Chapter 8

Value-at-Risk, Expected Shortfall and Density Forecasting

Alternative references for this chapter include Christoffersen (2003), which is a highly accessible introduction, Gourieroux and Jasiak (2009), who provide additional technical details, and, McNeil, Frey, and Embrechts (2005), who provide a comprehensive and technical treatment of risk measurement.

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 introduces three methods used to assess the riskiness of a portfolio: Value-at-Risk (VaR), Expected Shortfall, and modeling the entire density of the portfolio's return.

8.1 Defining Risk

Portfolios are exposed to multiple distinct sources of risk. The most important sources of risk can be classified into one of six categories.

Market Risk

Market risk describes the uncertainty about the future price of an asset due to changes in fundamentals or beliefs. For example, market risk captures changes in asset prices due to macroeconomics announcements such as FOMC policy rate updates or non-farm payroll releases.

Liquidity risk

Liquidity risk complements market risk by measuring the loss involved if a position must be rapidly unwound. For example, if a fund wished to sell 20,000,000 shares of IBM on a single day, which has a typical daily volume of 5,000,000, 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

Credit risk, also known as default risk, covers cases where a 3rd 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

Counterparty risk extends credit risk to instruments other than bonds and captures the event that a counterparty to a transaction, for example, the seller of an option contract, is unable to complete the transaction at expiration. Counterparty risk was a significant 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

Model risk represents an econometric form of risk that measures the uncertainty about the correct form of the model used to compute the price of the asset 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

Estimation risk captures an aspect of risk that is present whenever estimated parameters are used in econometric models to price securities or assess risk. Estimation risk is distinct from model risk since it is present even if a model when 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 more likely to be misspecified but may have less parameter estimation uncertainty.

This chapter deals exclusively with market risk. Liquidity, credit risk and counterparty risk all require specialized treatment beyond the scope of this course. Model evaluation, especially out-of-sample evaluation, is the primary tool for assessing model and estimation risks.

8.2 Value-at-Risk (VaR)

The most widely reported measure of risk is Value-at-Risk (VaR). The VaR of a portfolio is a measure of the risk in the left tail of portfolio's return over some period, often a day or a week. VaR provides a more sensible measure of the risk of the portfolio than variance since it focuses on losses. VaR is not a perfect measure of risk, and the issues with VaR are detailed in the context of coherent risk measures (section 8.8).

8.2.1 Value-at-Risk Defined

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

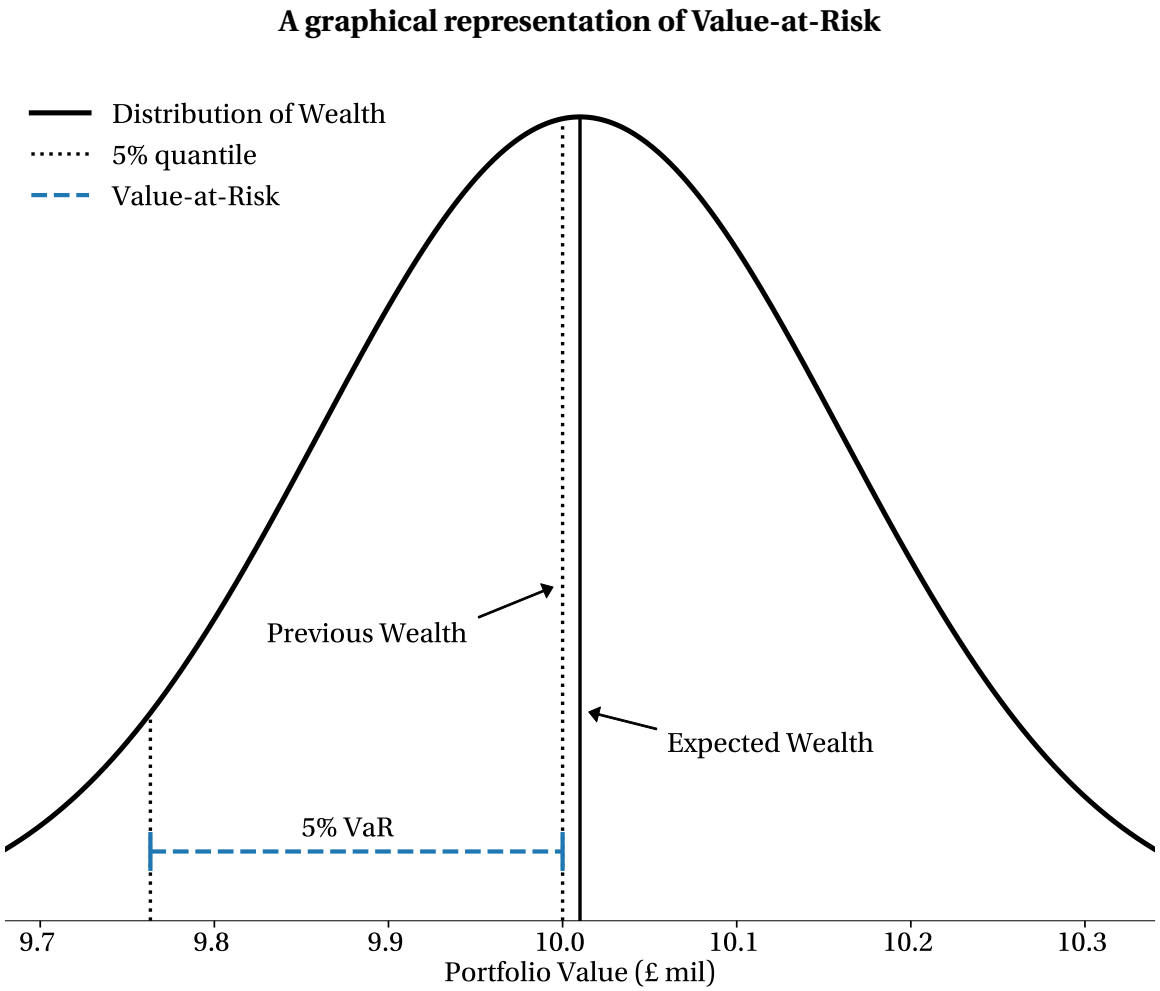


Figure 8.1: A graphical representation of Value-at-Risk. The VaR is represented by the magnitude of the horizontal bar and measures the distance between the value of the portfolio in the current period and the α-quantile of the portfolio value distribution. In this example, $\alpha = 5\%$, the value of the portfolio's assets is £10,000,000, 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 change in the portfolio such that the probability that the loss in portfolio value over a specified horizon is greater than the VaR is α,

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

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 is

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

where $\Phi^{-1}(\cdot)$ is the inverse CDF of a standard normal. This expression may appear backward – it is not. The negative sign on the mean indicates that increases in the mean *decrease* the VaR. 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 to remove the initial value portfolio from the measure.

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 a specified horizon is less than $-1\times$%VaR is α,

$$Pr(r_t < -\%\text{VaR}) = \alpha \tag{8.2}$$

where $r_t = (W_t - W_{t-1})/W_{t-1}$ is the return of 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 focuses on percentage VaR.

8.2.2 The relationship between VaR and quantiles

Understanding that VaR and quantiles are fundamentally related provides a key insight. If r is 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.[1]

8.3 Conditional Value-at-Risk

Most applications of VaR are used to measure risk over short horizons, and so require a conditional Value-at-Risk. Conditioning employs information up to time t to produce a VaR in period $t+h$.

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

$$Pr(r_{t+1} < -\text{VaR}_{t+1|t}|\mathcal{F}_t) = \alpha \tag{8.3}$$

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 (e.g., one day, two weeks, or a month)

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 are presented in the order of the strength of the assumptions required to justify the method, from strongest to weakest.

[1] It is theoretically possible for VaR to be negative. If the VaR of a portfolio is negative, either the portfolio has no risk, the portfolio manager extremely skillful, or most likely the model used to compute the VaR is badly misspecified.

8.3.1 RiskMetrics©

The RiskMetrics group has produced a simple, robust method for producing conditional VaR. The basic structure of the RiskMetrics model relies on a restricted GARCH(1,1) where $\alpha + \beta = 1$ and $\omega = 0$. The estimate of the portfolio's variance is

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

where r_t is the (percentage) return on the portfolio in period t. In the RiskMetrics specification σ_{t+1}^2 follows an EWMA which places weight $\lambda^j(1-\lambda)$ on r_{t-j}^2.[2] The RiskMetrics model does not include a conditional mean of the portfolio return, and so is only applicable to assets with returns that are close to zero. The restriction limits the applicability to applications where the risk-measurement horizon is short (e.g., one day to one month). The VaR is constructed from the α-quantile of a normal distribution,

$$\text{VaR}_{t+1} = -\sigma_{t+1}\Phi^{-1}(\alpha) \tag{8.5}$$

where $\Phi^{-1}(\cdot)$ is the inverse normal CDF. The RiskMetrics model has are no parameters to estimate; λ has been calibrated to .94 for daily data, 0.97 for weekly data, and .99 for monthly data.[3] This model can also be extended to multiple assets using 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, $\mathbf{\Sigma}_{t+1}$. The limitations of the RiskMetrics model are that the parameters aren't estimated (which is also an advantage), the model does not account for a leverage effect, and the VaR follows a random walk since $\lambda+(1-\lambda) = 1$.

8.3.2 Parametric ARCH Models

Parametric ARCH-family models provide a complete description of the future return distribution, and so can be applied to estimate the VaR of a portfolio. This model is highly adaptable since the mean, variance and distribution can all be tailored to the portfolio's historical returns. For simplicity, this example uses a constant mean and has a GARCH(1,1) variance process.[4]

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

where $F(0,1)$ is used to indicate that the distribution of innovations need not be normally distributed but must have mean 0 and variance 1. For example, F could be a standardized Student's

[2] An EWMA differs from a standard moving average in two ways. First, an EWMA places relatively more weight on recent observations than on observation in the distant past. Second, EWMAs depend on the entire history rather than a fixed-length window.

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

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

t with ν degrees of freedom or Hansen's skewed t with a 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}. The flexibility to build a model by specifying the mean, variance and distributions is the strength of this approach. 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 is 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.

8.3.3 Semiparametric ARCH Models/Filtered Historical Simulation

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

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

where $G(0,1)$ is an unknown distribution with mean zero and variance 1. Conditional VaR estimates from semiparametric models are also known as filtered Historical Simulation, due to their similarity to Historical Simulation (see). The ARCH model filters the return data by removing the conditional mean and volatility.

When the distribution of the standardized residuals $G(\cdot)$ is unknown, maximum likelihood estimation cannot be used to estimate model parameters. 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 converge to their true values for most any $G(\cdot)$. The model is estimated using QMLE by assuming that the errors are normally distributed and then the Value-at-Risk for the α-quantile can be computed

$$\text{VaR}_{t+1}(\alpha) = -\hat{\mu} - \hat{\sigma}_{t+1} \hat{G}_{\alpha}^{-1} \tag{8.6}$$

where $\hat{G}_{\alpha}^{-1}$ is the empirical α-quantile of of the standardized returns, $\{\hat{e}_{t+1}\}$. To estimate this quantile, define $\hat{e}_{t+1} = {}^{\hat{\epsilon}_{t+1}}/_{\hat{\sigma}_{t+1}}$. and order the errors such that

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

[5] Semiparametric estimators combine parametric and nonparametric estimators in a single model. In time-series applications, semiparametric estimators have parametric models for the dynamics of the mean and variance but use a nonparametric estimator of the distribution of the residuals.

Here 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] The estimate of G^{-1} is the α-quantile of the empirical distribution of $\hat{e}_{t+1}$ which is the value in position αn of the ordered standardized residuals.

Semiparametric ARCH models provide one clear advantage over their parametric ARCH cousins; the quantile, and hence the VaR, is 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 $\hat{G}_\alpha^{-1}$ may be poorly estimated – especially if α is very small (e.g., 1%). Semiparametric ARCH models also share the limitation they are only applicable when returns are generated by a location-scale distribution.

8.3.4 Cornish-Fisher Approximation

The Cornish-Fisher estimator of VaR lies between fully parametric model and the semiparametric model. The setup is identical to that of the semiparametric model,

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

where $G(\cdot)$ is an unknown distribution. The model parameters are estimated by QML assuming that the conditional distribution of residuals is normal to produce standardized residuals, $\hat{e}_{t+1} = {}^{\hat{\epsilon}_{t+1}}/_{\hat{\sigma}_{t+1}}$. The Cornish-Fisher approximation is a Taylor-series-like expansion of the α-quantile around the α-quantile of a normal and is given by

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

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

where ς and κ are the skewness and kurtosis of $\hat{e}_{t+1}$, respectively. From the expression for $G_{CF}^{-1}(\alpha)$, negative skewness and excess kurtosis ($\kappa > 3$, the kurtosis of a normal) decrease the 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 semiparametric 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 very heavy-tailed distributions, the third and fourth moments may not even exist.

[6] When 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.

8.3.5 Conditional Autoregressive Value-at-Risk (CaViaR)

Engle and Manganelli (2004) developed a family of ARCH-like models to estimate the conditional Value-at-Risk using quantile regression. CaViaR models have a similar structure to GARCH models. The α-quantile of the return distribution, $F^{-1}_{\alpha,t+1}$, is modeled as a weighted average of a constant, the previous value of the quantile, and a shock (or surprise). The shock can take many forms although a "*HIT*", defined as an exceedance of the previous Value-at-Risk, is the most natural.

$$HIT_{t+1} = I_{[r_{t+1}<F^{-1}_{t+1}]} - \alpha \tag{8.9}$$

where r_{t+1} the (percentage) return and $F^{-1}_{\alpha,t+1}$ is the time t α-quantile of this distribution. When $F^{-1}_{\alpha,t+1}$ is the conditional quantile of the return distribution, then a *HIT* is mean zero $\mathrm{E}_t\left[I_{[r_{t+1}<F^{-1}_{\alpha,t+1}]}\right] = \Pr\left(r_{t+1} < F^{-1}_{\alpha,t+1}\right) = \alpha$.

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

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

Other forms that have been explored include the symmetric absolute value,

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

the asymmetric absolute value,

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

the indirect GARCH,

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

The parameters of CaViaR models are estimated by minimizing the "tick" loss function

$$\begin{aligned}
&\arg\min_{\boldsymbol{\theta}} \; T^{-1} \sum_{t=1}^{T} \underbrace{\alpha(r_t - q_t)(1 - I_{[r_t<q_t]})}_{\text{Positive errors}} + \underbrace{(1-\alpha)(q_t - r_t)I_{[r_t<q_t]}}_{\text{Negative Errors}} = \\
&\arg\min_{\boldsymbol{\theta}} \; T^{-1} \sum_{t=1}^{T} \alpha(r_t - q_t) + (q_t - r_t)I_{[r_t<q_t]}
\end{aligned} \tag{8.14}$$

where $I_{[r_t<q_t]}$ is an indicator variable which is 1 if $r_t < q_t$ and 0 otherwise. The loss function is linear in the error $r_t - q_t$ and has a slope of α for positive errors and $1-\alpha$ for negative errors. Estimation of the parameters is complicated since the objective function may is non-differentiable and has many flat spots. Derivative-free methods, such as the Nelder-Mead simplex method or genetic algorithms, can be used to overcome this difficulty. The VaR in a CaViaR framework is then

$$\mathrm{VaR}_{t+1} = -q_{t+1} = -\hat{F}^{-1}_{t+1} \tag{8.15}$$

CaViaR model does not specify a distribution of returns or any moments, and so 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.

8.3.6 Weighted Historical Simulation

Weighted historical simulation constructs an empirical distribution where recent returns are given more weight than returns further in the past. The estimator is nonparametric since that no specific assumptions about either distribution or the dynamics of returns are made.

Weights are assigned using an exponentially declining function. If returns are available from $i = 1, \ldots, t$, then the weight given to data point i is

$$w_i = \lambda^{t-i}(1-\lambda)/\left(1-\lambda^t\right), \quad i = 1, 2, \ldots, t.$$

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 produce a VaR that is more "local" while larger values produce VaR estimates based most of the historical sample.

The weighted empirical CDF is then

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

The conditional VaR is then computed as the solution to

$$\text{VaR}_{t+1} = \min_r \hat{G}(r) \geq \alpha$$

which chooses the smallest value of r where the there is at least α probability below in the weighted cumulative distribution.

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

The concepts of VaR is illustrated using S&P 500 returns from January 1, 1999, until December 31, 2018. A variety of models have been estimated that all produce similar VaR estimates. Alternative distributional assumptions generally produce similar volatility parameter estimates in ARCH models, and so VaR estimates only differ due to differences in the quantiles. Table 8.1 reports parameter estimates from these models. The volatility parameters of the TARCH models were virtually identical across all three distributional assumptions. The degree of freedom parameter $\hat{\nu} \approx 8$ in both the standardized Student's t and the skewed t indicating that the standardizes residuals are leptokurtotic, and $\hat{\lambda} \approx -.1$, from the skewed t, indicating some negative skewness. The CaViaR estimates indicate little change in the conditional quantile for positive shock, a substantial increase in the VaR when the return is negative, and that the conditional quantile is highly persistent. The table also contains estimated quantiles using the parametric, semiparametric and Cornish-Fisher expansion of the normal. Since the fit conditional variances

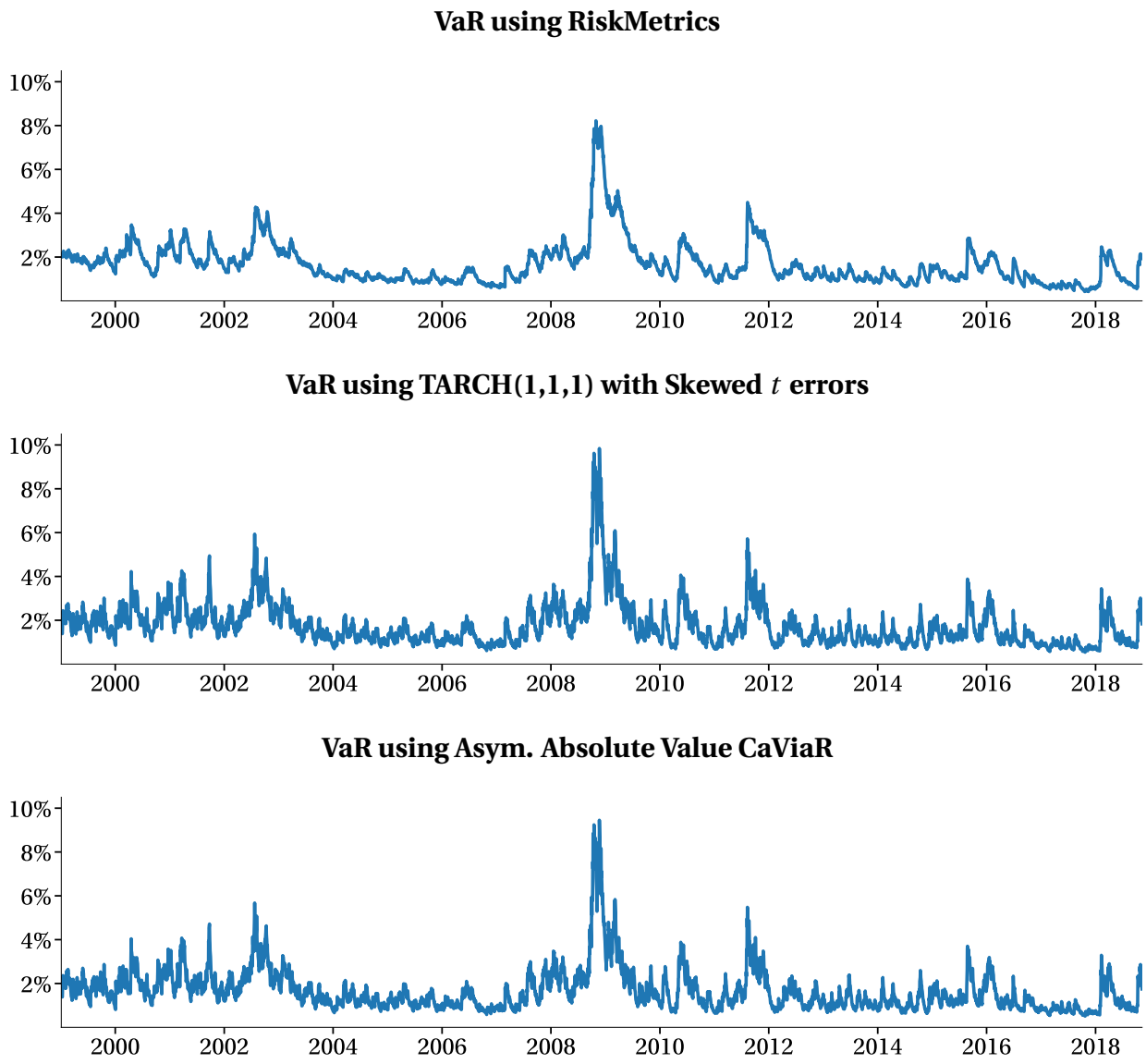


Figure 8.2: The figure contains the estimated 5% VaR for the S&P 500 using data from 1999 until the end of 2018. While these three models have different specifications for the evolution of the conditional VaR, the estimated VaRs are remarkably similar.

---

were nearly identical, the only meaningful difference in the VaRs comes from the differences in these quantiles. These are all qualitatively similar except at 1%.

Figure 8.2 plots the fitted VaRs from the RiskMetrics model, a TARCH with skewed t errors and an asymmetric absolute value CaViaR. All three plots appear very similar, and the TARCH and CaViaR model fits are virtually identical. This similarity is due to the common structure of the dynamics and the values of the estimated parameters. Figure 8.3 plots the conditional VaRs for the weighted Historical Simulation estimator for three values of the smoothing parameter λ. The three values of λ, 0.95, 0.99, and 0.995, places 90% of the weight on the most recent 45, 230 and 2280 observations, respectively. The different values of the smoothing parameter produce

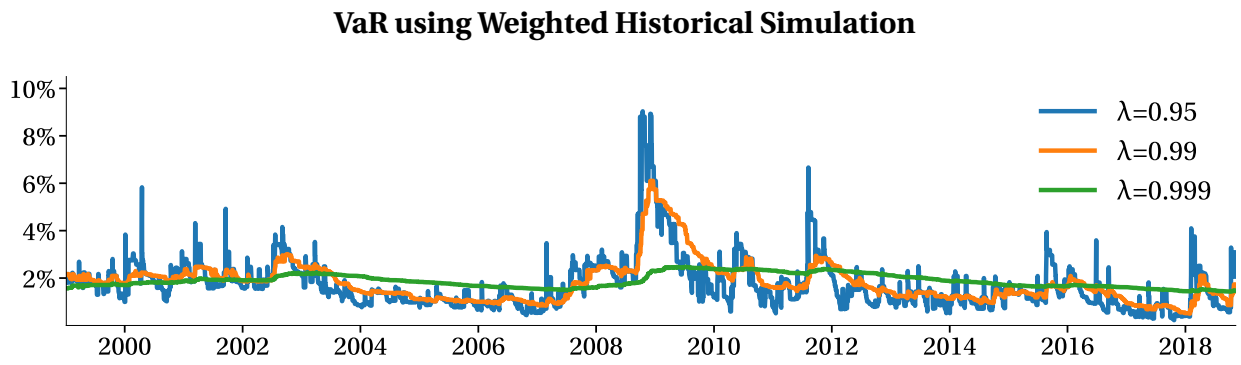


Figure 8.3: The estimated 5% VaR for the S&P 500 using weighted Historical Simulation for $\lambda \in \{0.95, 0.99, 0.999\}$. The three values of λ place 90% of the weight on the most recent 45, 230, and 2280 observations, respectively. Larger values of the decay parameter λ produce smoother conditional VaR estimates.

meaningfully different conditional VaR estimates. The smallest value appears to produce the conditional VaR estimates are most similar to those depicted in Figure 8.2.

8.4 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 models that do not make use of conditioning information to estimate the VaR directly from the historical return data.

8.4.1 Parametric Estimation

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

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

The parameters of the distribution are 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 $\boldsymbol{\theta}$ is assumed, $r_t \sim F(\boldsymbol{\theta})$ 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 general enough to capture the true distribution is difficult.

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	0.026	0.000	0.172	0.909		
Student's t	0.020	0.000	0.173	0.913	7.926	
Skew t	0.022	0.000	0.179	0.910	8.520	-0.123

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	0.035	0.002	0.290	0.910

Estimated Quantiles from Parametric and Semi-parametric TARCH models

	Semiparam.	Normal	Stud. t	Skew t	CF
1%	-2.656	-2.326	-2.510	-2.674	-2.918
5%	-1.705	-1.645	-1.610	-1.688	-1.739
10%	-1.265	-1.282	-1.209	-1.247	-1.237

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).

8.4.2 Nonparametric Estimation/Historical Simulation

At the other end of the spectrum is a simple nonparametric estimate of the unconditional VaR known as Historical Simulation. As was the case in the semiparametric conditional VaR, the first step is to sort the returns so that

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

where $n = T$ is used to denote an ordering not based on time. The VaR is estimated using $r_{\lfloor \alpha n \rfloor}$ or $r_{\lceil \alpha n \rceil}$ (or any value between the two). The estimate of the VaR is the α-quantile of the empirical distribution of $\{r_t\}$,

$$\text{VaR} = -\hat{G}_\alpha^{-1} \tag{8.17}$$

where $\hat{G}_\alpha^{-1}$ is the estimated quantile. The empirical CDF is defined

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

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 r.

Historical simulation estimates are rough, and a single new data point may produce very different VaR estimates when the sample size is small. 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. Smoothing the distribution is most beneficial when the sample size is small. See section 8.7.2 for more details.

The advantage of nonparametric estimates of VaR is that they are generally consistent under minimal assumptions about the distribution of returns 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).

8.4.3 Parametric Monte Carlo

Parametric Monte Carlo is meaningfully different from either parametric or nonparametric estimation of the unconditional distribution. Rather than fit a model to the returns directly, parametric Monte Carlo fits a parsimonious *conditional* model. This model 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,

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

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

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

where $\hat{\tilde{G}}^{-1}_{\alpha}$ is the empirical α-quantile of the simulated data, $\{\tilde{r}_t\}$. Generally, the amount of simulated data should be sufficiently large that the empirical quantile is an accurate estimate of the quantile of the unconditional distribution. There are two advantages to parametric Monte Carlo over other unconditional VaR estimators. First, this procedure exploits *conditioning* information that is ignored by the other estimators. Second, parsimonious conditional models, e.g., ARCH models with leverage, can generate rich families of unconditional distributions that are difficult to parameterize directly. The drawback of this procedure is that an incorrect conditional specification leads to an inconsistent estimate of the unconditional VaR.

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

Using the S&P 500 data, 3 unconditional parametric models, a normal, a Student's t and a skewed t were estimated. Estimates of the unconditional VaR using the Cornish-Fisher estimator and a

Unconditional Value-at-Risk

	HS	Normal	Stud. t	Skew t	CF
1% VaR	3.313	2.767	3.480	3.684	5.165
5% VaR	1.854	1.949	1.702	1.788	1.754
10% VaR	1.296	1.514	1.150	1.201	0.781

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 estimates of the 1% VaR differ.

Historical Simulation (nonparametric) estimator were also included. Estimates are reported in Table 8.2. The unconditional VaR estimates are similar except for the estimate computed using the Cornish-Fisher expansion. The kurtosis of the data was very high (23) which resulted in a very large 1% quantile. The others are broadly similar with the most substantial differences occurring at the 1% VaR. Figure 8.4 shows the estimated unconditional distribution from the normal and skewed t distributions and a nonparametric kernel density estimator. The key quantiles are similar despite meaningful differences in their shapes.

8.5 Evaluating VaR models

The process of evaluating the performance of VaR models is virtually identical to that of evaluating the specification of models of the conditional mean or variance. The key insight into VaR model evaluation comes from the tick loss function,

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

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

$$ge_t = I_{[r_t<F_{\alpha,t}^{-1}]} - \alpha \tag{8.20}$$

which is the time-t "HIT" (HIT_t).[7] When there is a VaR exceedance, $HIT_t = 1-\alpha$ and when there is no exceedance, $HIT_t = -\alpha$. If the model is correct, then α of the HITs should be $(1-\alpha)$ and $(1-\alpha)$ should be $-\alpha$, so that

[7]The 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 $L\left(y_{t+1}, \hat{y}_{t+1|t}\right)$, then the generalized error is the derivative of the loss function with respect to the second argument, that is

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

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

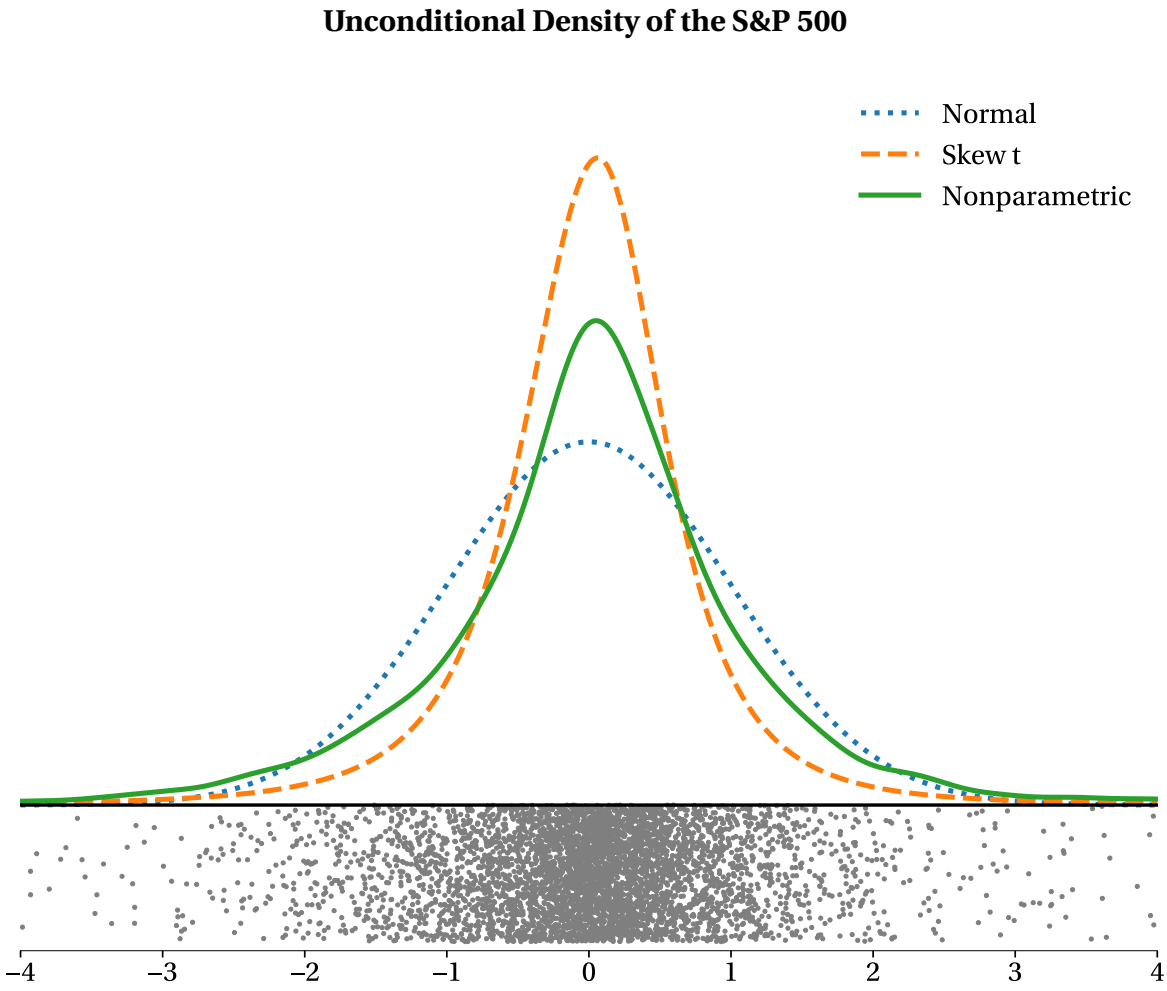


Figure 8.4: Plot of the S&P 500 returns as well as a parametric density using Hansen's skewed t and a nonparametric density estimator constructed using a kernel.

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

and the mean of HIT_t should be 0. Moreover, when the VaR is conditional on time t information , $\mathrm{E}_t[HIT_{t+1}] = 0$ which follows from the properties of optimal forecasts (see chapter 4).
A test that the conditional expectation is zero can be implemented using a generalized Mincer-Zarnowitz (GMZ) regression of $HIT_{t+1|t}$ on any time t available variable. For example, the estimated quantile $F^{-1}_{t+1|t}$ for $t+1$ could be included (since it is in the time-t information set) as well as lagged HITs to construct the regression model,

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

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. If the null is rejected, then either

the average number of violations is wrong, so that $\gamma_0 \neq 0$, or the VaR violations are predictable ($\gamma_j \neq 0$ for $j \geq 1$).

8.5.1 Likelihood Evaluation

VaR forecast evaluation can be improved by noting that VaR violations, $I_{[r_t<F_{\alpha,t}^{-1}]}$, are Bernoulli random variables which takes the value 1 with probability α and takes the value 0 with probability $1-\alpha$. A more powerful test can be constructed using a likelihood ratio test using the Bernoulli random variables $\widetilde{HIT}_t = I_{[r_t<F_{\alpha,t}^{-1}]}$. 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

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

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 a test of conditionally correct VaR by examining the dependence of *HIT*s. This testing strategy uses the properties of a Markov chain of Bernoulli random variables. A Markov chain is a modeling device which an ARMA model that models 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. so that $p_{00} = 1-\alpha$ and $p_{11} = \alpha$ in a correctly specified model.

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 consecutive *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 null hypothesis can be tested using the likelihood ratio

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

This test has an asymptotic χ^2_2 distribution since there are two restrictions under the null.

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. Both of these models are known as limited dependent variable models since the left-hand-side variables are always 0 or 1. For example, a specification test could be constructed using K lags of *HIT*, a constant and the forecast quantile as

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

Parameters are computed by maximizing the Bernoulli log-likelihood, which requires the estimated probabilities to satisfy

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

This restriction is imposed using one of two transformations, the normal CDF ($\Phi(z)$) which produces the probit model or the logistic function ($e^z/(1+e^z)$) which produces the logit model. Generally the choice between these two makes little difference. If $\mathbf{x}_t = [1 \;\; F_{t+1|t} \;\; \widetilde{HIT}_t \;\; \widetilde{HIT}_{t-1} \ldots \;\; \widetilde{HIT}_{t-K+1}]$, 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

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

The likelihood ratio for testing the null $H_0 : \gamma_0 = \Phi^{-1}(\alpha), \gamma_j = 0$ for all $j = 1, 2, \ldots, K$ against an alternative $H_1 = \gamma_0 \neq \Phi^{-1}(\alpha)$ or $\gamma_j \neq 0$ for some $j = 1, 2, \ldots, K$ can be computed

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

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).

8.5.2 Relative Comparisons

Diebold-Mariano tests can be used to rank the relative performance of VaR forecasting models (Diebold and Mariano, 1995). DM tests of VaR models are virtually identical to DM tests on the forecasts from two conditional mean or conditional variance models. The only important difference is the use of the VaR-specific tick loss function. 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

$$DM = \frac{\overline{d}}{\sqrt{\widehat{\mathrm{V}[d]}}} \tag{8.26}$$

where

$$d_t = L(r_{t+1}, \mathrm{VaR}^A_{t+1|t}) - L(r_{t+1}, \mathrm{VaR}^B_{t+1|t}),$$

VaR^A and VaR^B are the Value-at-Risks from models A and B respectively, $\overline{d} = R^{-1}\sum_{t=M+1}^{M+R} d_t$, 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 $\sqrt{\widehat{\mathrm{V}[d]}}$ is the long-run variance of d_t which requires the use of a HAC covariance estimator (e.g., Newey-West). Recall that DM is asymptotically normally distributed. The null is equal accuracy, $H_0 : \mathrm{E}[d_t] = 0$, and the composite alternative is $H_1^A : \mathrm{E}[d_t] < 0$ and $H_1^B : \mathrm{E}[d_t] > 0$. Large negative values (less than -2) indicate model A is superior while large positive values indicate the opposite; values close to zero indicate neither forecasting model outperforms the other.

Ideally the loss function, $L(\cdot)$, should reflect the user's preference over VaR forecast errors. In some circumstances there is no obvious choice, and the tick loss function,

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

is a theoretically sound choice. When the distribution of returns is continuous, the tick-loss is uniquely minimized at the conditional quantile. The tick-loss function has the same interpretation in a VaR model as the mean square error (MSE) does in conditional mean model evaluation or the QLIK loss function in volatility models evaluation.

8.6 Expected Shortfall

Expected shortfall – also known as tail VaR – combines aspects of VaR with additional 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

$$\begin{aligned} \text{ES} &= \text{E}\left[\frac{W_{t+1}-W_t}{W_t}\middle|\frac{W_{t+1}-W_t}{W_t}<-\text{VaR}\right] \\ &= \text{E}\left[r_{t+1}|r_{t+1}<-\text{VaR}\right] \end{aligned} \tag{8.28}$$

where W_t, is the value of the assets in the portfolio.[9]

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

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

$$\text{ES}_{t+1} = \text{E}_t\left[r_{t+1}|r_{t+1}<-\text{VaR}_{t+1}\right]. \tag{8.29}$$

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.)

Because the computation of Expected Shortfall requires both a quantile and an expectation, they are generally computed from density models, either parametric or semiparametric, rather than models focused on only the ES.

8.6.1 Evaluating Expected Shortfall models

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

$$\text{E}_t[\text{ES}_{t+1}] = \text{E}_t[r_{t+1}|r_{t+1}<-\text{VaR}_{t+1}].$$

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

[9]Just 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 here using the return.

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

$$(\text{ES}_{t+1|t} - r_{t+1})I_{[r_{t+1}<-\text{VaR}_{t+1|t}]} = \mathbf{x}_t\boldsymbol{\gamma} \tag{8.30}$$

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 regressors that capture important characteristics of the tail, such as recent volatility or the VaR forecast (VaR_{t+1}), 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 – the 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 due to the scarcity of data available to evaluate 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 misspecified Expected Shortfall models.

8.7 Density Forecasting

Value-at-Risk, a quantile, provides a narrow view into the riskiness of an asset. More importantly, VaR may not adequately describe the types of risk relevant to a forecast consumer. A density forecast, in contrast, summarizes *everything* there is to know about the riskiness of the asset. Density forecasts nest both VaR and Expected Shortfall as special cases.

In light of this relationship, it is not apparent that VaR or Expected Shortfall should be used. Density forecasting suffers from three distinct challenges:

- 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 density of future asset prices (e.g., pricing an option).

- Multi-step density forecasts are rarely analytically tractable since densities do not time aggregate, except in special cases that are too simple for most applications.

- Unless the user has preferences over the entire distribution, density forecasts inefficiently utilize information.

8.7.1 Density Forecasts from ARCH models

Density forecasting from ARCH models is identical to VaR forecasting from ARCH models. For simplicity, a model with a constant mean and GARCH(1,1) variances is used,

$$
\begin{aligned}
r_{t+1} &= \mu + \epsilon_{t+1} \\
\sigma^2_{t+1} &= \omega + \gamma_1 \epsilon^2_t + \beta_1 \sigma^2_t \\
\epsilon_{t+1} &= \sigma_{t+1} e_{t+1} \\
e_{t+1} &\overset{\text{i.i.d.}}{\sim} G(0,1).
\end{aligned}
$$

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. In practice, the mean and variance can be modeled using richer parameterizations that have been tailored so the historical data. Standard choices for $G(\cdot)$ include the standardized Student's t, the generalized error distribution, and Hansen's skewed t. The 1-step ahead density forecast is then

$$
\hat{F}_{t+1|t} \overset{d}{=} G(\hat{\mu}, \hat{\sigma}^2_{t+1|t}) \tag{8.31}
$$

where $F(\cdot)$ is the distribution of returns. This follow directly from the original model where $r_{t+1} = \mu + \sigma_{t+1} e_{t+1}$ and $e_{t+1} \overset{\text{i.i.d.}}{\sim} G(0,1)$.

8.7.2 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 $G(\cdot)$,

$$
\begin{aligned}
r_{t+1} &= \mu + \epsilon_{t+1} \\
\sigma^2_{t+1} &= \omega + \gamma_1 \epsilon^2_t + \beta_1 \sigma^2_t \\
\epsilon_{t+1} &= \sigma_{t+1} e_{t+1} \\
e_{t+1} &\overset{\text{i.i.d.}}{\sim} G(0,1).
\end{aligned}
$$

and estimates of $\hat{\sigma}^2_t$ 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 $\hat{e}_t = {}^{\hat{\epsilon}_t}/_{\hat{\sigma}_t}$. The final step is to compute the distribution. The simplest method to accomplish this is to compute the empirical CDF as

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

The function returns the percentage of the standardized residuals smaller than the value e. This method is trivial but has some limitations. First, the PDF does not exist since $G(\cdot)$ is not differentiable. This property makes some applications difficult, although a histogram provides a simple, but 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, more accurate 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 standardized residuals 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

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

where $K(\cdot)$ can be one of many kernels – the choice of which usually makes little difference. The two most common are the Gaussian,

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

and the Epanechnikov,

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

The choice of the bandwidth h determines the width of the window used to smooth the density. It plays a more substantial role than the choice of the kernel in the accuracy of a density estimate. In practice, Silverman's bandwidth,

$$h = 1.06\sigma T^{-\frac{1}{5}}, \tag{8.36}$$

is widely used where σ is the standard deviation of $\hat{e}_t$ (which is theoretically 1, but may differ if the model is misspecified). However, larger or smaller bandwidths can be used to produce smoother or rougher densities, respectively. The magnitude of the bandwidth represents a bias-variance tradeoff – a small bandwidth has little bias 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 is constructed by scaling the distribution G by $\sigma_{t+1|t}$ and adding the mean. The top panel of Figure 8.5 contains a plot of the empirical CDF and kernel smoothed CDF of TARCH(1,1,1)-standardized S&P 500 returns in 2018. The empirical CDF is jagged, and there are some large gaps in the observed returns. The bottom panel shows the histogram of the standardized returns where each bin contains 10 returns, and the smoothed kernel density estimate computed using Silverman's bandwidth and a Gaussian kernel.

8.7.3 Multi-step density forecasting and the fan plot

Multi-step ahead density forecasts do not time aggregate. For example, consider a simple GARCH(1,1) model with normal innovations,

$$\begin{aligned} r_{t+1} &= \mu + \epsilon_{t+1} \\ \sigma^2_{t+1} &= \omega + \gamma_1 \epsilon^2_t + \beta_1 \sigma^2_t \\ \epsilon_{t+1} &= \sigma_{t+1} e_{t+1} \\ e_{t+1} &\overset{\text{i.i.d.}}{\sim} N(0,1). \end{aligned}$$

The 1-step ahead density forecast of returns is

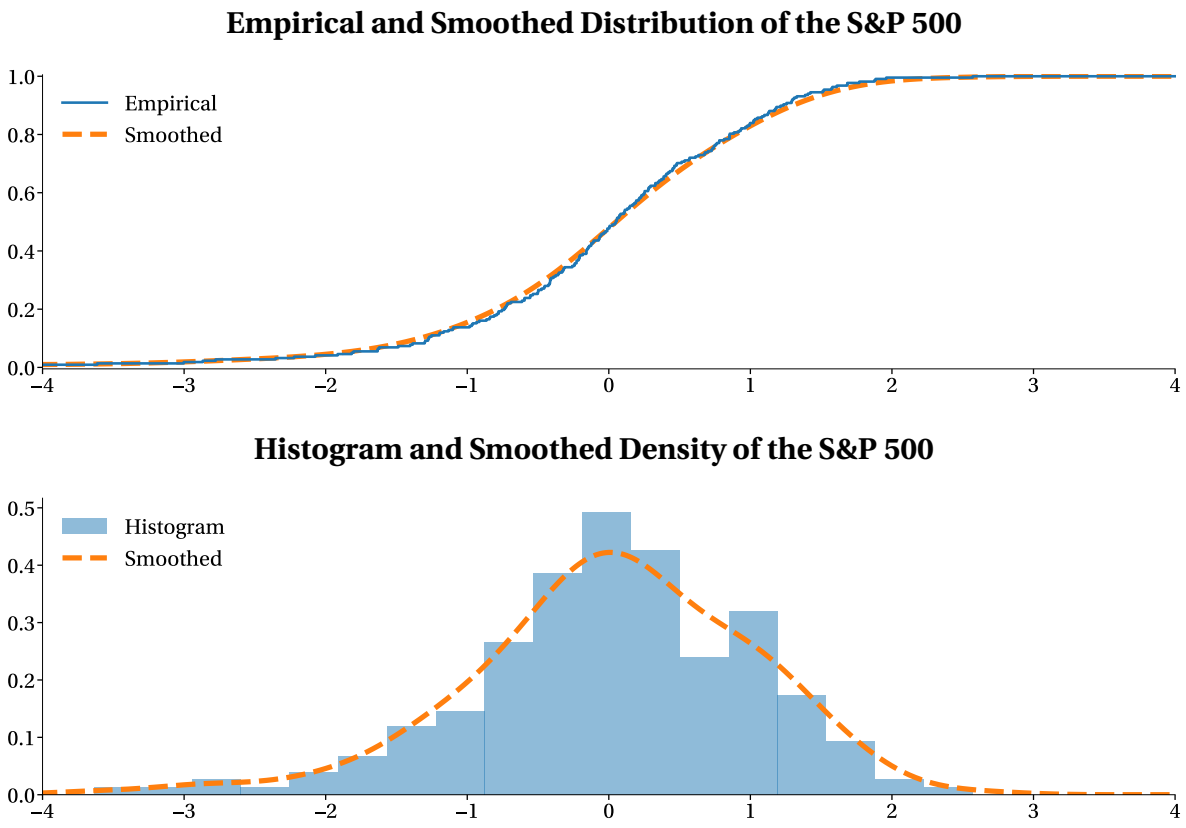


Figure 8.5: The top panel shows the rough empirical and smoothed empirical CDF for standardized returns of the S&P 500 in 2018 (standardized by a TARCH(1,1,1)). The bottom panel shows the histogram of the standardized returns using bins with 10 observations each and the smoothed kernel density.

$$r_{t+1}|\mathcal{F}_t \sim N(\mu, \sigma^2_{t+1|t}). \tag{8.37}$$

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

$$r_{t+2}|\mathcal{F}_t \sim N(\mu, \sigma^2_{t+2|t}). \tag{8.38}$$

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 $\sigma^2_{t+2|t}$, unlike $\sigma^2_{t+1|t}$, is a random variable and the uncertainty in $\sigma^2_{t+2|t}$ must be integrated out to determine the distribution of r_{t+2}. The correct form of the 2-step ahead density forecast is

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

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

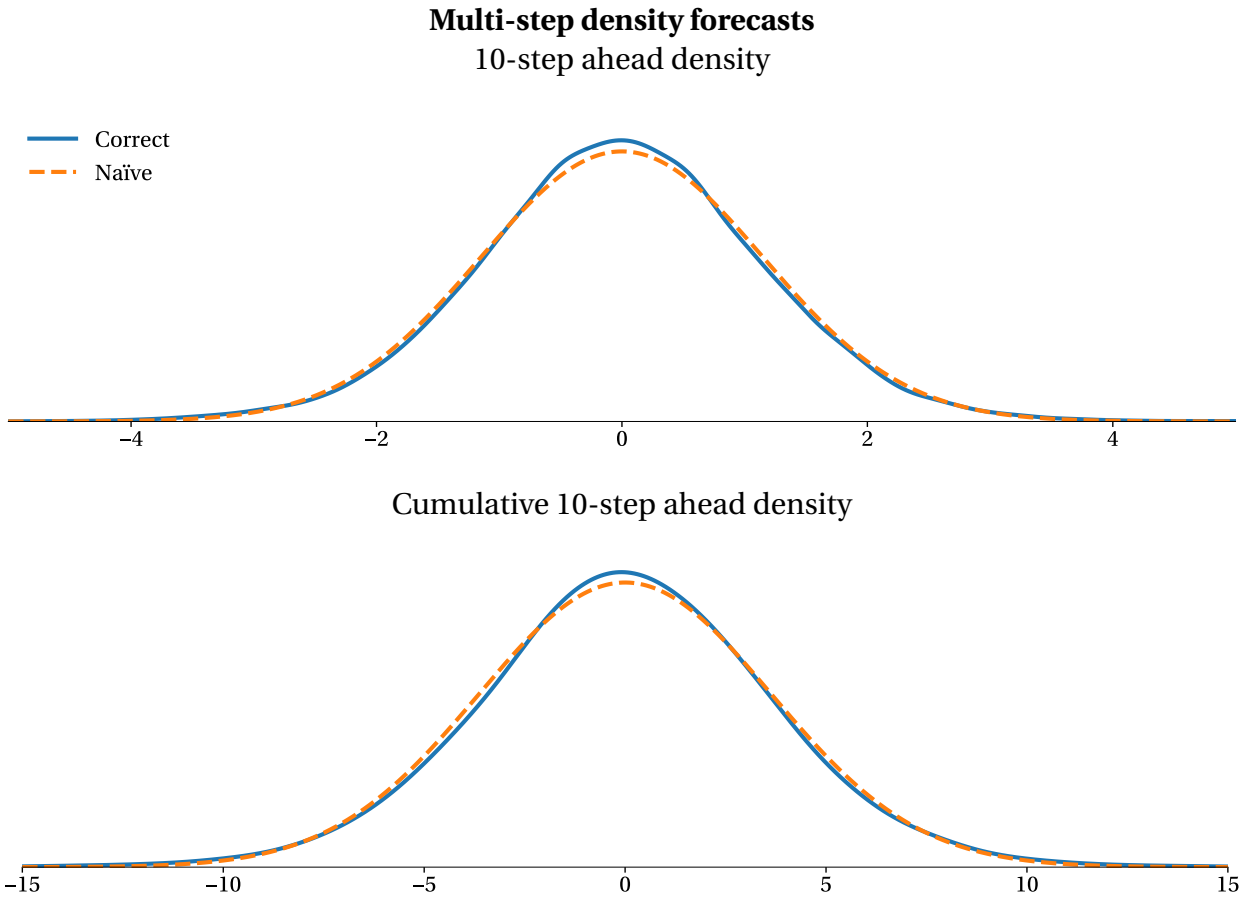


Figure 8.6: Naïve and correct 10-step ahead density forecasts from a simulated GARCH(1,1) model. The correct density forecasts have substantially fatter tails then the naïve forecast as evidenced by the central peak and cross-over of the density in the tails.

---

a mixture of normal random variables where the probability of getting a specific normal depends on $w(e) = \phi(e_{t+1})$,

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

Unless $w(e)$ is constant, the resulting distribution is not a normal. The top panel in Figure 8.6 contains the naïve 10-step ahead forecast and the correct 10-step ahead forecast for a simple GARCH(1,1) process,

$$\begin{aligned} r_{t+1} &= \epsilon_{t+1} \\ \sigma^2_{t+1} &= .02 + .2\epsilon^2_t + .78\sigma^2_t \\ \epsilon_{t+1} &= \sigma_{t+1} e_{t+1} \\ e_{t+1} &\overset{\text{i.i.d.}}{\sim} N(0,1) \end{aligned}$$

where $\epsilon_t^2 = \sigma_t = \bar{\sigma} = 1$ and hence $\mathrm{E}_t[\sigma_{t+h}] = 1$ 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

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

for $h = 1, 2, \ldots, 10$. 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 in the extremes.

Fan Plots

A fan plot is a graphical tool to convey information about future *changes* in uncertainty. The Bank of England has popularized these representations as a method to convey the uncertainty about the future of the economy. Figure 8.7 contains a fan plot of the forecast density for a persistent AR(2) with i.i.d. standard normal increments.[10] Darker regions indicate higher probability while progressively lighter regions indicate less likely events.

8.7.4 Quantile-Quantile (QQ) plots

A Quantile-Quantile, or QQ, plot is a graphical tool that is used 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 ordering the standardized residuals,

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

and then plotting the ordered residual $\hat{e}_j$ (x-axis) against its hypothetical value (y-axis) 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 graphical assessment of a distribution is formalized in the Kolmogorov-Smirnov test. Figure 8.8 contains 4 QQ plots for monthly S&P 500 returns against a normal, a Student's t, a skewed t, and a GED. The MLE of the density parameters were used to produce the QQ plots. The normal appears to be badly misspecified in the tails – as evidenced through deviations from the 45^o line. The other models appear adequate for the monthly returns. The skewed t performs especially well in the lower tail.

8.7.5 Evaluating Density Forecasts

All density evaluation strategies are derived from a basic property of continuous random variables: if $x \sim F$, then $u \equiv F(x) \sim U(0,1)$. That is, for any continuous random variable X, the cumulant of X has a Uniform distribution over $[0,1]$. The opposite of this results is also true, if $U \sim \text{Uniform}(0,1)$, $F^{-1}(U) = X \sim F$.[11]

[10] The density was generated from the AR(2) $Y_t = 1.8Y_{t-1} - 0.81Y_{t-2} + \epsilon_t$ where the final two in-sample values are $y_T = 4.83$ and $y_{T-1} = 1.37$.

[11] The latter result can be used as the basis of a random number generator. To generate a random number with a CDF of F, first generate a uniform value, u, and then compute the inverse CDF of 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.

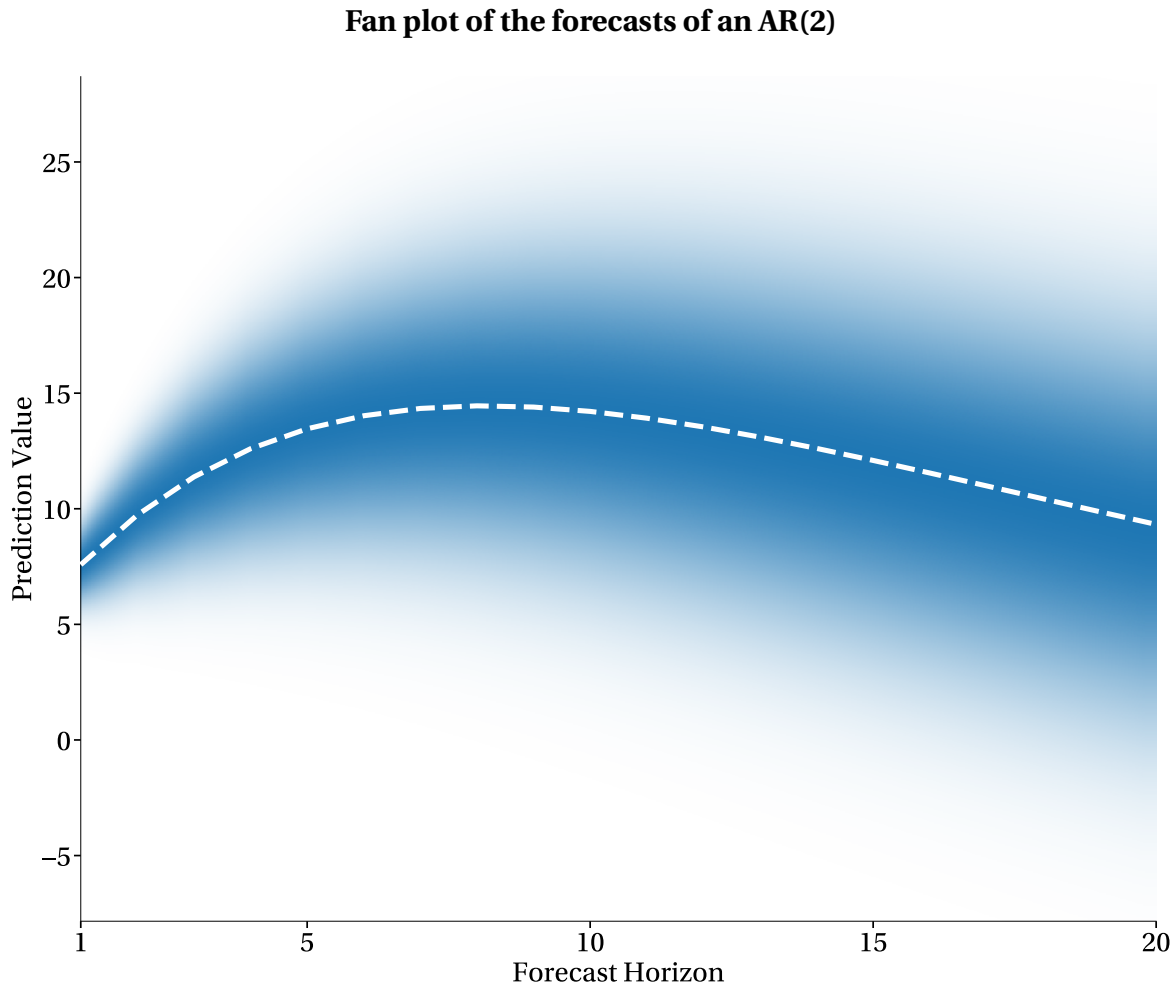


Figure 8.7: Future density of a persistent AR(2) with i.i.d. standard normal increments. Darker regions indicate higher probability while progressively lighter regions indicate less likely events.

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* $\Pr(Y \leq y) = y$, $0 < y < 1$.

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

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

□

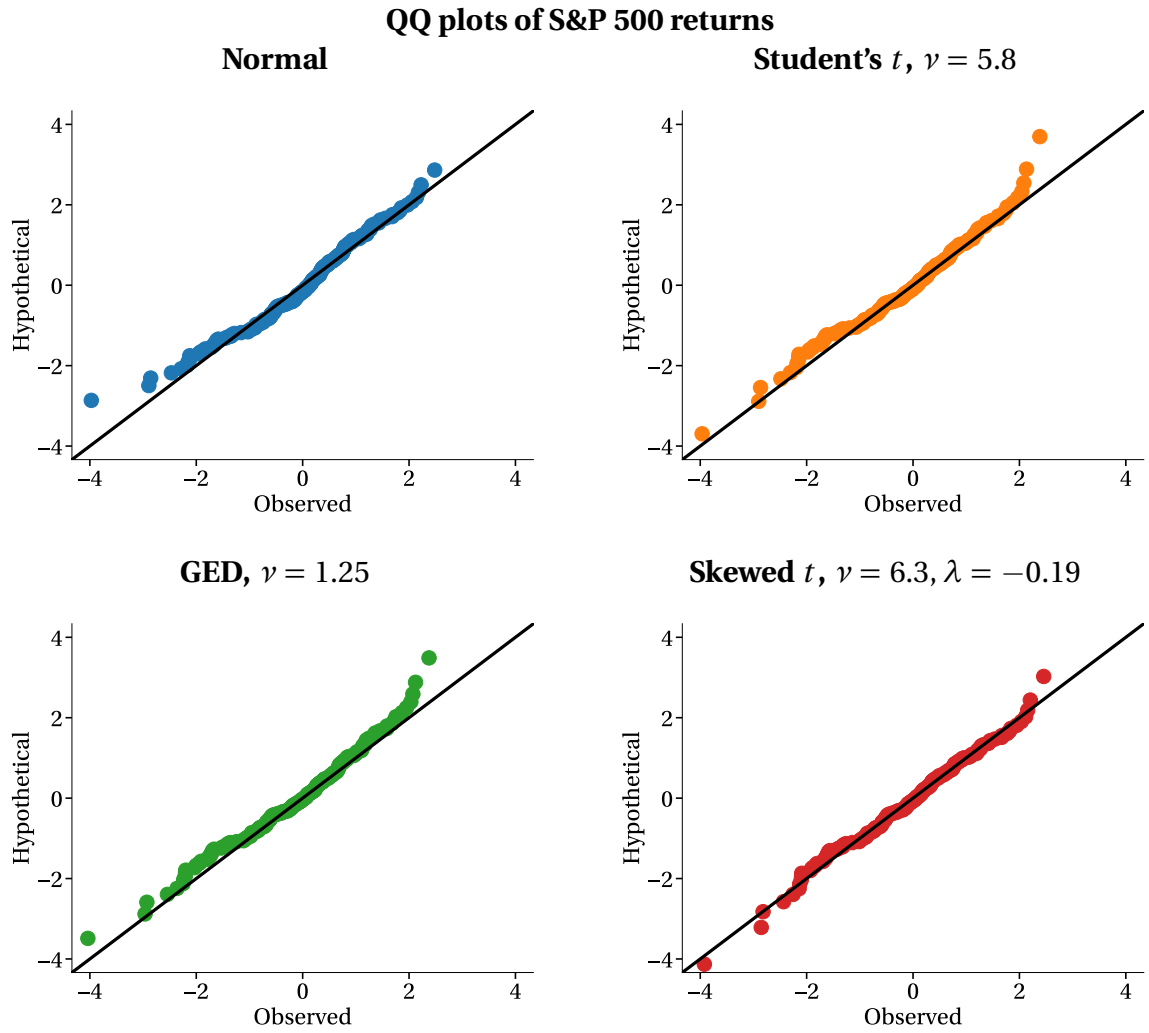


Figure 8.8: QQ plots of the studentized S&P 500 returns against fitted Normal, Student's t, GED and Skewed t distributions. Points along the 45^o indicate a good distributional fit.

The proof shows that $\Pr(F_X(X) \leq y) = y$ and so $Y = F_X(X)$ must be a uniform random variable (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* $\hat{u}_t = F(\hat{e}_t)$ from the standardized residuals and then sorting them

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

The KS test statistic is then computed from

$$KS = \max_{\tau} \left| \sum_{i=1}^{\tau} I_{[u_j < \frac{\tau}{T}]} - \frac{1}{T} \right| \tag{8.39}$$

$$= \max_{\tau} \left| \left(\sum_{i=1}^{\tau} I_{[u_j < \frac{\tau}{T}]} \right) - \frac{\tau}{T} \right|$$

The test statistic finds the maximum deviation between the number of u_j less than $\frac{\tau}{T}$ and the expected number of observations which should be less than $\frac{\tau}{T}$. Since the probability integral transformed residuals should be Uniform(0,1) when the model is correctly specified, 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 and simulated critical values are available in most software packages.

The KS test has a graphical interpretation as a QQ plot of the probability integral transformed residuals against a uniform. Figure 8.9 contains a representation of the KS test using data from two series: the first is standard normal and the second is a standardized Student's t_3. 95% confidence bands are denoted with dotted lines. The data from both series were assumed to be from a standard normal. The KS test rejects normality of the t_3 data as evidenced by the cumulants just crossing the confidence band.

Parameter Estimation Error and the KS Test

The critical values supplied by most packages *do not* account for parameter estimation error. KS tests on in-sample data from models with estimated parameters are *less* likely to reject than if the true 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, the desired size of 10% (corresponding to a 90% critical value) has an actual size closer to 0.1%. Using the wrong critical value distorts the size of the test and lowers the test's power – the test statistic is unlikely to reject the null hypothesis 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, to estimate the parameters, and to compute the KS test of the simulated standardized residuals (where the residuals are using estimated parameters). These steps can be repeated B times ($B > 1000$, possibly larger) and then the correct critical values can be computed from the empirical 90, 95 or 99% quantiles from KS_b, $b = 1, 2, \ldots, B$. These quantiles are the correct values to use under the null while accounting for parameter estimation uncertainty.

8.7.6 Evaluating conditional density forecasts

In a direct analogue to the unconditional case, if $X_{t+1}|\mathcal{F}_t \sim F$, then $\hat{u}_{t+1} \equiv F(\hat{x}_{t+1})|\mathcal{F}_t \overset{\text{i.i.d.}}{\sim} U(0,1)$. That is, the probability integral transformed residuals are *conditionally* i.i.d. Uniform (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 further transforming the probability integral transformed residuals into normals using the inverse Normal CDF . Specifically if $\hat{u}_{t+1} = F_{t+1|t}(\hat{e}_{t+1})$ are the residuals standardized by their forecast distributions, the Berkowitz test computes $\hat{y}_{t+1} = \Phi^{-1}(\hat{u}_{t+1}) = \Phi^{-1}(F_{t+1|t}(\hat{e}_{t+1}))$ which have the property, under the null of a correct specification that $\hat{y}_t \overset{\text{i.i.d.}}{\sim} N(0,1)$ – an i.i.d. sequence of standard normal random variables.

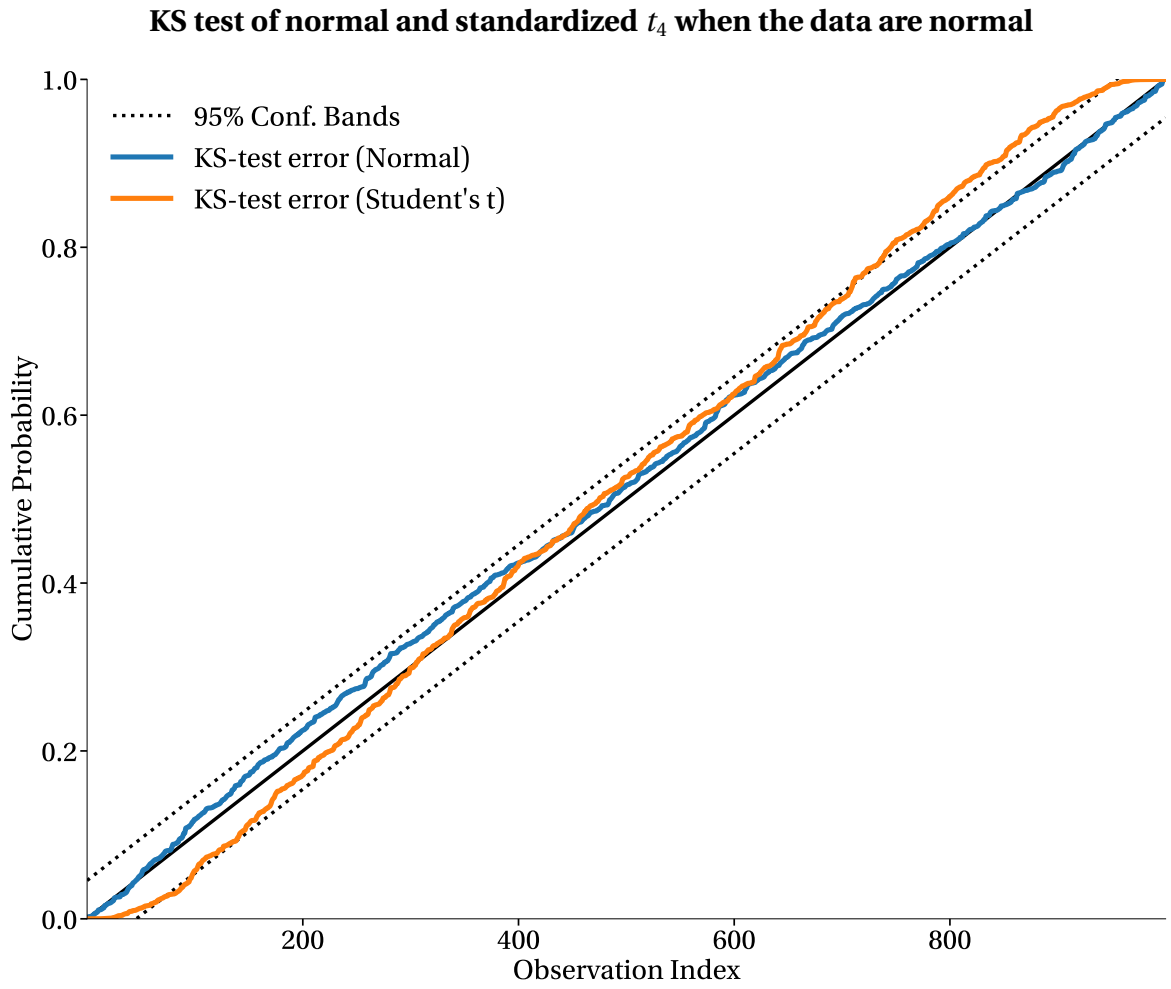


Figure 8.9: A KS test with simulated normal and t_3 data. In both cases, the null is that the data have normal distributions. The data generated from the t_3 crosses the confidence boundary indicating a rejection of this specification. An accurate density forecast should produce a cumulative distribution close to the 45^o line.

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 the likelihood ratio test

$$LR = 2(l(\hat{\boldsymbol{\theta}}; \mathbf{y}) - l(\boldsymbol{\theta}_0; \mathbf{y})) \sim \chi^2_3 \tag{8.40}$$

where $\boldsymbol{\theta}_0$ are the parameters if the null is true, $\phi_0 = \phi_1 = 0$ 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 $H_0 : \boldsymbol{\gamma} = 0, \sigma^2 = 1$ and the alternative is the unrestricted estimate from the alternative specification. The likelihood ratio test statistic in the general case would have a χ^2_{K+1} distribution where K is the number of elements in $\mathbf{x}_t$ (the +1 comes from the restriction that $\sigma^2 = 1$).

8.8 Coherent Risk Measures

With multiple measures of risk available, which should be chosen: variance, VaR, or Expected Shortfall? Recent research into risk measurement has identified four desirable properties of any risk measure. Let ρ be any measure of risk, e.g., VaR or ES, that maps the riskiness of a portfolio to the reserves required to cover regularly occurring losses. P, P_1 and P_2 are portfolios of assets.

Drift Invariance

The requires reserved for portfolio P satisfies

$$\rho\left(P + C\right) = \rho\left(P\right) - c$$

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

Homogeneity

The required reserved are linear homogeneous,

$$\rho(\lambda P) = \lambda \rho(P) \quad \text{for any } \lambda > 0. \tag{8.41}$$

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 (P_1 FOSD P_2), the required reserves for P_1 must be less than those of P_2since

$$\rho(P_1) \leq \rho(P_2). \tag{8.42}$$

If P_1 FOSD P_2 then the value of portfolio P_1 is 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 than the required reserves for each treated separately

$$\rho(P_1 + P_2) \leq \rho(P_1) + \rho(P_2). \tag{8.43}$$

Definition 8.6 (Coherent Risk Measure)**.** Any risk measure which satisfies these four properties is *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 always has a higher return, and so 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 a portfolio for non-linear exposures. The simplest example where subadditivity fails is in portfolios of defaults bonds. Suppose P_1 and P_2 are portfolios where each contains a single bond with a face value of \$1,000 paying 0% interest. The bonds in the portfolios are from two companies. Assume that the default of one company is independent of the default of the other and that each company defaults with probability 3%. If a company defaults, only 60% of the bond value is recovered. The 5% VaR of both P_1 and P_2 is 0 since the companies pays the full \$1,000 97% of the time. The VaR of $P_3 = 50\% \times P_1 + 50\% \times P_2$, however, is \$200 since at least one company defaults 5.91% of the time. The distribution of P_3 is:

Probability	Portfolio Value
0.09%	\$600
5.82%	\$800
94.09%	\$1,000

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 proof that Expected Shortfall is coherent is straight forward in specific models (for example, if the returns are jointly normally distributed). The proof for an arbitrary distribution is challenging and provides little intuition. However, coherency alone does not make Expected Shortfall a better choice than VaR for measuring portfolio risk. VaR has many advantages as a risk measure: it only requires the modeling of a quantile of the return distribution, VaR always exists and is finite, and there are many well-established methodologies for accurately estimating VaR. Expected Shortfall requires an estimate of the mean in the tail which is more difficult to estimate accurately than the VaR. The ES may not exist in some cases if the distribution is very heavy-tailed. Additionally, in most realistic cases, increases in the Expected Shortfall is accompanied with increases in the VaR, and so these two measures often agree about the risk in a 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 Historical Simulation?

Problem 8.3. Define Expected Shortfall. How does this extend the idea of Value-at-Risk? Why is it preferred to 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.

Problem 8.6. How are Value-at-Risk forecasts assessed? Describe two methods that can be sued to detect flawed Value-at-Risk models.

Longer Exercises

Exercise 8.1. Precisely answer the following questions.

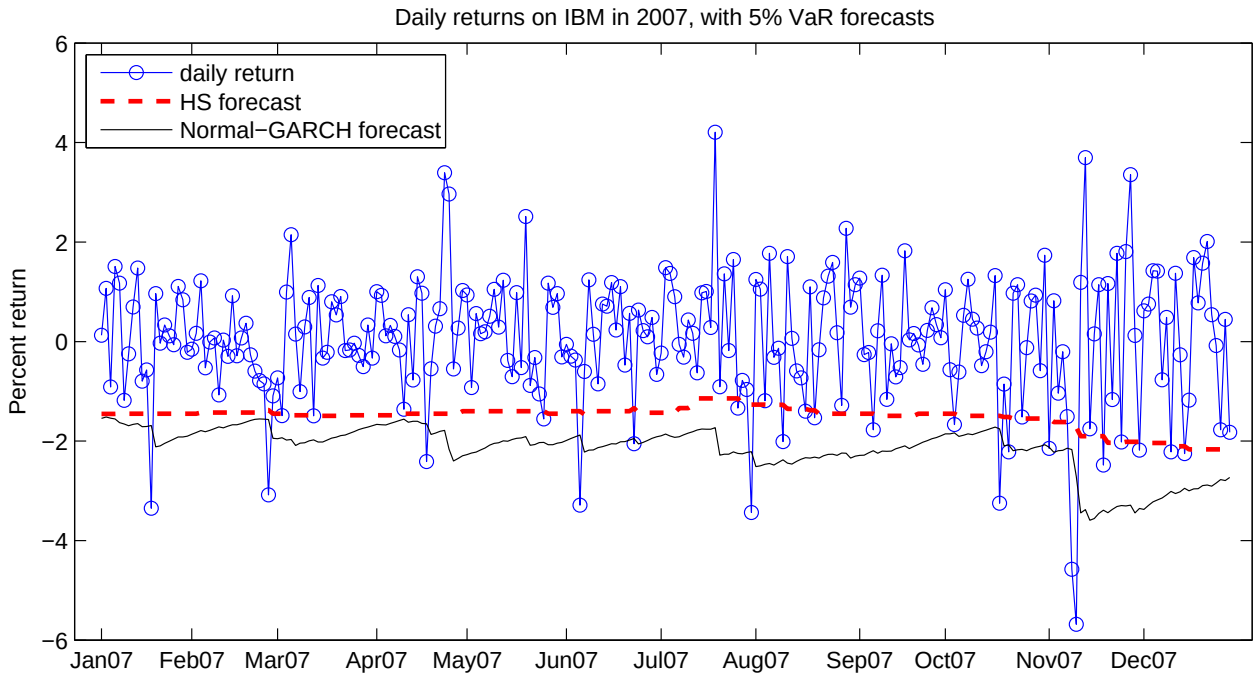
1. What is VaR?

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).



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):

$$\begin{aligned} HIT_t^{HS} &= \underset{(0.0186)}{0.0956} + u_t \\ HIT_t^{GARCH} &= \underset{(0.0129)}{0.0438} + u_t \end{aligned}$$

   (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):

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

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 8.10 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):

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

(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):

$$\begin{aligned} HIT_t^{HS} &= \underset{(0.0136)}{0.0462} + \underset{(0.1176)}{0.1845} HIT_{t-1}^{HS} + u_t \\ HIT_t^{GARCH} &= \underset{(0.0106)}{0.0285} - \underset{(0.0201)}{0.0233} HIT_{t-1}^{GARCH} + u_t \end{aligned}$$

A joint test that the intercept is 0.05 and the slope coefficient is zero yielded a chi-squared statistic of 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 differences between what you found when testing using only a constant and when using a constant and the lagged HIT.

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

2. Assume that $V \sim Uniform(0,1)$, 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 $\beta \in (0,\infty)$:

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

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

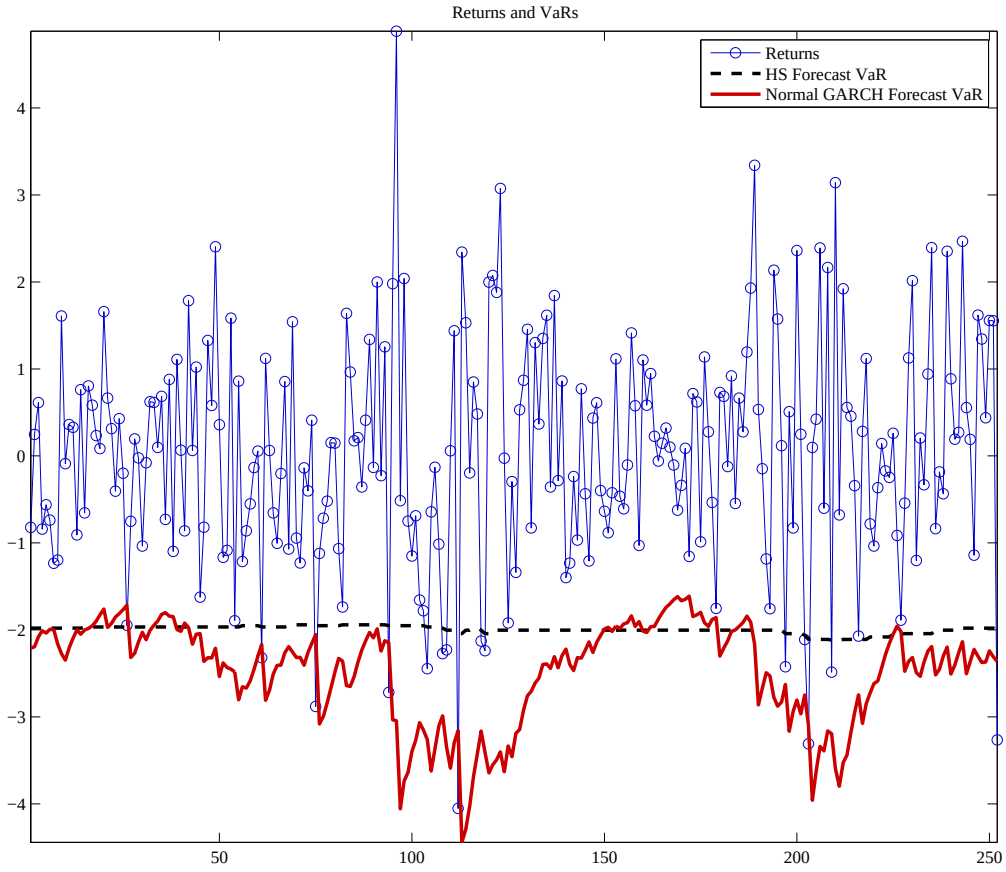


Figure 8.10: Returns, Historical Simulation VaR and Normal GARCH VaR.

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 $n \to \infty$. Show that $\Pr[U \leq u] \xrightarrow{p} u$ as $n \to \infty$, 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

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

where $\hat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^{T}\hat{\eta}_t^2$, $\hat{\mathbf{\Sigma}}_{XX} = \frac{1}{T}\mathbf{X}'\mathbf{X}$ 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 misspecified, 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_v^2 are

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

Exercise 8.6. Suppose you have a sample of 500 observations to evaluate a Value-at-Risk model using Out-of-Sample forecasts. You observe 36 95% VaR violations in this period.

1. What features the VaR violations of a correctly specified VaR model have?

2. Perform a test that the model is well specified using the sample average.

3. The likelihood of a Bernoulli(p) random variable is

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

How can you use this likelihood to implement a better test? Compute the test statistic and draw conclusions about the accuracy of the model.

4. Explain the differences between these two approaches.

5. Fully describe one method that would allow you to use the time series of VaR violations to test whether the model has correctly specified dynamics.