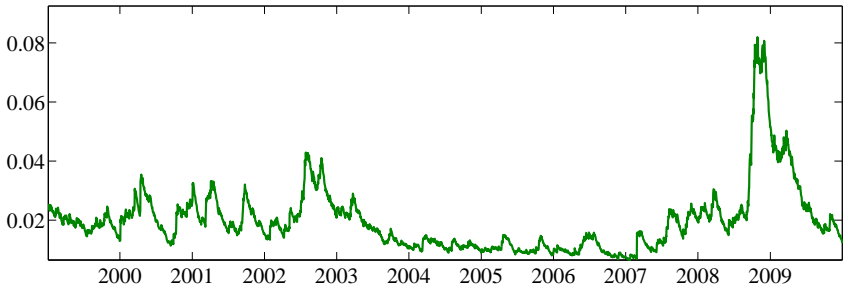
$$\hat{G}(r) = \sum_{i=1}^t w_i I_{[r < r_i]}$$

Clearly the CDF of a return as large as the largest is 1 since the weights sum to 1, and the CDF of the a return smaller than the smallest has CDF value 0. The VaR is then computed as the solution to

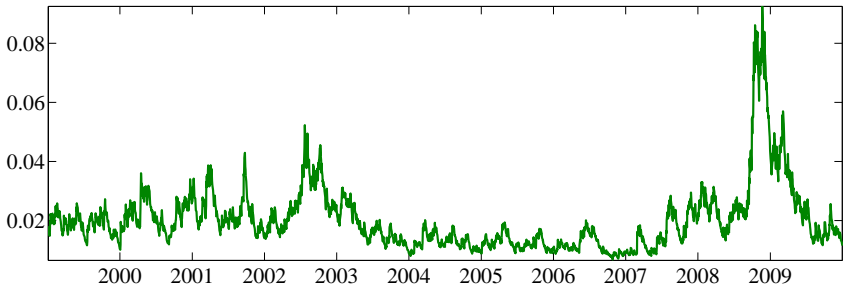
$$VaR_{t+1|t} = \min_r \hat{G}(r) \geq \alpha$$

which chooses the smallest value of r where the weighted cumulative distribution is just as large as α .

Fit Percentage Value-at-Risk using $\alpha = 5\%$
% VaRusing RiskMetrics



% VaRusing TARCH(1,1,1) with Skew t errors



% VaRusing Asymmetric CaViaR

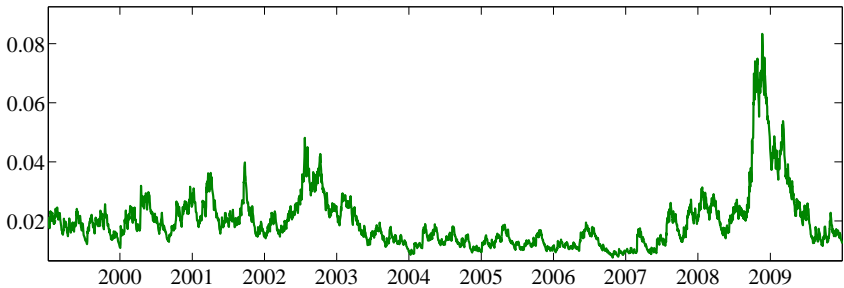


Figure 8.2: The figure contains the estimated % VaR for the S&P 500 using data from 1999 until the end of 2009. While these three models are clearly distinct, the estimated VaRs are remarkably similar.

8.2.2.7 Example: Conditional Value-at-Risk for the S&P 500

The concepts of VaR will be illustrated using S&P 500 returns from January 1, 1999 until December 31, 2009, the same data used in the univariate volatility chapter. A number of models have been

Model Parameters						
TARCH(1,1,1)						
	$\sigma_{t+1} = \omega + \gamma_1 r_t + \gamma_2 r_t I_{[r_t < 0]} + \beta\sigma_t$					
	ω	γ_1	γ_2	β	ν	λ
Normal	0.016	0.000	0.120	0.939		
Student's t	0.015	0.000	0.121	0.939	12.885	
Skew t	0.016	0.000	0.125	0.937	13.823	-0.114
CaViaR						
	$q_{t+1} = \omega + \gamma_1 r_t + \gamma_2 r_t I_{[r_t < 0]} + \beta q_t$					
	ω	γ_1	γ_2	β		
Asym CaViaR	-0.027	0.028	-0.191	0.954		

Estimated Quantiles from Parametric and Semi-parametric TARCH models

	Semiparam.	Normal	Stud. t	Skew t	CF
1%	-3.222	-2.326	-2.439	-2.578	-4.654
5%	-1.823	-1.645	-1.629	-1.695	-1.734
10%	-1.284	-1.282	-1.242	-1.274	-0.834

Table 8.1: Estimated model parameters and quantiles. The choice of distribution for the standardized shocks makes little difference in the parameters of the TARCH process, and so the fit conditional variances are virtually identical. The only difference in the VaRs from these three specifications comes from the estimates of the quantiles of the standardized returns (bottom panel).

estimated which produce similar VaR estimates. Specifically the GARCH models, whether using a normal likelihood, a Students t , a semiparametric or Cornish-Fisher approximation all produce very similar fits and generally only differ in the quantile estimated. Table 8.1 reports parameter estimates from these models. Only one set of TARCH parameters are reported since they were similar in all three models, $\hat{\nu} \approx 12$ in both the standardized Student's t and the skewed t indicating that the standardizes residuals are leptokurtotic, and $\hat{\lambda} \approx -0.1$, from the skewed t , indicating little skewness. The CaViaR estimates indicate little change in the conditional quantile for a symmetric shock (other than mean reversion), a large decrease when the return is negative and that the conditional quantile is persistent.

The table also contains estimated quantiles using the parametric, semiparametric and Cornish-Fisher estimators. Since the fit conditional variances were similar, the only meaningful difference in the VaRs comes from the differences in the estimated quantiles.

8.2.3 Unconditional Value at Risk

While the conditional VaR is often the object of interest, there may be situations which call for the unconditional VaR (also known as marginal VaR). Unconditional VaR expands the set of choices from the conditional to include ones which do not make use of conditioning information to estimate the VaR directly from the unmodified returns.

8.2.3.1 Parametric Estimation

The simplest form of VaR specifies a parametric model for the unconditional distribution of returns and derives the VaR from the α -quantile of this distribution. For example, if $r_t \sim N(\mu, \sigma^2)$, the α -VaR is

$$VaR = -\mu - \sigma\Phi^{-1}(\alpha) \quad (8.16)$$

and the parameters can be directly estimated using Maximum likelihood with the usual estimators,

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

In a general parametric VaR model, some distribution for returns which depends on a set of unknown parameters θ is assumed, $r_t \sim F(\theta)$ and parameters are estimated by maximum likelihood. The VaR is then $-F_\alpha^{-1}$, where F_α^{-1} is the α -quantile of the estimated distribution. The advantages and disadvantages to parametric unconditional VaR are identical to parametric conditional VaR. The models are parsimonious and the parameters estimates are precise yet finding a specification which necessarily includes the true distribution is difficult (or impossible).

8.2.3.2 Nonparametric Estimation/Historical Simulation

At the other end of the spectrum is a pure nonparametric estimate of the unconditional VaR. As was the case in the semiparametric conditional VaR, the first step is to sort the returns such that

$$r_1 < r_2 < \dots < r_{n-1} < r_n$$

where $n = T$ is used to denote an ordering not based on time. The VaR is estimated using $r_{\lfloor \alpha n \rfloor}$ or alternatively $r_{\lceil \alpha n \rceil}$ or an average of the two where $\lfloor x \rfloor$ and $\lceil x \rceil$ denote the floor (largest integer smaller than) and ceiling (smallest integer larger than) of x , respectively. In other words, the estimate of the VaR is the α -quantile of the empirical distribution of $\{r_t\}$,

$$VaR = -\hat{G}_\alpha^{-1} \quad (8.17)$$

where \hat{G}_α^{-1} is the estimated quantile. This follows since the empirical CDF is defined as

$$G(r) = T^{-1} \sum_{t=1}^T I_{[r < r_t]}$$

where $I_{[r < r_t]}$ is an indicator function that takes the value 1 if r is less than r_t , and so this function counts the percentage of returns which are smaller than any value r .

Historical simulation estimates are rough and a single new data point may produce very different VaR estimates. Smoothing the estimated quantile using a kernel density generally improves the precision of the estimate when compared to one calculated directly on the sorted returns. This is particularly true if the sample is small. See section 8.4.2 for more details.

The advantage of nonparametric estimates of VaR is that they are generally consistent under very weak conditions and that they are trivial to compute. The disadvantage is that the VaR estimates can be poorly estimated – or equivalently that very large samples are needed for estimated quantiles to be accurate – particularly for 1% VaRs (or smaller).

8.2.3.3 Parametric Monte Carlo

Parametric Monte Carlo is meaningfully different from either straight parametric or nonparametric estimation of the density. Rather than fit a model to the returns directly, parametric Monte Carlo fits a parsimonious *conditional* model which is then used to simulate the *unconditional* distribution. For example, suppose that returns followed an AR(1) with GARCH(1,1) errors and normal innovations,

$$\begin{aligned} r_{t+1} &= \phi_0 + \phi_1 r_t + \epsilon_{t+1} \\ \sigma_{t+1}^2 &= \omega + \gamma \epsilon_t^2 + \beta \sigma_t^2 \\ \epsilon_{t+1} &= \sigma_{t+1} e_{t+1} \\ e_{t+1} &\stackrel{\text{i.i.d.}}{\sim} N(0, 1). \end{aligned}$$

Parametric Monte Carlo is implemented by first estimating the parameters of the model, $\hat{\theta} = [\hat{\phi}_0, \hat{\phi}_1, \hat{\omega}, \hat{\gamma}, \hat{\beta}]'$, and then simulating the process for a long period of time (generally much longer than the actual number of data points available). The VaR from this model is the α -quantile of the *simulated* data \tilde{r}_t .

$$VaR = -\hat{G}_\alpha^{-1} \tag{8.18}$$

where \hat{G}_α^{-1} is the empirical α -quantile of the simulated data, $\{\tilde{r}_t\}$. Generally the amount of simulated data should be sufficient that no smoothing is needed so that the empirical quantile is an accurate estimate of the quantile of the unconditional distribution. The advantage of this procedure is that it efficiently makes use of *conditioning* information which is ignored in either parametric or nonparametric estimators of unconditional VaR and that rich families of unconditional distributions can be generated from parsimonious conditional models. The

Unconditional Value-at-Risk					
	HS	Normal	Stud. t	Skew t	CF
1% VaR	2.832	3.211	3.897	4.156	5.701
5% VaR	1.550	2.271	2.005	2.111	2.104
10% VaR	1.088	1.770	1.387	1.448	1.044

Table 8.2: Unconditional VaR of S&P 500 returns estimated assuming returns are Normal, Student's t or skewed t , using a Cornish-Fisher transformation or using a nonparametric quantile estimator. While the 5% and 10% VaR are similar, the estimators of the 1% VaR differ.

obvious drawback of this procedure is that an incorrect conditional specification leads to an inconsistent estimate of the unconditional VaR.

8.2.3.4 Example: Unconditional Value-at-Risk for the S&P 500

Using the S&P 500 data, 3 parametric models, a normal, a Student's t and a skewed t , a Cornish-Fisher estimator based on the studentized residuals ($\hat{e}_t = (r_t - \hat{\mu})/\hat{\sigma}$) and a nonparametric estimator were used to estimate the unconditional VaR. The estimates are largely similar although some differences can be seen at the 1% VaR.

8.2.4 Evaluating VaR models

Evaluating the performance of VaR models is not fundamentally different from the evaluation of either ARMA or GARCH models. The key insight of VaR evaluation comes from the loss function for VaR errors,

$$\sum_{t=1}^T \alpha(r_t - F_t^{-1})(1 - I_{[r_t < F_t^{-1}]}) + (1 - \alpha)(F_t^{-1} - r_t)I_{[r_t < F_t^{-1}]} \quad (8.19)$$

where r_t is the return in period t and F_t^{-1} is α -quantile of the return distribution in period t . The *generalized error* can be directly computed from this loss function by differentiating with respect to *VaR*, and is

$$ge_t = I_{[r_t < F_t^{-1}]} - \alpha \quad (8.20)$$

which is the time- t "HIT" (HIT_t).⁷ When there is a VaR exceedance, $HIT_t = 1 - \alpha$ and when

⁷The generalized error extends the concept of an error in a linear regression or linear time-series model to nonlinear estimators. Suppose a loss function is specified as $L(y_{t+1}, \hat{y}_{t+1|t})$, then the generalized error is the derivative of the loss function with respect to the second argument, that is

$$ge_t = \frac{L(y_{t+1}, \hat{y}_{t+1|t})}{\partial \hat{y}_{t+1|t}} \quad (8.21)$$

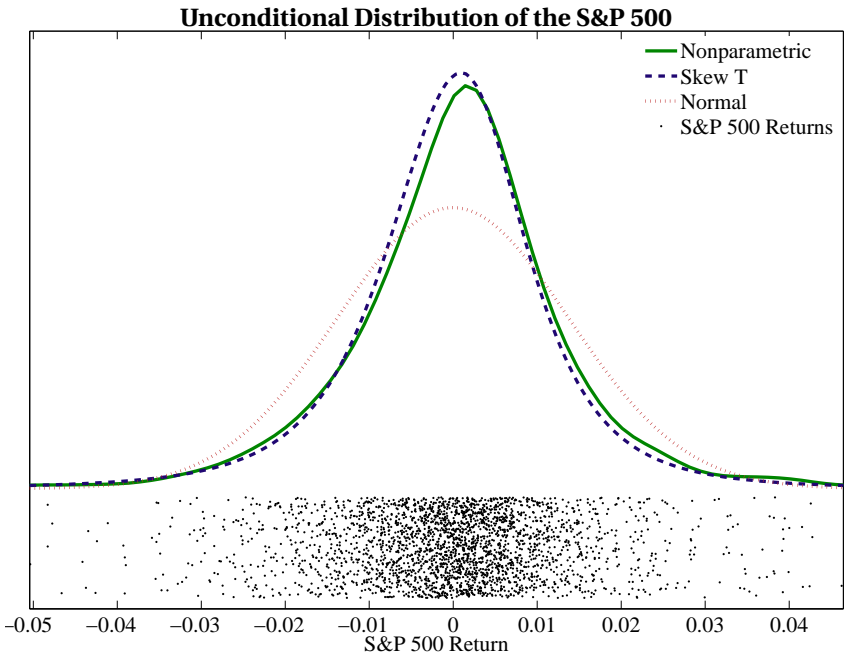


Figure 8.3: Plot of the S&P 500 returns as well as a parametric density using Hansen's skewed t and a nonparametric density estimator constructed using a kernel.

there is no exceedance, $HIT_t = -\alpha$. If the model is correct, then α of the HIT s should be $(1 - \alpha)$ and $(1 - \alpha)$ should be $-\alpha$,

$$\alpha(1 - \alpha) - \alpha(1 - \alpha) = 0,$$

and the mean of HIT_t should be 0. Moreover, when the VaR is conditional on time t information, $E_t[HIT_{t+1}] = 0$ which follows from the properties of optimal forecasts (see chapter 4). A test that the conditional expectation is zero can be performed using a generalized Mincer-Zarnowitz (GMZ) regression of $HIT_{t+1|t}$ on any time t available variable. For example, the estimated quantile $F_{t+1|t}^{-1}$ for $t + 1$ could be included (which is in the time- t information set) as well as lagged HIT s to form a regression,

$$HIT_{t+1|t} = \gamma_0 + \gamma_1 F_{t+1|t}^{-1} + \gamma_2 HIT_t + \gamma_3 HIT_{t-1} + \dots + \gamma_K HIT_{t-K+2} + \eta_t$$

If the model is correctly specified, all of the coefficients should be zero and the null $H_0 : \boldsymbol{\gamma} = 0$ can be tested against an alternative that $H_1 : \gamma_j \neq 0$ for some j .

where it is assumed that the loss function is differentiable at this point.

8.2.4.1 Likelihood Evaluation

While the generalized errors can be tested in the GMZ framework, VaRforecast evaluation can be improved by noting that HIT_t is a Bernoulli random variable which takes the value $1 - \alpha$ with probability α and takes the value $-\alpha$ with probability $1 - \alpha$. Defining $\widetilde{HIT}_t = I_{[r_t < F_t]}$, this modified HIT is exactly a Bernoulli(α), and a more powerful test can be constructed using a likelihood ratio test. Under the null that the model is correctly specified, the likelihood function of a series of \widetilde{HIT} s is

$$f(\widetilde{HIT}; p) = \prod_{t=1}^T p^{\widetilde{HIT}_t} (1 - p)^{1 - \widetilde{HIT}_t}$$

and the log-likelihood is

$$l(p; \widetilde{HIT}) = \sum_{t=1}^T \widetilde{HIT}_t \ln(p) + (1 - \widetilde{HIT}_t) \ln(1 - p).$$

If the model is correctly specified, $p = \alpha$ and a likelihood ratio test can be performed as

$$LR = 2(l(\hat{p}; \widetilde{HIT}) - l(p = \alpha; \widetilde{HIT})) \quad (8.22)$$

where $\hat{p} = T^{-1} \sum_{t=1}^T \widetilde{HIT}_t$ is the maximum likelihood estimator of p under the alternative. The test has a single restriction and so has an asymptotic χ_1^2 distribution.

The likelihood-based test for unconditionally correct VaR can be extended to conditionally correct VaR by examining the sequential dependence of HIT s. This testing strategy uses properties of a Markov chain of Bernoulli random variables. A Markov chain is a modeling device which resembles ARMA models yet is more general since it can handle random variables which take on a finite number of values – such as a HIT . A simple 1st order binary valued Markov chain produces Bernoulli random variables which are not necessarily independent. It is characterized by a transition matrix which contains the probability that the state stays the same. In a 1st order binary valued Markov chain, the transition matrix is given by

$$\begin{bmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{bmatrix} = \begin{bmatrix} p_{00} & 1 - p_{00} \\ 1 - p_{11} & p_{11} \end{bmatrix},$$

where p_{ij} is the probability that the next observation takes value j given that this observation has value i . For example, p_{10} indicates that the probability that the next observation is a not a HIT given the current observation is a HIT . In a correctly specified model, the probability of a HIT in the current period should not depend on whether the previous period was a HIT or not. In other words, the sequence $\{HIT_t\}$ is i.i.d., and so that $p_{00} = 1 - \alpha$ and $p_{11} = \alpha$ when the model is conditionally correct.

Define the following quantities,

$$\begin{aligned}
n_{00} &= \sum_{t=1}^{T-1} (1 - \widetilde{HIT}_t)(1 - \widetilde{HIT}_{t+1}) \\
n_{10} &= \sum_{t=1}^{T-1} \widetilde{HIT}_t(1 - \widetilde{HIT}_{t+1}) \\
n_{01} &= \sum_{t=1}^{T-1} (1 - \widetilde{HIT}_t)\widetilde{HIT}_{t+1} \\
n_{11} &= \sum_{t=1}^{T-1} \widetilde{HIT}_t\widetilde{HIT}_{t+1}
\end{aligned}$$

where n_{ij} counts the number of times $\widetilde{HIT}_{t+1} = i$ after $\widetilde{HIT}_t = j$.

The log-likelihood for the sequence two VaR exceedances is

$$l(p; \widetilde{HIT}) = n_{00} \ln(p_{00}) + n_{01} \ln(1 - p_{00}) + n_{11} \ln(p_{11}) + n_{10} \ln(1 - p_{11})$$

where p_{11} is the probability of two sequential HIT s and p_{00} is the probability of two sequential periods without a HIT . The null is $H_0 : p_{11} = 1 - p_{00} = \alpha$. The maximum likelihood estimates of p_{00} and p_{11} are

$$\begin{aligned}
\hat{p}_{00} &= \frac{n_{00}}{n_{00} + n_{01}} \\
\hat{p}_{11} &= \frac{n_{11}}{n_{11} + n_{10}}
\end{aligned}$$

and the hypothesis can be tested using the likelihood ratio

$$LR = 2(l(\hat{p}_{00}, \hat{p}_{11}; \widetilde{HIT}) - l(p_{00} = 1 - \alpha, p_{11} = \alpha; \widetilde{HIT})) \quad (8.23)$$

and is asymptotically χ^2_2 distributed.

This framework can be extended to include conditioning information by specifying a *probit* or *logit* for \widetilde{HIT}_t using any time- t available information. For example, a specification test could be constructed using K lags of HIT , a constant and the forecast quantile as

$$\widetilde{HIT}_{t+1|t} = \gamma_0 + \gamma_1 F_{t+1|t} + \gamma_2 \widetilde{HIT}_t + \gamma_3 \widetilde{HIT}_{t-1} + \dots + \gamma_K \widetilde{HIT}_{t-K+1}.$$

To implement this in a Bernoulli log-likelihood, it is necessary to ensure that

$$0 \leq \gamma_0 + \gamma_1 F_{t+1|t} + \gamma_2 \widetilde{HIT}_t + \gamma_3 \widetilde{HIT}_{t-1} + \dots + \gamma_K \widetilde{HIT}_{t-K+1} \leq 1.$$

This is generally accomplished using one of two transformations, the normal CDF ($\Phi(z)$)

which produces a probit or the logistic function ($e^z/(1+e^z)$) which produces a logit. Generally the choice between these two makes little difference. If $\mathbf{x}_t = [1 \ F_{t+1|t} \ \widetilde{HIT}_t \ \widetilde{HIT}_{t-1} \ \dots \ \widetilde{HIT}_{t-K+1}]$, the model for \widetilde{HIT} is

$$\widetilde{HIT}_{t+1|t} = \Phi(\mathbf{x}_t \boldsymbol{\gamma})$$

where the normal CDF is used to map from $(-\infty, \infty)$ to $(0,1)$, and so the model is a conditional probability model. The log-likelihood is then

$$l(\boldsymbol{\gamma}; \widetilde{HIT}, \mathbf{x}) = \sum_{t=1}^T \widetilde{HIT}_t \ln(\Phi(\mathbf{x}_t \boldsymbol{\gamma})) - (1 - \widetilde{HIT}_t) \ln(1 - \Phi(\mathbf{x}_t \boldsymbol{\gamma})). \quad (8.24)$$

The likelihood ratio for testing the null $H_0 : \boldsymbol{\gamma}_0 = \Phi^{-1}(\alpha), \boldsymbol{\gamma}_j = 0$ for all $j = 1, 2, \dots, K$ against an alternative $H_1 = \boldsymbol{\gamma}_0 \neq \Phi^{-1}(\alpha)$ or $\boldsymbol{\gamma}_j \neq 0$ for some $j = 1, 2, \dots, K$ can be computed

$$LR = 2 \left(l(\hat{\boldsymbol{\gamma}}; \widetilde{HIT}) - l(\boldsymbol{\gamma}_0; \widetilde{HIT}) \right) \quad (8.25)$$

where $\boldsymbol{\gamma}_0$ is the value under the null ($\boldsymbol{\gamma} = \mathbf{0}$) and $\hat{\boldsymbol{\gamma}}$ is the estimator under the alternative (i.e. the unrestricted estimator from the probit).

8.2.5 Relative Comparisons

Diebold-Mariano tests can be used to relatively rank VaR forecasts in an analogous manner as how they are used to rank conditional mean or conditional variance forecasts (Diebold & Mariano 1995). If $L(r_{t+1}, VaR_{t+1|t})$ is a loss function defined over VaR, then a Diebold-Mariano test statistic can be computed

$$DM = \frac{\bar{d}}{\sqrt{\widehat{V}[\bar{d}]}} \quad (8.26)$$

where

$$d_t = L(r_{t+1}, VaR_{t+1|t}^A) - L(r_{t+1}, VaR_{t+1|t}^B),$$

VaR^A and VaR^B are the Value-at-Risks from models A and B respectively, $\bar{d} = R^{-1} \sum_{t=M+1}^{M+R} d_t$, M (for modeling) is the number of observations used in the model building and estimation, R (for reserve) is the number of observations held back for model evaluation, and $\sqrt{\widehat{V}[\bar{d}]}$ is the long-run variance of d_t which requires the use of a HAC covariance estimator (e.g. Newey-West). Recall that DM has an asymptotical normal distribution and that the test has the null $H_0 : E[d_t] = 0$ and the composite alternative $H_1^A : E[d_t] < 0$ and $H_1^B : E[d_t] > 0$. Large negative values (less than -2) indicate model A is superior while large positive values indicate the opposite; values close to zero indicate neither forecast outperforms the other.

Ideally the loss function, $L(\cdot)$, should reflect the user's loss over VaR forecast errors. In some circumstances there is not an obvious choice. When this occurs, a reasonable choice for the loss function is the VaR optimization objective,

$$L(r_{t+1}, VaR_{t+1|t}) = \alpha(r_{t+1} - VaR_{t+1|t})(1 - I_{[r_{t+1} < VaR_{t+1|t}]}) + (1 - \alpha)(VaR_{t+1|t} - r_{t+1})I_{[r_{t+1} < VaR_{t+1|t}]} \quad (8.27)$$

which has the same interpretation in a VaR model as the mean square error (MSE) loss function has in conditional mean model evaluation or the Q-like loss function has for comparing volatility models.

8.3 Expected Shortfall

Expected shortfall – also known as tail VaR – combines aspects of the VaR methodology with more information about the distribution of returns in the tail.⁸

Definition 8.4 (Expected Shortfall). Expected Shortfall (ES) is defined as the expected value of the portfolio loss *given* a Value-at-Risk exceedance has occurred. The *unconditional* Expected Shortfall is defined

$$\begin{aligned} ES &= E \left[\frac{W_1 - W_0}{W_0} \mid \frac{W_1 - W_0}{W_0} < -VaR \right] \\ &= E [r_{t+1} | r_{t+1} < -VaR] \end{aligned} \quad (8.28)$$

where W_t , $t = 0, 1$, is the value of the assets in the portfolio and 1 and 0 measure an arbitrary length of time (e.g. one day or two weeks).⁹

The conditional, and generally more useful, Expected Shortfall is similarly defined.

Definition 8.5 (Conditional Expected Shortfall). Conditional Expected Shortfall is defined

$$ES_{t+1} = E_t [r_{t+1} | r_{t+1} < -VaR_{t+1}]. \quad (8.29)$$

where r_{t+1} return on a portfolio at time $t + 1$. Since t is an arbitrary measure of time, $t + 1$ also refers to an arbitrary unit of time (day, two-weeks, 5 years, etc.)

⁸Expected Shortfall is a special case of a broader class of statistics known as *exceedance measures*. Exceedance measures all describe a common statistical relationship *conditional* on one or more variable being in its tail. Expected shortfall it is an *exceedance mean*. Other exceedance measures which have been studied include exceedance variance, $V[x | x < q_\alpha]$, exceedance correlation, $\text{Corr}(x, y | x < q_{\alpha,x}, y < q_{\alpha,y})$, and exceedance β , $\text{Cov}(x, y | x < q_{\alpha,x}, y < q_{\alpha,y}) / (V[x | x < q_{\alpha,x}]V[y | y < q_{\alpha,y}])^{\frac{1}{2}}$ where q_α is the α -quantile of the distribution of x or y .

⁹Just like VaR, Expected Shortfall can be equivalently defined in terms of returns or in terms of wealth. For consistency with the VaR discussion, Expected Shortfall is presented in terms of the return.

Because computation of Expected Shortfall requires both a quantile and an expectation, they are generally computed from density models, either parametric or semi-parametric, rather than simpler and more direct specifications.

8.3.1 Evaluating Expected Shortfall models

Expected Shortfall models can be evaluated using standard techniques since Expected Shortfall is a conditional mean,

$$E_t[ES_{t+1}] = E_t[r_{t+1} | r_{t+1} < -VaR_{t+1}].$$

A generalized Mincer-Zarnowitz regression can be used to test whether this mean is zero. Let $I_{[r_t < VaR_t]}$ indicate that the portfolio return was less than the VaR. The GMZ regression for testing Expected Shortfall is

$$(ES_{t+1|t} - r_{t+1})I_{[r_{t+1} < -VaR_{t+1|t}]} = \mathbf{x}_t \boldsymbol{\gamma} \quad (8.30)$$

where \mathbf{x}_t , as always, is any set of time t measurable instruments. The natural choices for \mathbf{x}_t include a constant and $ES_{t+1|t}$, the forecast expected shortfall. Any other time- t measurable regressor that captures some important characteristic of the tail, such as recent volatility ($\sum_{i=0}^{\tau} r_{t-i}^2$) or VaR (VaR_{t-i}), may also be useful in evaluating Expected Shortfall models. If the Expected Shortfall model is correct, the null that none of the regressors are useful in predicting the difference, $H_0 : \boldsymbol{\gamma} = \mathbf{0}$, should not be rejected. If the left-hand side term – Expected Shortfall “surprise” – in eq. (8.30) is predictable, then the model can be improved.

Despite the simplicity of the GMZ regression framework to evaluate Expected Shortfall, their evaluation is difficult owing to a lack of data regarding the exceedance mean; Expected Shortfall can only be measured when there is a VaR exceedance and so 4 years of data would only produce 50 observations where this was true. The lack of data about the tail makes evaluating Expected Shortfall models difficult and can lead to a failure to reject in many cases even when using badly misspecified Expected Shortfall models.

8.4 Density Forecasting

VaR (a quantile) provides a narrow view into the riskiness of an asset. More importantly, VaR may not adequately describe the types of risk an arbitrary forecast consumer may care about. The same cannot be said for a density forecast which summarizes *everything* there is to know about the riskiness of the asset. Density forecasts also nest both VaR and Expected Shortfall as special cases.

In light of this relationship, it may be tempting to bypass VaR or Expected Shortfall forecasting and move directly to density forecasts. Unfortunately density forecasting also suffers from a number of issues, including:

- The density contains all of the information about the random variable being studied, and so a flexible form is generally needed. The cost of this flexibility is increased parameter estimation error which can be magnified when computing the expectation of nonlinear functions of a forecast asset price density (e.g. pricing an option).
- Multi-step density forecasts are difficult (often impossible) to compute since densities do not time aggregate, except in special cases which are usually too simple to be of any interest. This contrasts with standard results for ARMA and ARCH models.
- Unless the user has preferences over the entire distribution, density forecasting inefficiently utilizes information.

8.4.1 Density Forecasts from ARCH models

Producing density forecasting from ARCH models is virtually identical to producing VaR forecasts from ARCH models. For simplicity, only a model with a constant mean and GARCH(1,1) variances will be used, although the mean and variance can be modeled using richer, more sophisticated processes.

$$\begin{aligned}
 r_{t+1} &= \mu + \epsilon_{t+1} \\
 \sigma_{t+1}^2 &= \omega + \gamma_1 \epsilon_t^2 + \beta_1 \sigma_t^2 \\
 \epsilon_{t+1} &= \sigma_{t+1} e_{t+1} \\
 e_{t+1} &\stackrel{\text{i.i.d.}}{\sim} g(0, 1).
 \end{aligned}$$

where $g(0, 1)$ is used to indicate that the distribution of innovations need not be normal but must have mean 0 and variance 1. Standard choices for $g(\cdot)$ include the standardized Student's t , the generalized error distribution, and Hansen's skew t . The 1-step ahead density forecast is

$$\hat{f}_{t+1|t} \stackrel{d}{=} g(\hat{\mu}, \hat{\sigma}_{t+1|t}^2) \tag{8.31}$$

where $f(\cdot)$ is the distribution of returns. This follows directly from the original model since $r_{t+1} = \mu + \sigma_{t+1} e_{t+1}$ and $e_{t+1} \stackrel{\text{i.i.d.}}{\sim} g(0, 1)$.

8.4.2 Semiparametric Density forecasting

Semiparametric density forecasting is also similar to its VaR counterpart. The model begins by assuming that innovations are generated according to some unknown distribution $g(\cdot)$,

$$\begin{aligned}
r_{t+1} &= \mu + \epsilon_{t+1} \\
\sigma_{t+1}^2 &= \omega + \gamma_1 \epsilon_t^2 + \beta_1 \sigma_t^2 \\
\epsilon_{t+1} &= \sigma_{t+1} e_{t+1} \\
e_{t+1} &\stackrel{\text{i.i.d.}}{\sim} g(0, 1).
\end{aligned}$$

and estimates of $\hat{\sigma}_t^2$ are computed assuming that the innovations are conditionally normal. The justification for this choice follows from the strong consistency of the variance parameter estimates even when the innovations are not normal. Using the estimated variances, standardized innovations are computed as $\hat{e}_t = \frac{\epsilon_t}{\hat{\sigma}_t}$. The final step is to compute the density. The simplest method to accomplish this is to compute the empirical CDF as

$$G(e) = \sum_{t=1}^T I_{[\hat{e}_t < e]} \quad (8.32)$$

which simply sums up the number of standardized residuals than e . This method is trivial but has some limitations. First, the PDF does not exist since $G(\cdot)$ is not differentiable. This makes some applications difficult, although a histogram provides a simple, if imprecise, method to work around the non-differentiability of the empirical CDF. Second, the CDF is jagged and is generally an inefficient estimator, particularly in the tails.

An alternative, and more efficient estimator, can be constructed using a kernel to smooth the density. A kernel density is a local average of the number of \hat{e}_t in a small neighborhood of e . The more in this neighborhood, the higher the probability in the region, and the larger the value of the kernel density. The kernel density estimator is defined

$$g(e) = \frac{1}{Th} \sum_{t=1}^T K\left(\frac{\hat{e}_t - e}{h}\right) \quad (8.33)$$

where $K(\cdot)$ can be one of many kernels - the choice of which usually makes little difference - and the normal

$$K(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2) \quad (8.34)$$

or the Epanechnikov

$$K(x) = \begin{cases} \frac{3}{4}(1 - x^2) & -1 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases} \quad (8.35)$$

are the most common. The choice of the bandwidth (h) is more important and is usually set to Silverman's bandwidth, $h = 1.06\sigma T^{-\frac{1}{5}}$ where σ is the standard deviation of \hat{e}_t . However, larger or smaller bandwidths can be used to produce smoother or rougher densities, respectively, and the size of the bandwidth represents a bias-variance tradeoff - a small bandwidth has little bias

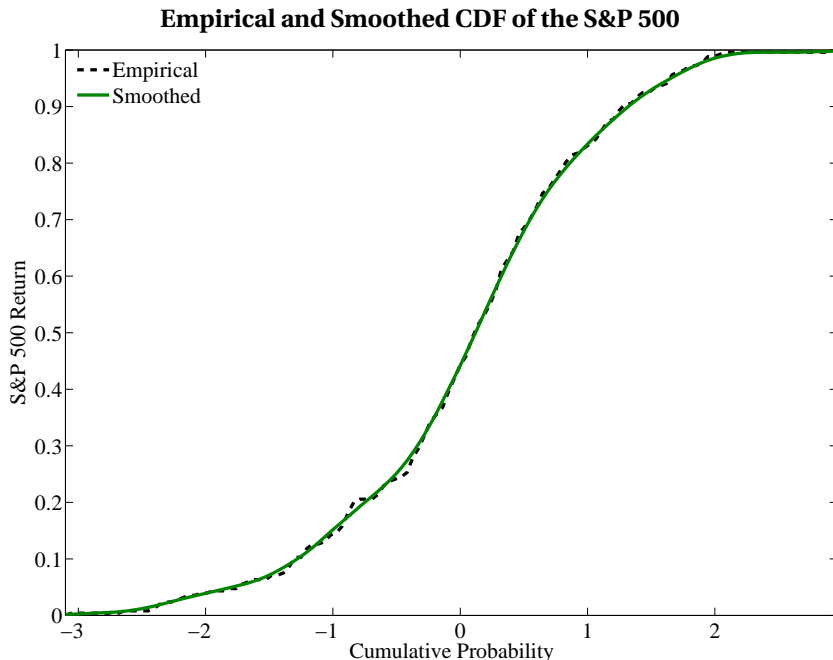


Figure 8.4: The rough empirical and smoothed empirical CDF for standardized returns of the S&P 500 in 2009 (standardized by a TARCH(1,1,1)).

but is very jagged (high variance), while a large bandwidth produces an estimate with substantial bias but very smooth (low variance). If the CDF is needed, $g(e)$ can be integrated using numerical techniques such as a trapezoidal approximation to the Riemann integral.

Finally the density forecast can be constructed by scaling the distribution G by $\sigma_{t+1|t}$ and adding the mean. Figure 8.4 contains a plot of the smooth and non-smoothed CDF of TARCH(1,1,1)-standardized S&P 500 returns in 2009. The empirical CDF is jagged and there are some large gaps in the observed returns.

8.4.3 Multi-step density forecasting and the fan plot

Multi-step ahead density forecasting is not trivial. For example, consider a simple GARCH(1,1) model with normal innovations,

$$\begin{aligned}
 r_{t+1} &= \mu + \epsilon_{t+1} \\
 \sigma_{t+1}^2 &= \omega + \gamma_1 \epsilon_t^2 + \beta_1 \sigma_t^2 \\
 \epsilon_{t+1} &= \sigma_{t+1} e_{t+1}
 \end{aligned}$$

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

The 1-step ahead density forecast of returns is

$$r_{t+1} | \mathcal{F}_t \sim N(\mu, \sigma_{t+1|t}^2). \quad (8.36)$$

Since innovations are conditionally normal and $E_t \left[\sigma_{t+2|t}^2 \right]$ is simple to compute, it is tempting to construct a 2-step ahead forecast using a normal,

$$r_{t+2} | \mathcal{F}_t \sim N(\mu, \sigma_{t+2|t}^2). \quad (8.37)$$

This forecast is not correct since the 2-step ahead distribution is a *variance-mixture* of normals and so is itself non-normal. This reason for the difference is that $\sigma_{t+2|t}^2$, unlike $\sigma_{t+1|t}^2$, is a random variable and the uncertainty must be integrated out to determine the distribution of r_{t+2} . The correct form of the 2-step ahead density forecast is

$$r_{t+2} | \mathcal{F}_t \sim \int_{-\infty}^{\infty} \phi(\mu, \sigma^2(e_{t+1})_{t+2|t+1}) \phi(e_{t+1}) de_{t+1}.$$

where $\phi(\cdot)$ is a normal probability density function and $\sigma^2(e_{t+1})_{t+2|t+1}$ reflects the explicit dependence of $\sigma_{t+2|t+1}^2$ on e_{t+1} . While this expression is fairly complicated, a simpler way to view it is as a mixture of normal random variables where the probability of getting a specific normal depends on $w(e)$,

$$r_{t+2} | \mathcal{F}_t \sim \int_{-\infty}^{\infty} w(e) f(\mu, \sigma(e_{t+1})_{t+2|t+1}) de.$$

Unless $w(e)$ is constant, the resulting distribution will not be a normal. The top panel in figure 8.5 contains the naïve 10-step ahead forecast and the correct 10-step ahead forecast for a simple GARCH(1,1) process,

$$\begin{aligned} r_{t+1} &= \epsilon_{t+1} \\ \sigma_{t+1}^2 &= .02 + .2\epsilon_t^2 + .78\sigma_t^2 \\ \epsilon_{t+1} &= \sigma_{t+1} e_{t+1} \\ e_{t+1} &\stackrel{\text{i.i.d.}}{\sim} N(0, 1) \end{aligned}$$

where $\epsilon_t^2 = \sigma_t = \bar{\sigma} = 1$ and hence $E_t[\sigma_{t+h}] = 1$ for all h . The bottom panel contains the plot of the density of a cumulative 10-day return (the sum of the 10 1-day returns). In this case the naïve model assumes that

$$r_{t+h} | \mathcal{F}_t \sim N(\mu, \sigma_{t+h|t})$$

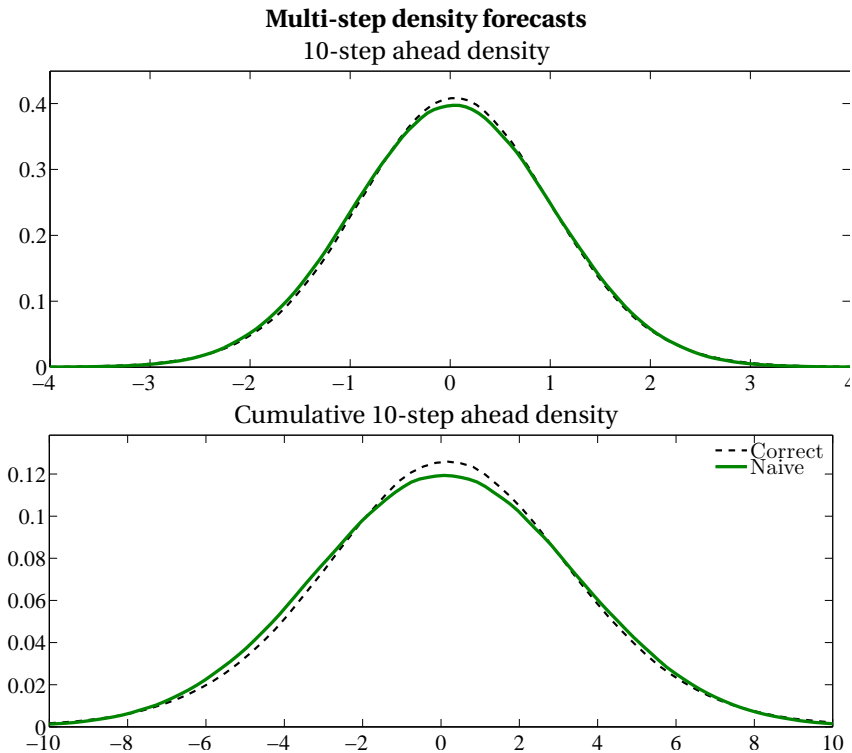


Figure 8.5: Naïve and correct 10-step ahead density forecasts from a simulated GARCH(1,1) model. The correct density forecasts have substantially fatter tails than the naïve forecast as evidenced by the central peak and cross-over of the density in the tails.

for $h = 1, 2, \dots, 10$. The correct forecast has heavier tails than the naïve forecast which can be verified by checking that the solid line is above the dashed line for large deviations.

8.4.3.1 Fan Plots

Fan plots are a graphical method to convey information about future *changes* in uncertainty. Their use has been popularized by the Bank of England and they are a good way to “wow” an audience unfamiliar with this type of plot. Figure 8.6 contains a simple example of a fan plot which contains the density of an irregular ARMA(6,2) which begins at 0 and has i.i.d. standard normal increments. Darker regions indicate higher probability while progressively lighter regions indicate less likely events.

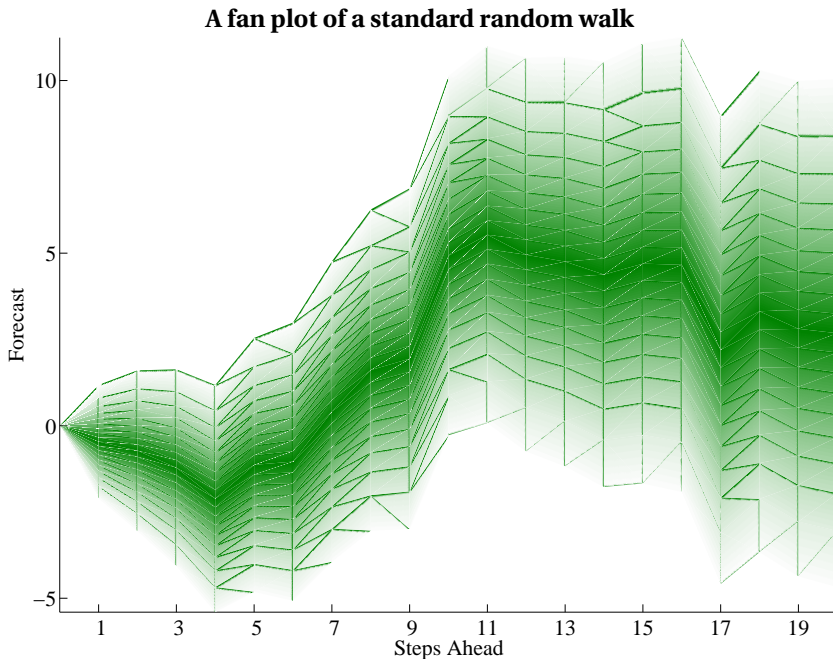


Figure 8.6: Future density of an ARMA(6,2) beginning at 0 with i.i.d. standard normal increments. Darker regions indicate higher probability while progressively lighter regions indicate less likely events.

8.4.4 Quantile-Quantile (QQ) plots

Quantile-Quantile, or QQ, plots provide an informal but simple method to assess the fit of a density or a density forecast. Suppose a set of standardized residuals \hat{e}_t are assumed to have a distribution F . The QQ plot is generated by first ordering the standardized residuals,

$$\hat{e}_1 < \hat{e}_2 < \dots < \hat{e}_{n-1} < \hat{e}_n$$

and then plotting the ordered residual \hat{e}_j against its hypothetical value if the correct distribution were F , which is the inverse CDF evaluated at $\frac{j}{T+1}$, $\left(F^{-1}\left(\frac{j}{T+1}\right)\right)$. This informal assessment of a distribution will be extended into a formal statistic in the Kolmogorov-Smirnov test. Figure 8.7 contains 4 QQ plots for the raw S&P 500 returns against a normal, a t_8 , a t_3 and a GED with $\nu = 1.5$. The normal, t_8 and GED appear to be badly misspecified in the tails – as evidenced through deviations from the 45° line – while the t_3 appears to be a good approximation (the MLE estimate of the Student's t degree of freedom, ν , was approximately 3.1).

QQ plots of S&P 500 returns

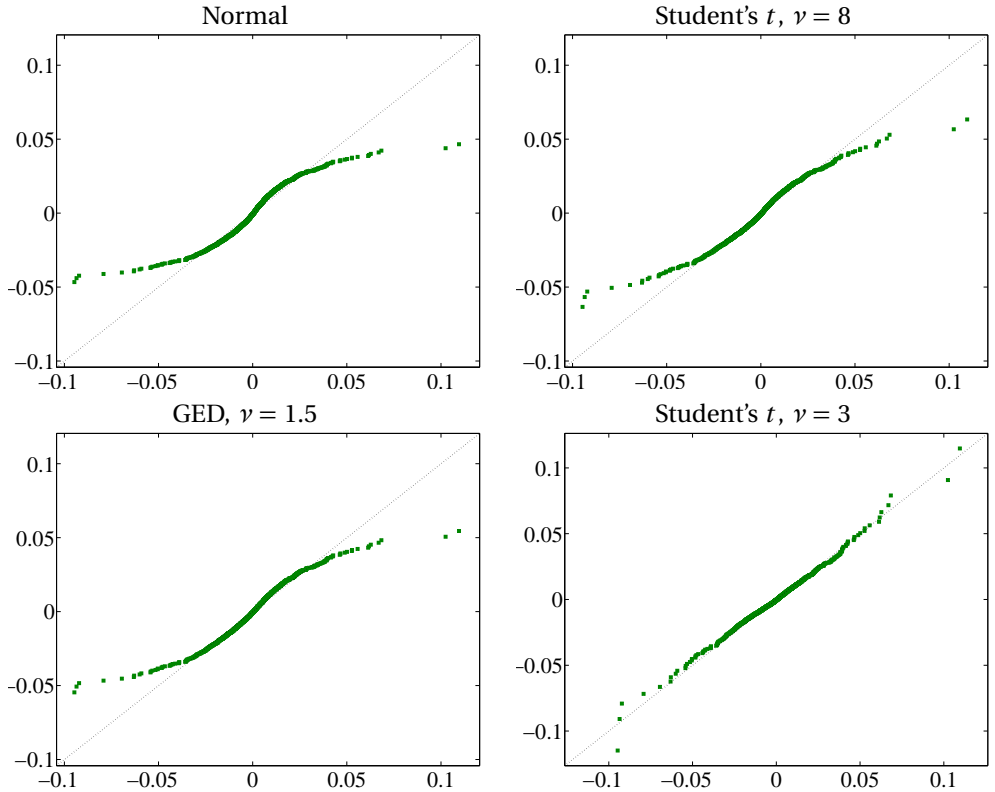


Figure 8.7: QQ plots of the raw S&P 500 returns against a normal, a t_8 , a t_3 and a GED with $\nu = 1.5$. Points along the 45° indicate a good distributional fit.

8.4.5 Evaluating Density Forecasts

All density evaluation strategies are derived from a basic property of random variables: if $x \sim F$, then $u \equiv F(x) \sim U(0, 1)$. That is, for any random variable x , the cumulant of x has a Uniform distribution over $[0, 1]$. The opposite of this results is also true, if $u \sim U(0, 1)$, $F^{-1}(u) = x \sim F$.¹⁰

Theorem 8.1 (Probability Integral Transform). *Let a random variable X have a continuous, increasing CDF $F_X(x)$ and define $Y = F_X(X)$. Then Y is uniformly distributed and $\Pr(Y \leq y) = y$, $0 < y < 1$.*

¹⁰The latter result can be used as the basis of a random number generator. To generate a random number with a CDF of F , a first generate a uniform, u , and then compute the inverse CDF at u to produce a random number from F , $y = F^{-1}(u)$. If the inverse CDF is not available in closed form, monotonicity of the CDF allows for quick, precise numerical inversion.

Theorem 8.1. For any $y \in (0, 1)$, $Y = F_X(X)$, and so

$$\begin{aligned}
 F_Y(y) &= \Pr(Y \leq y) = \Pr(F_X(X) \leq y) \\
 &= \Pr(F_X^{-1}(F_X(X)) \leq F_X^{-1}(y)) && \text{Since } F_X^{-1} \text{ is increasing} \\
 &= \Pr(X \leq F_X^{-1}(y)) && \text{Invertible since strictly increasing} \\
 &= F_X(F_X^{-1}(y)) && \text{Definition of } F_X \\
 &= y
 \end{aligned}$$

□

The proof shows that $\Pr(F_X(X) \leq y) = y$ and so this must be a uniform distribution (by definition).

The Kolmogorov-Smirnov (KS) test exploits this property of residuals from the correct distribution to test whether a set of observed data are compatible with a specified distribution F . The test statistic is calculated by first computing the *probability integral transformed residuals* $\hat{u}_t = F(\hat{\epsilon}_t)$ from the standardized residuals and then sorting them

$$u_1 < u_2 < \dots < u_{n-1} < u_n.$$

The KS test is computed as

$$\begin{aligned}
 KS &= \max_{\tau} \left| \sum_{i=1}^{\tau} I_{[u_j < \frac{\tau}{T}]} - \frac{1}{T} \right| \\
 &= \max_{\tau} \left| \left(\sum_{i=1}^{\tau} I_{[u_j < \frac{\tau}{T}]} \right) - \frac{\tau}{T} \right|
 \end{aligned} \tag{8.38}$$

The test finds the point where the distance between the observed cumulant distribution The objects being maximized over is simply the number of u_j less than $\frac{\tau}{T}$ minus the expected number of observations which should be less than $\frac{\tau}{T}$. Since the probability integral transformed residuals should be $U(0,1)$ when the model is correct, the number of probability integral transformed residuals expected to be less than $\frac{\tau}{T}$ is $\frac{\tau}{T}$. The distribution of the KS test is nonstandard but many software packages contain the critical values. Alternatively, simulating the distribution is trivial and precise estimates of the critical values can be computed in seconds using only uniform *pseudo*-random numbers.

The KS test has a graphical interpretation as a QQ plot of the probability integral transformed residuals against a uniform. Figure 8.8 contains a representation of the KS test using data from two series, the first is standard normal and the second is a standardized students t_4 . 95% confidence bands are denoted with dotted lines. The data from both series were assumed to be standard normal and the t_4 just rejects the null (as evidenced by the cumulants touching the confidence band).

A KS test of normal and standardized t_4 when the data are normal

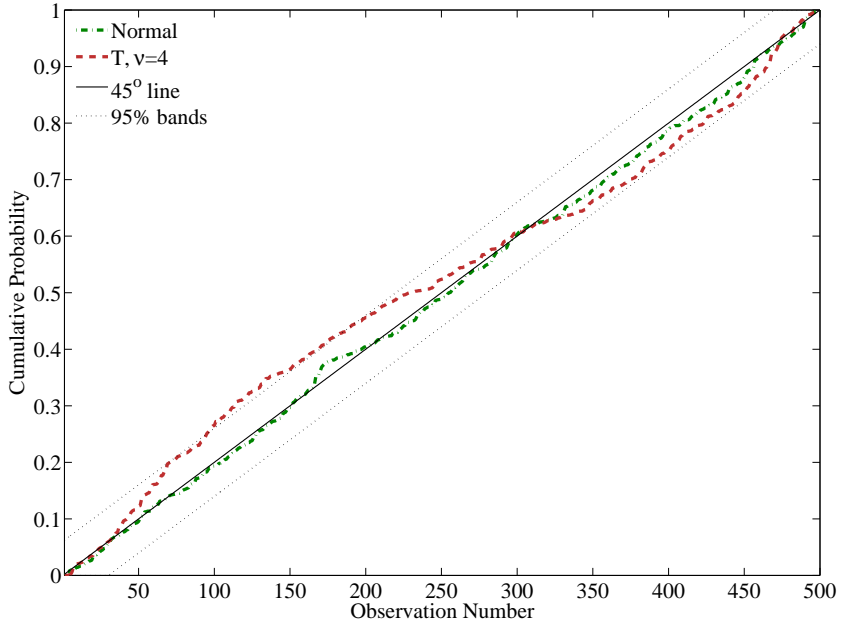


Figure 8.8: A simulated KS test with a normal and a t_4 . The t_4 crosses the confidence boundary indicating a rejection of this specification. A good density forecast should have a cumulative distribution close to the 45° line.

8.4.5.1 Parameter Estimation Error and the KS Test

The critical values supplied by most packages *do not* account for parameter estimation error and KS tests with estimated parameters are generally *less* likely to reject than if the parameters are known. For example, if a sample of 1000 random variables are i.i.d. standard normal and the mean and variance are known to be 0 and 1, the 90, 95 and 99% CVs for the KS test are 0.0387, 0.0428, and 0.0512. If the parameters are not known and must be estimated, the 90, 95 and 99% CVs are reduced to 0.0263, 0.0285, 0.0331. Thus, a desired size of 10% (corresponding to a critical value of 90%) has an actual size closer 0.1% and the test will not reject the null in many instances where it should.

The solution to this problem is simple. Since the KS-test requires knowledge of the entire distribution, it is simple to simulate a sample with length T , and to estimate the parameters and to compute the KS test on the simulated standardized residuals (where the residuals are using estimated parameters). Repeat this procedure B times ($B > 1000$, possibly larger) and then compute the empirical 90, 95 or 99% quantiles from KS_b , $b = 1, 2, \dots, B$. These quantiles are the correct values to use under the null while accounting for parameter estimation uncertainty.

8.4.5.2 Evaluating conditional density forecasts

In a direct analogue to the unconditional case, if $x_t | \mathcal{F}_{t-1} \sim F$, then $\hat{u}_t \equiv F(\hat{x}_t) | \mathcal{F}_{t-1} \stackrel{\text{i.i.d.}}{\sim} U(0, 1)$. That is, probability integral transformed residuals are *conditionally* i.i.d. uniform on $[0, 1]$. While this condition is simple and easy to interpret, direct implementation of a test is not. The Berkowitz (2001) test works around this by re-transforming the probability integral transformed residuals into normals using the inverse Normal CDF. Specifically if $\hat{u}_t = F_{t|t-1}(\hat{e}_t)$ are the residuals standardized by their forecast distributions, the Berkowitz test computes $\hat{y}_t = \Phi^{-1}(\hat{u}_t) = \Phi^{-1}(F_{t|t-1}(\hat{e}_t))$ which have the property, under the null of a correct specification that $\hat{y}_t \stackrel{\text{i.i.d.}}{\sim} N(0, 1)$, an i.i.d. sequence of standard normal random variables.

Berkowitz proposes using a regression model to test the y_t for i.i.d. $N(0, 1)$. The test is implementing by estimating the parameters of

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

via maximum likelihood. The Berkowitz test is computing using a the likelihood ratio test

$$LR = 2(l(\hat{\boldsymbol{\theta}}; \mathbf{y}) - l(\boldsymbol{\theta}_0; \mathbf{y})) \sim \chi_3^2 \quad (8.39)$$

where $\boldsymbol{\theta}_0$ are the parameters if the null is true, corresponding to parameter values of $\phi_0 = \phi_1 = 0$ and $\sigma^2 = 1$ (3 restrictions). In other words, that the y_t are independent normal random variables with a variance of 1. As is always the case in tests of conditional models, the regression model can be augmented to include any time $t - 1$ available instrument and a more general specification is

$$y_t = \mathbf{x}_t \boldsymbol{\gamma} + \eta_t$$

where \mathbf{x}_t may contains a constant, lagged y_t or anything else relevant for evaluating a density forecast. In the general specification, the null is $H_0 : \boldsymbol{\gamma} = \mathbf{0}, \sigma^2 = 1$ and the alternative is the unrestricted estimate from the alternative specification. The likelihood ratio test statistic in the case would have a χ_{K+1}^2 distribution where K is the number of elements in \mathbf{x}_t (the +1 comes from the restriction that $\sigma^2 = 1$).

8.5 Coherent Risk Measures

With multiple measures of risk available, which should be chosen: variance, VaR, or Expected Shortfall? Recent research into risk measures have identified a number of desirable properties for measures of risk. Let ρ be a generic measure of risk that maps the riskiness of a portfolio to an amount of required reserves to cover losses that regularly occur and let P, P_1 and P_2 be portfolios of assets.

Drift Invariance The requires reserved for portfolio P satisfies

$$\rho(P + c) = \rho(P) - c$$

That is, adding a constant return c to P decreases the required reserved by that amount

Homogeneity The required reserved are linear homogeneous,

$$\rho(\lambda P) = \lambda \rho(P) \quad \text{for any } \lambda > 0 \quad (8.40)$$

The homogeneity property states that the required reserves of two portfolios with the same relative holdings of assets depends linearly on the scale – doubling the size of a portfolio while not altering its relative composition generates twice the risk, and requires twice the reserves to cover regular losses.

Monotonicity If P_1 first order stochastically dominates P_2 , the required reserves for P_1 must be less than those of P_2

$$\rho(P_1) \leq \rho(P_2) \quad (8.41)$$

If P_1 FOSD P_2 then the value of portfolio P_1 will be larger than the value of portfolio P_2 in every state of the world, and so the portfolio must be less risky.

Subadditivity The required reserves for the combination of two portfolios is less then the required reserves for each treated separately

$$\rho(P_1 + P_2) \leq \rho(P_1) + \rho(P_2) \quad (8.42)$$

Definition 8.6 (Coherent Risk Measure). Any risk measure which satisfies these four properties is known as *coherent*.

Coherency seems like a good thing for a risk measure. The first three conditions are indisputable. For example, in the third, if P_1 FOSD P_2 , then P_1 will always have a higher return and must be less risky. The last is somewhat controversial.

Theorem 8.2 (Value-at-Risk is not Coherent). *Value-at-Risk is not coherent since it fails the subadditivity criteria. It is possible to have a VaR which is superadditive where the Value-at-Risk of the combined portfolio is greater than the sum of the Values-at-Risk of either portfolio.*

Examples of the superadditivity of VaR usually require the portfolio to depend nonlinearly on some assets (i.e. hold derivatives). Expected Shortfall, on the other hand, is a coherent measure of risk.

Theorem 8.3 (Expected Shortfall is Coherent). *Expected shortfall is a coherent risk measure.*

The coherency of Expected Shortfall is fairly straight forward to show in many cases (for example, if the returns are jointly normally distributed) although a general proof is difficult and provides little intuition. However, that Expected Shortfall is coherent and VaR is not does not make Expected Shortfall a better choice. VaR has a number of advantages for measuring risk since it only requires the modeling of a quantile of the return distribution, VaR always exists and

is finite and there are many widely tested methodologies for estimating VaR. Expected Shortfall requires an estimate of the mean in the tail which is substantially harder than simply estimating the VaR and may not exist in some cases. Additionally, in most realistic cases, increases in the Expected Shortfall will be accompanied with increases in the VaR and they will both broadly agree about the risk of a the portfolio.

Shorter Problems

Problem 8.1. Discuss any properties the generalized error should have when evaluating Value-at-Risk models.

Problem 8.2. Define and contrast Historical Simulation and Filtered Historic Simulation?

Problem 8.3. Define expected shortfall. How does this extend the idea of Value-at-Risk?

Problem 8.4. Why are HITs useful for testing a Value-at-Risk model?

Problem 8.5. Define conditional Value-at-Risk. Describe two methods for estimating this and compare their strengths and weaknesses.

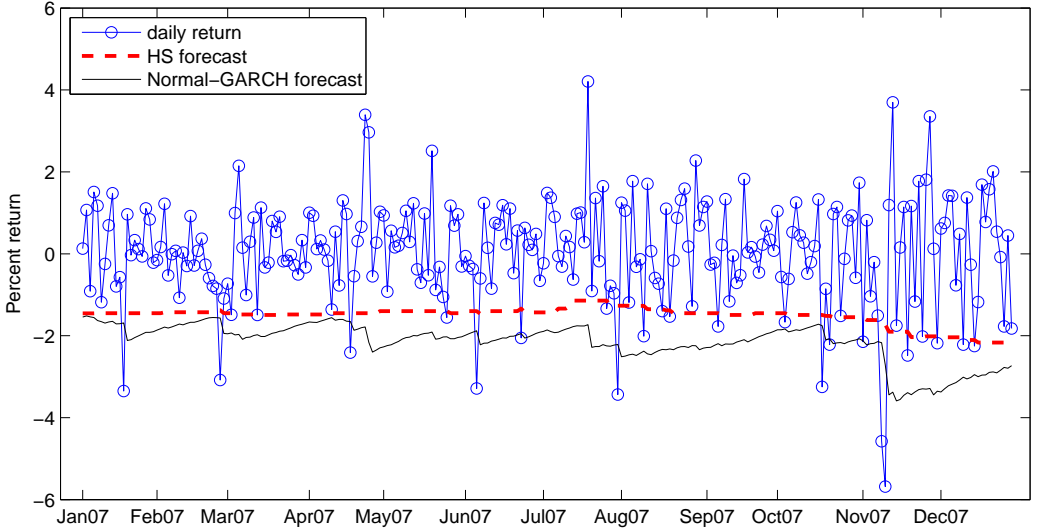
Longer Exercises

Exercise 8.1. Precisely answer the following questions

1. What is VaR?
2. What is expected shortfall?
3. Describe two methods to estimate the VaR of a portfolio? Compare the strengths and weaknesses of these two approaches.
4. Suppose two bankers provide you with VaR forecasts (which are different) and you can get data on the actual portfolio returns. How could you test for superiority? What is meant by better forecast in this situation?

Exercise 8.2. The figure below plots the daily returns on IBM from 1 January 2007 to 31 December 2007 (251 trading days), along with 5% Value-at-Risk (VaR) forecasts from two models. The first model (denoted “HS”) uses ‘historical simulation’ with a 250-day window of data. The second model uses a GARCH(1,1) model, assuming that daily returns have a constant conditional mean, and are conditionally Normally distributed (denoted ‘Normal-GARCH’ in the figure).

Daily returns on IBM in 2007, with 5% VaR forecasts



1. Briefly describe **one** other model for VaR forecasting, and discuss its pros and cons relative to the 'historical simulation' model and the Normal-GARCH model.
2. For each of the two VaR forecasts in the figure, a sequence of 'hit' variables was constructed:

$$Hit_t^{HS} = \mathbf{1} \left\{ r_t \leq \widehat{VaR}_t^{HS} \right\}$$

$$Hit_t^{GARCH} = \mathbf{1} \left\{ r_t \leq \widehat{VaR}_t^{GARCH} \right\}$$

where $\mathbf{1} \{ r_t \leq a \} = \begin{cases} 1, & \text{if } r_t \leq a \\ 0, & \text{if } r_t > a \end{cases}$

and the following regression was run (standard errors are in parentheses below the parameter estimates):

$$Hit_t^{HS} = 0.0956 + u_t$$

(0.0186)

$$Hit_t^{GARCH} = 0.0438 + u_t$$

(0.0129)

- (a) How can we use the above regression output to test the accuracy of the VaR forecasts from these two models?
 - (b) What do the tests tell us?
3. Another set of regressions was also run (standard errors are in parentheses below the

parameter estimates):

$$\begin{aligned} Hit_t^{HS} &= \underset{(0.0196)}{0.1018} - \underset{(0.0634)}{0.0601} Hit_{t-1}^{HS} + u_t \\ Hit_t^{GARCH} &= \underset{(0.0133)}{0.0418} + \underset{(0.0634)}{0.0491} Hit_{t-1}^{GARCH} + u_t \end{aligned}$$

A joint test that the intercept is 0.05 and the slope coefficient is zero yielded a chi-squared statistic of 6.9679 for the first regression, and 0.8113 for the second regression.

- (a) Why are these regressions potentially useful?
- (b) What do the results tell us? (The 95% critical values for a chi-squared variable with q degrees of freedom are given below:)

q	95% critical value
1	3.84
2	5.99
3	7.81
4	9.49
5	11.07
10	18.31
25	37.65
249	286.81
250	287.88
251	288.96

Exercise 8.3. Figure 8.9 plots the daily returns from 1 January 2008 to 31 December 2008 (252 trading days), along with 5% Value-at-Risk (VaR) forecasts from two models. The first model (denoted “HS”) uses ‘historical simulation’ with a 250-day window of data. The second model uses a GARCH(1,1) model, assuming that daily returns have a constant conditional mean, and are conditionally Normally distributed (denoted ‘Normal-GARCH’ in the figure).

1. Briefly describe **one** other model for VaR forecasting, and discuss its pros and cons relative to the ‘historical simulation’ model and the Normal-GARCH model.
2. For each of the two VaR forecasts in the figure, a sequence of ‘hit’ variables was constructed:

$$\begin{aligned} Hit_t^{HS} &= \mathbf{1} \left\{ r_t \leq \widehat{VaR}_t^{HS} \right\} \\ Hit_t^{GARCH} &= \mathbf{1} \left\{ r_t \leq \widehat{VaR}_t^{GARCH} \right\} \\ \text{where } \mathbf{1} \{ r_t \leq a \} &= \begin{cases} 1, & \text{if } r_t \leq a \\ 0, & \text{if } r_t > a \end{cases} \end{aligned}$$

and the following regression was run (standard errors are in parentheses below the parameter estimates):

$$\begin{aligned} Hit_t^{HS} &= 0.0555 + u_t \\ &\quad (0.0144) \\ Hit_t^{GARCH} &= 0.0277 + u_t \\ &\quad (0.0103) \end{aligned}$$

- (a) How can we use the above regression output to test the accuracy of the VaR forecasts from these two models?
- (b) What do the tests tell us?

3. Another set of regressions was also run (standard errors are in parentheses below the parameter estimates):

$$\begin{aligned} Hit_t^{HS} &= 0.0462 + 0.1845Hit_{t-1}^{HS} + u_t \\ &\quad (0.0136) \quad (0.1176) \\ Hit_t^{GARCH} &= 0.0285 - 0.0285Hit_{t-1}^{GARCH} + u_t \\ &\quad (0.0106) \quad (0.0106) \end{aligned}$$

A joint test that the intercept is 0.05 and the slope coefficient is zero yielded a chi-squared statistic of 8.330 for the first regression, and 4.668 for the second regression.

- (a) Why are these regressions potentially useful?
- (b) What do the results tell us? (The 95% critical values for a chi-squared variable with q degrees of freedom are given below:)

4. Comment on the similarities and/or differences between what you found in (b) and (c).

q	95% critical value
1	3.84
2	5.99
3	7.81
4	9.49
5	11.07
10	18.31
25	37.65
249	286.81
250	287.88
251	288.96

Exercise 8.4. Answer the following question:

1. Assume that X is distributed according to some distribution F , and that F is continuous and strictly increasing. Define $U \equiv F(X)$. Show that $U \sim \text{Uniform}(0, 1)$.

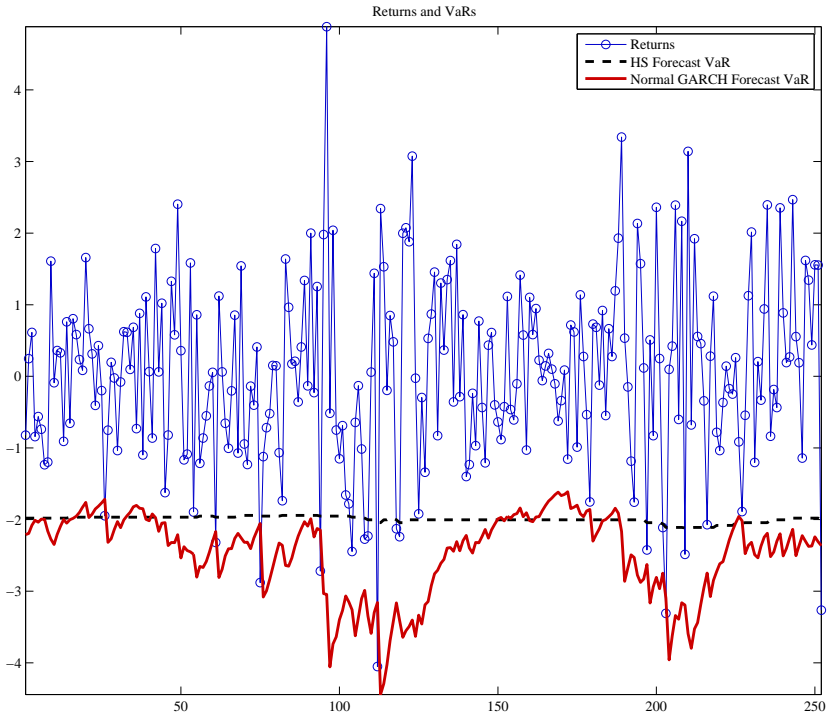


Figure 8.9: Returns, Historical Simulation VaR and Normal GARCH VaR.

2. Assume that $V \sim \text{Uniform}(0, 1)$, and that G is some continuous and strictly increasing distribution function. If we define $Y \equiv G^{-1}(V)$, show that $Y \sim G$.

For the next two parts, consider the problem of forecasting the time taken for the price of a particular asset (P_t) to reach some threshold (P^*). Denote the time (in days) taken for the asset to reach the threshold as Z_t . Assume that the true distribution of Z_t is Exponential with parameter $\beta \in (0, \infty)$:

$$Z_t \sim \text{Exponential}(\beta)$$

$$\text{so } F(z; \beta) = \begin{cases} 1 - \exp\{-\beta z\}, & z \geq 0 \\ 0, & z < 0 \end{cases}$$

Now consider a forecaster who gets the distribution correct, but the parameter wrong. Denote her distribution forecast as $\hat{F}(z) = \text{Exponential}(\hat{\beta})$.

3. If we define $U \equiv \hat{F}(Z)$, show that $\Pr[U \leq u] = 1 - (1 - u)^{\beta/\hat{\beta}}$ for $u \in (0, 1)$, and interpret.

4. Now think about the case where $\hat{\beta}$ is an estimate of β , such that $\hat{\beta} \xrightarrow{P} \beta$ as $n \rightarrow \infty$. Show that $\Pr[U \leq u] \xrightarrow{P} u$ as $n \rightarrow \infty$, and interpret.

Exercise 8.5. A Value-at-Risk model was fit to some return data and the series of 5% VaR violations was computed. Denote these \widetilde{HIT}_t . The total number of observations was $T = 50$, and the total number of violations was 4.

1. Test the null that the model has unconditionally correct coverage using a t -test.
2. Test the null that the model has unconditionally correct coverage using a LR test. The likelihood for a Bernoulli(p) random Y is

$$f(y; p) = p^y (1 - p)^{1-y}.$$

The following regression was estimated

$$\widetilde{HIT}_t = 0.0205 + 0.7081\widetilde{HIT}_{t-1} + \hat{\eta}_t$$

The estimated asymptotic covariance of the parameters is

$$\hat{\sigma}^2 \hat{\Sigma}_{XX}^{-1} = \begin{bmatrix} 0.0350 & -0.0350 \\ -0.0350 & 0.5001 \end{bmatrix}, \text{ and } \hat{\Sigma}_{XX}^{-1} \hat{\mathbf{S}} \hat{\Sigma}_{XX}^{-1} = \begin{bmatrix} 0.0216 & -0.0216 \\ -0.0216 & 2.8466 \end{bmatrix}$$

where $\hat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^T \hat{\eta}_t^2$, $\hat{\Sigma}_{XX} = \frac{1}{T} \mathbf{X}'\mathbf{X}$ and $\hat{\mathbf{S}} = \frac{1}{T} \sum_{t=1}^T \hat{\eta}_t^2 \mathbf{x}_t' \mathbf{x}_t$.

3. Is there evidence that the model is dynamically mis-specified, ignoring the unconditional rate of violations?
4. Compute a joint test that the model is completely correctly specified. Note that

$$\begin{bmatrix} a & b \\ b & c \end{bmatrix}^{-1} = \frac{1}{ac - b^2} \begin{bmatrix} c & -b \\ -b & a \end{bmatrix}.$$

Note: The 5% critical values of a χ^2_v are

ν	CV
1	3.84
2	5.99
3	7.81
47	64.0
48	65.1
49	66.3
50	67.5

