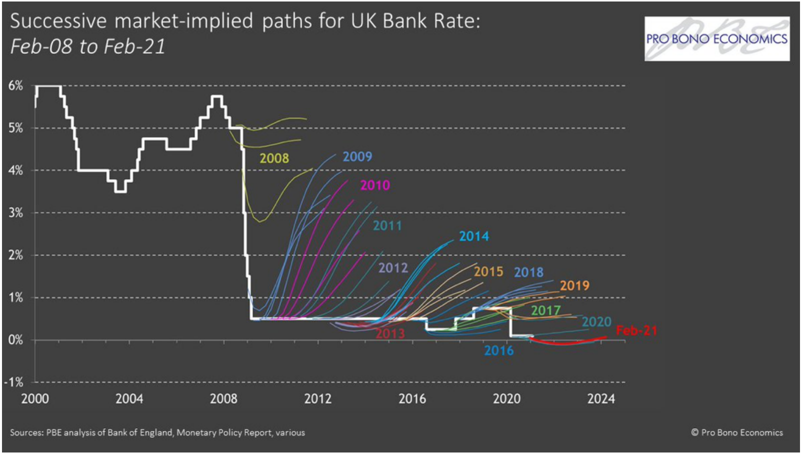
Forecasting Economic and Financial Time Series
Week 8:Volatility, interval and density forecasts

Ron Smith
MSc/PGCE Option, EMS Birkbeck.

Spring 2021

Beware Hedgehogs



Beyond point forecasts

- The need to provide more than point forecasts follows from not knowing users' loss functions.
- If they care about the uncertainty they may want a **volatility forecast,** needed to get optimal forecast with linex loss.
- Using an estimate of the variance (and perhaps higher moments) and a given distribution we can give a parametric **interval forecast**, e.g. a 95% probability that it lies between $y^-_{t+h,t}$ and $y^+_{t+h,t}$
- We can give a **probability** forecast of a value below $y^-_{t+h,t}$
- We can give a **density forecast** of the probability distribution of future values, this may be
  - parametric based on an assumed distribution
  - bootstrapped, drawn from the distribution of past residuals
  - judgemental.

GARCH, volatility forecasts

- The generalized autoregressive conditional heteroskedasticity GARCH(1,1)

$$\begin{aligned} y_t &= \beta' x_t + \varepsilon_t; \quad E(\varepsilon_t^2 \mid I_{t-1}) = h_t^2 \\ h_t^2 &= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \phi_1 h_{t-1}^2 \end{aligned}$$

  $\alpha_0 > 0$, α_1 is the ARCH term, ϕ_1 is the GARCH term. It is stationary if $\mid \alpha_1 + \phi_1 \mid < 1$, the case where it equals 1 is Integrated GARCH IGARCH.

- GARCH is widely used for financial data where the normality assumptions is often not appropriate: tails are fatter than Gaussian, excess kurtosis. Alternative distributions are t and Generalised Error Distribution. The t distribution has moments of the order ofits degrees of freedom, ν, so it needs $\nu > 2$ for a variance to exist and if $\nu > 30$ it is indistinguishable from a normal.

EGARCH

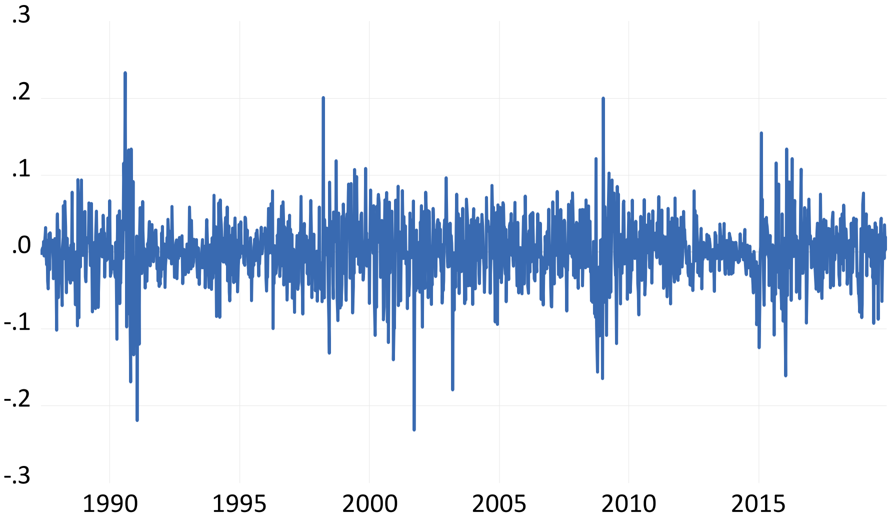
- There are lots of other GARCH variants, exponential GARCH, EGARCH(1,1,1), It is

$$\log h_t^2 = \alpha_0 + \alpha_1 \left(\frac{\varepsilon_{t-1}}{h_{t-1}} \right) + \alpha_2 \left(\left| \frac{\varepsilon_{t-1}}{h_{t-1}} \right| - \mu \right) + \phi_1 h_{t-1}^2$$

- It allows for leverage effects, the sign of the shock matters, negative shocks increase the conditional volatility more than positive shocks. Called leverage effects because a negative shock to returns reduces equity values and increases the leverage of the firm as measured by debt-equity ratios thereby making equity more risky.
- There are restrictions on the parameters of GARCH models to ensure the variance is positive, since EGARCH uses log variance this problem does not arise.
- Can have ARCH in mean effects, with $\widehat{\sigma}_{t,t-1}$ appearing in the mean equation

Brent

- Look at the change in log Brent oil price, weekly data until end November 2019
- Graph shows definite time varying volatility and fat tails
- The GARCH family of models allows more flexibility in forecasting through having
  - a mean equation, we assume that it is ARIMA(1,1,0) (try GARCH in mean using log variance, not significant)
  - a variance equation try GARCH and EGARCH
  - more flexible distributions, assume t.



.3
.2
.1
.0
-.1
-.2
-.3
1990
1995
2000
2005
2010
2015

Dependent Variable: DLOGB
Method: ML ARCH - Student's t distribution (BFGS / Marquardt steps)
Date: 02/14/20 Time: 18:05
Sample: 1/02/2015 11/29/2019
Included observations: 257
Convergence achieved after 34 iterations
Coefficient covariance computed using outer product of gradients
Presample variance: backcast (parameter = 0.7)
GARCH = C(3) + C(4)*RESID(-1)^2 + C(5)*GARCH(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.001780	0.002424	0.734019	0.4629
DLOGB(-1)	0.244871	0.063318	3.867301	0.0001
		Variance Equation		
C	0.000107	8.36E-05	1.280340	0.2004
RESID(-1)^2	0.084576	0.049614	1.704683	0.0883
GARCH(-1)	0.843414	0.081609	10.33477	0.0000
T-DIST. DOF	15.27849	17.39008	0.878575	0.3796

R-squared	0.060801	Mean dependent var	0.000382
Adjusted R-squared	0.057118	S.D. dependent var	0.042823
S.E. of regression	0.041582	Akaike info criterion	-3.579974
Sum squared resid	0.440906	Schwarz criterion	-3.497116
Log likelihood	466.0267	Hannan-Quinn criter.	-3.546653
Durbin-Watson stat	1.931564		

Dependent Variable: DLOGB
Method: ML ARCH - Student's t distribution (BFGS / Marquardt steps)
Date: 03/13/20 Time: 15:39
Sample (adjusted): 5/29/1987 11/29/2019
Included observations: 1697 after adjustments
Convergence achieved after 41 iterations
Coefficient covariance computed using outer product of gradients
Presample variance: backcast (parameter = 0.7)
GARCH = C(4) + C(5)*RESID(-1)^2 + C(6)*GARCH(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
LOG(GARCH)	0.002491	0.001401	1.778204	0.0754
C	0.018125	0.009700	1.868659	0.0617
DLOGB(-1)	0.228101	0.024606	9.270076	0.0000
		Variance Equation		
C	3.44E-05	1.13E-05	3.045565	0.0023
RESID(-1)^2	0.102704	0.017306	5.934658	0.0000
GARCH(-1)	0.880340	0.018258	48.21590	0.0000
T-DIST. DOF	8.951118	1.483036	6.035672	0.0000

R-squared	0.035069	Mean dependent var	0.000737
Adjusted R-squared	0.033930	S.D. dependent var	0.041733
S.E. of regression	0.041019	Akaike info criterion	-3.758088
Sum squared resid	2.850304	Schwarz criterion	-3.735662
Log likelihood	3195.738	Hannan-Quinn criter.	-3.749785
Durbin-Watson stat	2.059782		

Dependent Variable: DLOGB
Method: ML ARCH - Student's t distribution (BFGS / Marquardt steps)
Date: 02/14/20 Time: 18:06
Sample: 1/02/2015 11/29/2019
Included observations: 257
Failure to improve likelihood (non-zero gradients) after 79 iterations
Coefficient covariance computed using outer product of gradients
Presample variance: backcast (parameter = 0.7)
LOG(GARCH) = C(3) + C(4)*ABS(RESID(-1)/@SQRT(GARCH(-1))) + C(5)
*RESID(-1)/@SQRT(GARCH(-1)) + C(6)*LOG(GARCH(-1))

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-0.001187	0.001914	-0.620192	0.5351
DLOGB(-1)	0.253842	0.062474	4.063188	0.0000
		Variance Equation		
C(3)	-0.179582	0.036458	-4.925756	0.0000
C(4)	-0.122384	0.049574	-2.468699	0.0136
C(5)	-0.171096	0.032865	-5.206074	0.0000
C(6)	0.957304	2.12E-05	45248.31	0.0000
T-DIST. DOF	89.67955	785.5968	0.114155	0.9091

R-squared	0.060719	Mean dependent var	0.000382
Adjusted R-squared	0.057036	S.D. dependent var	0.042823
S.E. of regression	0.041584	Akaike info criterion	-3.684483
Sum squared resid	0.440944	Schwarz criterion	-3.587815
Log likelihood	480.4561	Hannan-Quinn criter.	-3.645608
Durbin-Watson stat	1.947267		

GARCH forecasts

- Estimated by a non-linear maximum likelihood procedure. May not converge. Always check convergence. Programs differ in how they handle non-convergence. EViews prints out the non-converged results that can be very misleading.
- You can use the mean equation to forecast the expected value in the usual way; then the variance equation to forecast the variance;

$$\begin{aligned} h_t^2 &= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \phi_1 h_{t-1}^2 \\ h_{T+1/T}^2 &= \widehat{\alpha}_0 + \widehat{\alpha}_1 \widehat{\varepsilon}_T^2 + \widehat{\phi}_1 \widehat{h}_T^2 \\ h_{T+2/T}^2 &= \widehat{\alpha}_0 + \widehat{\phi}_1 h_{T+1/T}^2 \end{aligned}$$

- Then using the variance and t distribution degrees of freedom you can get probabilities, intervals and densities from a t distribution.

Exponential (Holt Winters) Smoothing 1

- Before we did it for a variable to estimate a local mean, with variants to allow for trend and seasonality
- Initialise at time $t = 1$: $\bar{y}_1 = y_1$
- Update: $\bar{y}_t = \alpha y_t + (1 - \alpha)\bar{y}_{t-1}$ equivalent to updating the forecast proportional to the previous period forecast error $\bar{y}_t - \bar{y}_{t-1} = \alpha(y_t - \bar{y}_{t-1})$
- Forecast: $\bar{y}_{T+h,T} = \bar{y}_T$
- Note
$$\bar{y}_t = \sum_{j=0}^{t-1} \alpha(1-\alpha)^j y_{t-j}$$
- Choose $0 < \alpha < 1$ to reflect the relative importance of signal in y_t (large α) relative to noise (small α).

Exponential (Holt Winters) Smoothing 2

- We can do the same thing with variances rather than the variable, consider daily returns of stocks i, j, r_{it}, r_{jt} where they have means close to zero, then like Riskmetrics we can apply exponential smoothing

$$y^f_{t+1,t} = \alpha y_t + (1-\alpha) y^f_{t,t-1}$$

  to daily estimates
  - of variances
  $$y_t = r^2_{it}$$
  - and covariances
  $$y_t = r_{it} r_{jt}.$$
  - No problem of negative variances.

Parameteric interval and probability forecasts

- Given a known distribution, e.g. t, and estimated parameters, such as.(i) conditional mean (expectation), (ii) conditional variance, (iii) degrees of freedom,
  - We can construct confidence interval at a specified probability level.
  - We can also make probability forecasts, e.g. the probability that returns will be negative.
- These estimates are dependent on choosing the right distribution, particularly for tail events. With $z < -1.64$, p=5% with a normal distribution and p=10% with a t distribution with 3 degrees of freedom, similarly for $z < -2.35$, p=0.94% and p=5%.

Value at Risk VaR

- Value at Risk, VaR, for a given portfolio, time horizon, and probability p, is defined as
  - a threshold loss value, V, such that the probability that the loss on the portfolio over the given horizon exceeds V is p.
- A one day, 1%, VaR of £1million, means that there is a 1% chance that losses will exceed £1m in the next day. Losses should exceed £1m, one day in 100.

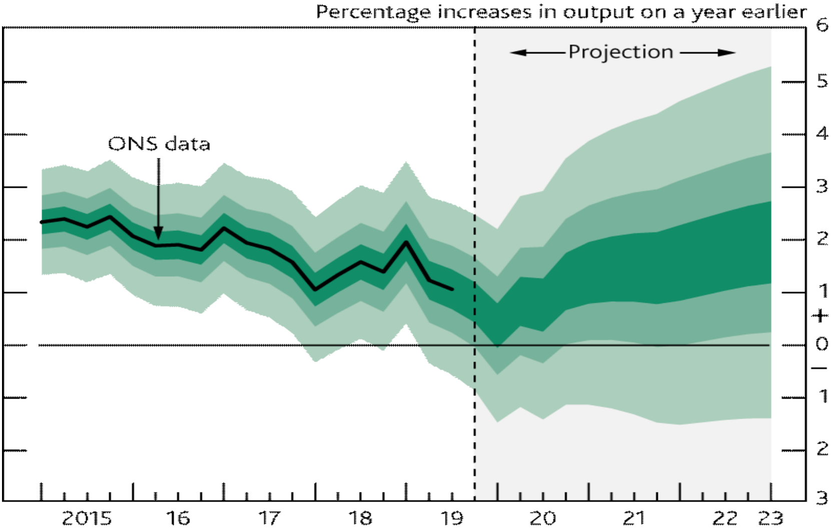
Density Forecasting

- A **density forecast** of the realisation of a random variable at some future time is an **estimate of the probability distribution of the possible future values of that variable.**
- It provides a complete description of the uncertainty associated with a forecast. A point forecast, which contains no indication of the associated uncertainty an interval forecast some indication.
- Density forecasts are now more common in macroeconomic forecasting e.g. fan charts for inflation.
- Density forecasts seem to avoid the problem of specifying loss function, but a loss function is required to estimate the parameters of the distribution.

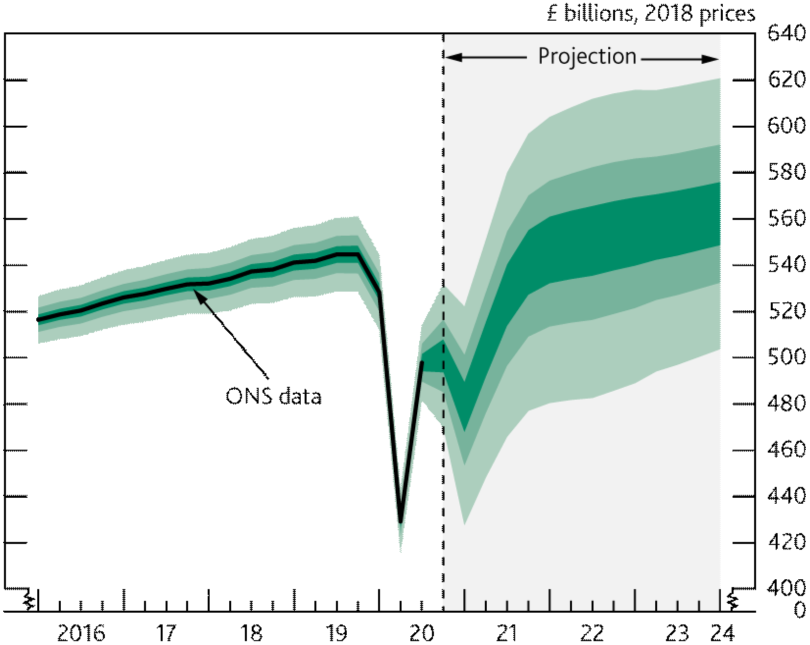
Density Forecasting under general loss functions

- The importance of publishing density forecasts then follows from the fact that we tend, in reality, not to know users' loss functions. Central banks, for example, do not quantify, explicitly at least, their loss functions, but we should not expect these (unknown to us) functions to be quadratic.
- For example, we should expect the range of uncertainty to matter to the Bank since it possibly does not care equally about inflation above and below the target. BoE symmetric target, ECB asymetric target.
- When the forecast user's loss function is asymmetric, such that positive and negative forecasting errors have differing costs, the user's "optimal" forecast need not equal the conditional mean; e.g. linex loss. or use of the mode

BofE Jan 2020 GDP Fan chart: nobody knows anything



February 2021 Bank forecast, note the change of scale



Why Density Forecasting?

- Provides more information to clients with different loss functions.
- Interest may lie in the dispersion or tails of the density itself; for example, inflation targets often focus the attention of monetary authorities to the probability of future inflation falling within some pre-defined target range, while users of growth forecasts may be concerned about the probability of recession
- These probability event forecasts can readily be extracted from the density forecast
- Bank doesn't give probabilties of tail events, outside the 90%, so while you can get mode or median you cannot get mean. Most people assume it follows the normal tails.
- Desnity forecast reminds the clients that the forecasters themselves expect the point forecasts to be 'wrong'
- However, clients often hate density forecasts: many just want a number.

Subjective Density Forecasts: Bank of England

- Based on a two-piece normal with parameters μ, σ_1 and σ_2,

$$f(y) = \begin{cases} A \exp[-(y-\mu)^2/2\sigma_1^2] & \text{if } y \leq \mu \\ A \exp[-(y-\mu)^2/2\sigma_2^2] & \text{if } y > \mu \end{cases}$$

  where $A = \sqrt{2\pi}(\sigma_1 + \sigma_2)^{-1}$.

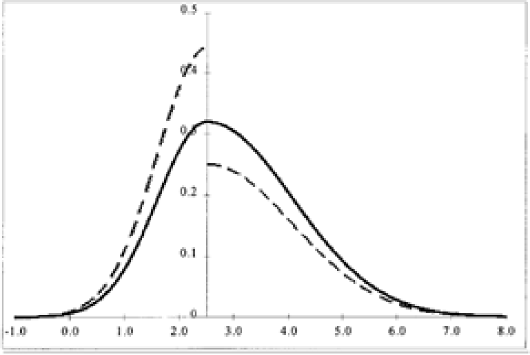
- With $\sigma_1 > \sigma_2$ this has positive skewness, which can be seen by noting that the mean is

$$E(Y) = \mu + \sqrt{\frac{2}{\pi}}(\sigma_2 - \sigma_1)$$

  and so exceeds the mode, μ, if $\sigma_2 > \sigma_1$ or the third moment about the mean is a positive multiple of $(\sigma_2 - \sigma_1)$

- The coefficient of Kurtosis exceeds 3 when ever $\sigma_2 \neq \sigma_1$, so the distribution shows excess kurtosis, is leptokurtic, has fatter tails than the normal.

Subjective Density Forecasts: Bank of England



Pdf of the two-pieces normal distribution: $\mu = 2.5$, $\sigma_1 = 0.902$ (left) and $\sigma_2 = 1.592$ (right).

Subjective Density Forecasts: Bank of England

- It is a convenient way to represent departures from the symmetry of the normal distribution since probability calculations can still be based on standard normal tables, with suitable scaling - however it has no convenient multivariate generalisation
- Would need multivariate generalisation to estimate the probability that both growth and inflation were below zero. They are not independent. Recession makes deflation more likely.
- This distribution is arrived at as the subjective assessment by the Bank's Monetary Policy Committee (MPC), based partly on past forecast errors, partly on a range of formal models, and partly on subjective judgements regarding the asymmetry of risk in the forecast

Subjective Density Forecasts: NIESR

- Most of the density forecasts which are produced regularly by national or international institutions are, however, constructed in a much less formal way.
- NIESR began publishing regular density forecasts in 1996, although it had been publishing a mean absolute error for its forecasts of inflation since the second quarter of 1992
- NIESR, in contrast to the Bank, imposes a normal distribution around their point forecast, with the variance determined on the basis of past forecast errors
- The window used to calculate this variance is quite important - there is uncertainty about the variance used to quantify the degree of uncertainty inherent in the fan-chart if the variance is changing

Model Based Density Forecasting

- In the absence of parameter uncertainty, the calculation of a probability forecast for a specified event is closely related to the forecast confidence interval.
- We can assume normality or an alternative parameteric distribution, like the Student's t-distribution, to allow for excess kurtosis, characteristic of finance and risk manangememt applications.
- One technique, among many, is the use of the bootstrap in the form of a stochastic simulation

Bootstrap methods

- Using bootstrap methods we can relax the normality assumption, suppose for example, we want a 1-step-ahead interval forecast for an AR(1) process. We know the future observation of interest is:

$$y_{T+1} = \phi y_T + \varepsilon_{T+1}$$

- We know y_T, and we can estimate ϕ and then proceed as if ϕ were known - here we abstract from parameter uncertainty - using the operational forecast $\widehat{y}_{T+1,T} = \widehat{\phi} y_T$

Bootstrap methods

- Using a normal distribution we would compute $\widehat{y}_{T+1,T} \pm z_{\alpha/2}\widehat{\sigma}$, alternatively we can proceed as follows
  1. Imagine that we could sample from the distribution of ε_{T+1} - whatever that distribution might be. We simply assign probability $1/T$ to each of the observed residuals (which are our estimates of the unobserved errors) and draw from then S times with replacement. Take S draws, $\left\{\varepsilon_{T+1}^{(i)}\right\}_{i=1}^{S}$, where S is a large number, say 10,000
  2. For each draw, construct the corresponding forecast of y_{T+1} as

$$\widehat{y}_{T+1,T}^{(i)} = \widehat{\phi} y_T + \varepsilon_{T+1}^{(i)}$$

  3. Then form a histogram of the $\widehat{y}_{T+1,T}^{(i)}$ values, which is the density - from which can can construct an interval statement of our choosing

Generating Simulated Errors

Shocks used in stochastic simulations can be generated in two basic ways:

1. The first is a *parametric* method where the errors are drawn from an assumed probability distribution function
2. Alternatively, one could employ a *non-parametric* procedure.

- An obvious non-parametric approach to generating the simulated errors, $u_{T+1}^{(s)}$, is simply to take random draws with replacements from the in-sample residual vectors $\{\widehat{u}_1, \ldots, \widehat{u}_T\}$
- These simulated errors have the same distribution properties and, in the multivariate case (e.g. with VAR forecasts), covariance structure as that observed in the original sample

The Evaluation of Interval Forecasts

- A ‘good’ interval forecast should, have correct *unconditional* coverage *ex post*; i.e., the outturn should fall in the interval the predicted proportion of times. (e.g 95% for a 95% confidence interval)
- In time-series it should also have correct *conditional* coverage, so for example in volatile periods, the interval is wider than in less volatile periods
- Let $I_t = 1$ if the outcome falls in the forecast interval t, 0 otherwise
- For an interval forecast with coverage probability p, forecasts have correct *conditional* coverage given information set Ω_{t-1}, if $\mathrm{E}(I_t \mid \Omega_{t-1}) = p$
- If $\Omega_{t-1} = \{I_{t-1}, I_{t-2}, ...\}$ then this implies $\{I_t\}$ is $i.i.d.$ Bernoulli with parameter p

The Evaluation of Interval Forecasts

- Regression-based tests of interval forecasts involve estimating:

$$I_t = \alpha + \beta\Omega_{t-1} + \varepsilon_t,$$

  where the set of interval forecasts are conditionally efficient when $\alpha = p$ and $\beta = 0$.
- Similarly to Mincer-Zarnowitz regressions, these regression-based tests distinguish between conditional and unconditional objectives
- The forecasts have correct unconditional coverage when $\alpha = p$, and are conditionally efficient when the forecast 'errors' are uncorrelated with information available at the time the forecast was made, i.e., $\alpha = p$ and $\beta = 0$

The Evaluation of Probability event Forecasts

- The same equation also serves as the basis for tests of probability event forecasts. Let $p_{t|t-1}$ to be the probability forecast made 1-period ahead of an event (such as a breach of the inflation target, a) happening at time t; $p_{t|t-1} = P(y_t \geq a \mid \Omega_{t-1})$
- Conditional efficiency, $\mathrm{E}\left[I_t \mid \Omega_{t-1}\right] = p_{t|t-1}$, then implies $\lambda = 1$, $\alpha = 0$ and $\beta = 0$ in the following variant:

$$I_t = \lambda p_{t|t-1} + \alpha + \beta \Omega_{t-1} + \varepsilon_t,$$

where the indicator variable I_t is re-defined with respect to the event forecasts

Probability Integral Transforms (pit's) Density Forecast Evaluation

- Suppose we have a series of 1-step-ahead (rolling) forecast densities $\{g(y_t \mid \Omega_{t-h})\}_{t=1}^{T}$, the task is then to determine whether this estimated density is equal to the true density $\{f(y_t \mid \Omega_{t-h})\}_{t=1}^{T}$
- *The* **probability integral transform** *is the cumulative density function corresponding to the density $g(y_t)$ evaluated at the observed value y_t*, given by

$$z_{t|t-h} = \int_{-\infty}^{y_t} g(u \mid \Omega_{t-h}) du = G(y_t \mid \Omega_{t-h}); \; (t = 1, ..., T).$$

- Call the density of z_t, $q(z_t)$. If the density and distribution fuinction $G(y_t) = F(y_t)$, the true value then $q(z_t)$ is the $U(0,1)$ density

Probability Integral Transforms (pit's) Density Forecast Evaluation

- When $h = 1$, $z_{t|t-h}$ is not just uniform but also independently distributed. In other words, one-step ahead density forecasts are optimal and capture all aspects of the distribution of y_t only when the $z_{t|t-1}$ are independently and uniformly distributed
- When $h > 1$ we should expect serial dependence in $z_{t|t-h}$ even for correctly specified density forecasts. Again this is analogous to expecting dependence (an $MA(h-1)$ process) when evaluating a sequence of rolling optimal h-step ahead point forecasts
- There is not, however, a one-for-one relationship between the point forecast errors and $z_{t|t-h}$
- One can test density forecasts unconditionally, via a distributional test, and conditionally via a test for independence.

Goodness-of-fit tests: in practice

- Evaluation tests are often based on the difference between the empirical distribution of $z_{t|t-h}$ and the cumulative distribution function of a uniform random variable on [0,1], i.e., the 45-degree line
- In many empirical studies, this has simply involved the application of a Kolmogorov-Smirnov or Anderson-Darling test for uniformity.
- For one-step ahead forecasts this is often supplemented with a separate test for the independence of $z_{t|t-h}$. e.g. LM tests for independence, and LR tests for both uniformity/normality and independence
- Many of the popular distributional tests, such as the K-S and A-D tests, are not robust to dependence, their properties having been developed under independence
- Plot the histogram of the PITs statistics and see if it looks like a uniform distribution.

Scoring Rules

- Scoring rules are (specific) loss functions assigning numerical scores based on the density forecast and the subsequent realisation. Measure relative not absolute performance
- Can be related to the Kullback-Leibler Information Criterion (KLIC) or distance
- KLIC measures "divergence", between the "true" but unknown conditional density $f(y_t \mid \Psi_{t-h})$, with information set Ψ_{t-h}, and the density forecast $g(y_t \mid \Omega_{it-h})$, defined with respect to forecaster i's information set Ω_{it-h}:

$$\begin{aligned} KLIC^i_{t|t-h} &= \mathrm{E}\left[\ln\ f(y_t \mid \Psi_{t-h}) - \ln\ g(y_t \mid \Omega_{it-h})\right] \\ &= \int f(y_t \mid \Psi_{t-h}) \ln\left\{\frac{f(y_t \mid \Psi_{t-h})}{g(y_t \mid \Omega_{it-h})}\right\} dy_t. \end{aligned}$$

- $KLIC^i_{t|t-h} = 0$ if and only if $g(y_t \mid \Omega_{it-h}) = f(y_t \mid \Psi_{t-h})$

The logarithmic scoring rule

- Consider the logarithmic scoring rule: $S(g(y_t \mid \Omega_{it-h}), y_t) = \ln g(y_t \mid \Omega_{it-h})$, where the density forecast is evaluated at the realisation of the random variable
- The logarithmic scoring rule is intuitively appealing as it gives a high score to a density forecast that provides a high probability to the value y_t that materialises
- Competing density forecasts can be ranked according to the size of $S(g(y_t \mid \Omega_{it-h}), y_t)$, with higher values indicating better performance

Log Score Tests

- Suppose there are two forecast densities of inflation, $p(\pi_\tau \mid I_{1,\tau})$ and $p(\pi_\tau \mid I_{2,\tau})$, so that the KLIC differential between them is the expected difference in their log scores:

$$d_\tau = \ln p(\pi_\tau \mid I_{1,\tau}) - \ln p(\pi_\tau \mid I_{2,\tau})$$

- The null hypothesis of equal forecast performance is $\mathcal{H}_0 : E(d_\tau) = 0$
- A test can then be constructed since the mean of d_τ over the evaluation period, $\overline{d}_\tau$, under appropriate assumptions, has the limiting distribution: $\sqrt{T}\overline{d}_\tau \rightarrow N(0, \Omega)$, where Ω is a consistent estimator of the asymptotic variance of d_τ
- Note: we can simply implement this test by taking the time series of log scores, for each recursion, and regressing the difference between the competing models log score on a constant and then testing the null of whether the constant is zero using a Newey-West adjusted standard errors

2018 Exam question on volatility .

Below are given results for two models for weekly data on the log of the Brent oil price, LB, where DLB is the first difference of LB In both cases the mean equation is an AR(4) and a t distribution is assumed. The first model assumes the variance follows a GARCH(1,1) process and the second model that it follows an EGARCH process.

(a) Show that the mean specification is a reparameterisation of an AR(4). Is log Brent I(0) or I(1)?

(b) Briefly explain the procedure for obtaining maximum likelihood estimates for non-linear models like these and comment on the issues in the case of these two models.

(c) Explain the difference between the GARCH and EGARCH models and comment on the results. Which model would you choose for forecasting and why?

(d) Explain how you would use the GARCH model to provide a 90% interval forecast for log Brent.

(e) Briefly describe at least three other procedures that might be used to forecast the Brent crude oil price.

Dependent Variable: DLB
Method: ML ARCH - Student's t distribution (BFGS / Marquardt steps)
Date: 02/19/18 Time: 17:36
Sample: 2/14/2014 2/09/2018
Included observations: 209
Convergence achieved after 42 iterations
Coefficient covariance computed using outer product of gradients
Presample variance: backcast (parameter = 0.7)
GARCH = C(6) + C(7)*RESID(-1)^2 + C(8)*GARCH(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.029469	0.024883	1.184304	0.2363
LB(-1)	-0.006975	0.005779	-1.207049	0.2274
DLB(-1)	0.363834	0.073824	4.928391	0.0000
DLB(-2)	-0.129121	0.079369	-1.626845	0.1038
DLB(-3)	0.046641	0.075975	0.613894	0.5393
		Variance Equation		
C	2.12E-05	1.79E-05	1.178729	0.2385
RESID(-1)^2	0.110597	0.053309	2.074646	0.0380
GARCH(-1)	0.887332	0.048123	18.43886	0.0000
T-DIST. DOF	8.626359	5.230666	1.649189	0.0991

R-squared	0.099032	Mean dependent var	-0.002381
Adjusted R-squared	0.081366	S.D. dependent var	0.043063
S.E. of regression	0.041274	Akaike info criterion	-3.739667
Sum squared resid	0.347522	Schwarz criterion	-3.595739
Log likelihood	399.7952	Hannan-Quinn criter.	-3.681476
Durbin-Watson stat	2.074339		

Dependent Variable: DLB
Method: ML ARCH - Student's t distribution (BFGS / Marquardt steps)
Date: 02/19/18 Time: 17:01
Sample: 2/14/2014 2/09/2018
Included observations: 209
Failure to improve likelihood (non-zero gradients) after 117 iterations
Coefficient covariance computed using outer product of gradients
Presample variance: backcast (parameter = 0.7)
LOG(GARCH) = C(6) + C(7)*ABS(RESID(-1)/@SQRT(GARCH(-1))) + C(8)
*RESID(-1)/@SQRT(GARCH(-1)) + C(9)*LOG(GARCH(-1))

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.027043	0.011622	2.326813	0.0200
LB(-1)	-0.006963	0.002589	-2.688968	0.0072
DLB(-1)	0.302606	0.071510	4.231644	0.0000
DLB(-2)	-0.104367	0.072949	-1.430678	0.1525
DLB(-3)	0.076877	0.068439	1.123291	0.2613
		Variance Equation		
C(6)	-0.118746	0.072768	-1.631840	0.1027
C(7)	-0.143436	0.090906	-1.577855	0.1146
C(8)	-0.182048	0.028849	-6.310453	0.0000
C(9)	0.964832	1.0E-103	9.6E+102	0.0000
T-DIST. DOF	34.30841	138.3080	0.248058	0.8041

R-squared	0.105767	Mean dependent var	-0.002381
Adjusted R-squared	0.088233	S.D. dependent var	0.043063
S.E. of regression	0.041119	Akaike info criterion	-3.899789
Sum squared resid	0.344924	Schwarz criterion	-3.739869
Log likelihood	417.5279	Hannan-Quinn criter.	-3.835132
Durbin-Watson stat	1.958783		

Sketch answer (a-b)

(a)

$$\begin{aligned} y_t &= \alpha + \rho_1 y_{t-1} + \rho_2 y_{t-2} + \rho_3 y_{t-3} + \rho_4 y_{t-4} + \varepsilon_t \\ y_t - y_{t-1} &= \alpha + (\rho_1 + \rho_2 + \rho_3 + \rho_4 - 1) y_{t-1} \\ &\quad -(\rho_2 + \rho_3 + \rho_4)(y_{t-1} - y_{t-2}) \\ &\quad -(\rho_3 + \rho_4)(y_{t-2} - y_{t-3}) - \rho_4 (y_{t-3} - y_{t-4}) + +\varepsilon_t \end{aligned}$$

It is clearly I(1) since the coefficient of LB is not significant even at standard critical values, let alone ADF ones.

(b) A picture showing an iterative procedure from initial starting values would be OK, talk about convergence, difficulty of estimating Hessian and the use of OPG formula instead. The EGARCH did not converge and is giving silly values Standard error on the GARCH term c(9) of zero. So use GARCH.

Sketch answer (c-e)

(c) Just give and interpret the equations for the variance processes.
(d) The mean gives a point forecast, the variance term gives a forecast of the variance, so you can use the t distribution to get an interval with 90% coverage for the forecast variance.
(e) Read the paper on Moodle and discuss for instance a VAR a DSGE supply-demand model and the futures price.

2017 Exam Question

Consider a series y_t which, conditional on information at $t-1$, is distributed $N(\mu_t, h_t)$, $t = 1, 2, ..., T$.

1.
   1.1 Explain how you would specify and estimate a GARCH model for the processes determining μ_t and h_t. Explain how to construct an interval forecast for y_{T+1}.
   1.2 Explain how to evaluate a time series of interval forecasts once sufficient data on realisations were available.
   1.3 Give a definition of a density forecast and explain how to construct a density forecast for a GARCH model.
   1.4 Describe the Probability Integral Transform (PITs) and how they might be used to evaluate density forecasts.
   1.5 Above a Gaussian distribution was assumed. How should interval and density forecasts be constructed if the distribution was not known.

Sketch answer (a)

A GARCH model requires specification of a process for μ_t. and h_t, for instance and AR1 for the mean and a GARCH(1,1) for the variance,

$$\begin{aligned} y_t &= \beta_0 + \beta_1 y_{t-1} + u_t \\ h_t &= \alpha_0 + \alpha_1 h_{t-1} + \alpha_2 \widehat{u}_{t-1} \end{aligned}$$

Estimation is by maximum likelihood, using the assumed normality (or whatever distribution is specified). Explain non-linear estimation.
An interval forecast for y_{T+1} takes the form

$$\begin{aligned} & \widehat{\mu}_{T+1} + k\widehat{h}_{T+1} \text{ to } \widehat{\mu}_{T+1} - k\widehat{h}_{T+1} \\ \widehat{\mu}_{T+1} &= \hat{\beta}_0 + \hat{\beta}_1 y_T \\ h_{T+1} &= \hat{\alpha}_0 + \hat{\alpha}_1 h_T + \hat{\alpha}_2 \widehat{u}_T \end{aligned}$$

where k would be 1.96.for a 95% confidence interval and a Gaussian distribution.

Sketch answer (b-c)

- For a 95% confidence intervals the actual should lie within the forecast interval 95% of the times. You could test whether the actual proportion was significantly different from the expected proportion, see earlier slides..
- A density forecast gives probabilities over the whole range like the Bank fan chart, here you would use the normal distribution assumption to plot out the density given the mean and variance for each period.

Sketch answer (d-e)

- *The* **probability integral transform** *is the cumulative density function corresponding to the density $g(y_t)$ evaluated at the observed value y_t*, given by

$$z_{t|t-h} = \int_{-\infty}^{y_t} g(u \mid \Omega_{t-h}) du = G(y_t \mid \Omega_{t-h}); \ (t = 1, ..., T).$$

  If $g(y_t) = f(y_t)$, then $q(z_t)$ is the $U(0,1)$ density. Thus this suggests we evaluate densities by assessing whether the PIT series $\{z_t\}_{t=1}^{T}$ is $i.i.d. U(0,1)$.
- Judgementally like the Bank or by drawing from the observed residuals. Expand this.