2. Volatility models

Javier Mencía
javier.mencia@bde.es

Banco de España, Financial Stability Department

Outline of the presentation

Outline

1. Introduction
2. Volatility models: ARCH, GARCH
3. Volatility and non-normal distributions to model tail events

Conditional and unconditional moments

- Volatility is defined as the conditional standard deviation of the underlying asset return y_t.
- The unconditional variance is the variance of the time series without controlling for any information: $\sigma^2 = Var(y_t)$. In practice, it is the sample variance, which is a constant that yields the long term variance over the sample.
- In contrast, the conditional variance will change at every point in time because it depends on the history of returns up to that point.

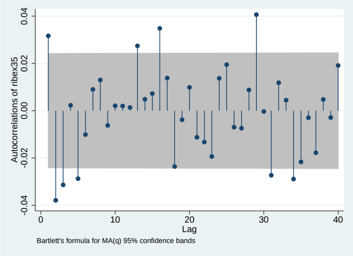
$$\sigma_t = \sqrt{Var(y_t|I_{t-1})}.$$

- Volatility is sometimes expressed in annualised terms ($\times$250 for weekly data, $\times$12 for monthly data).

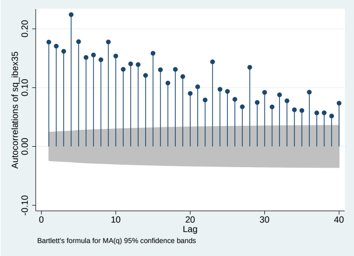
Motivation

- In part 1, we have treated volatility as constant, but the volatility of financial asset returns changes over time.
- Furthermore, it usually follows a regular pattern, with periods of high (low) volatility often followed by periods of high (low) volatility.
- Since σ_t is not directly observable, we need to develop models to filter it out from the data.
- Volatility modelling is key for many different applications: option pricing, risk management, asset allocation, systemic risk measurement.

Ibex-35 returns correlogram

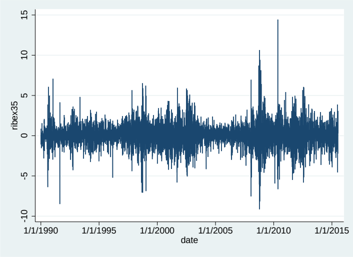


Ibex-35 squared returns correlogram

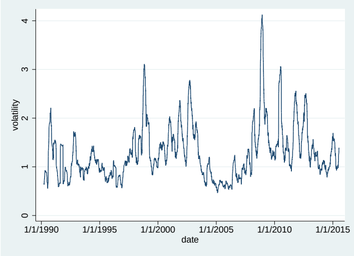


- Volatility evolves from periods of unusually high volatility to periods of low volatility (high and low vol observations cluster together).

Ibex-35 daily returns (%)



Ibex-35 volatility, estimated with a 60-day moving window



- Volatility reacts differently to a big price increase than to a big price drop.

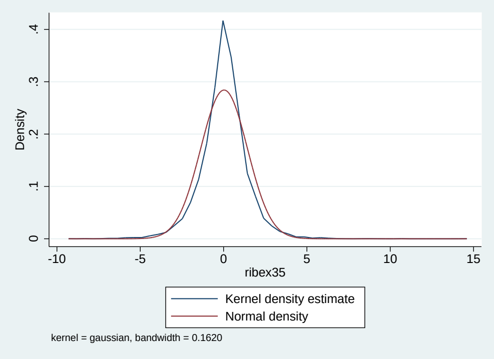
```
. regress sq_ibex35 l.sq_ibex35 l2.sq_ibex35 l3.sq_ibex35 l.ribex35, r;

Linear regression                                      Number of obs =    2564
                                                       F(  4,  2559) =    7.51
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.0836
                                                       Root MSE      =  4.5916
```

sq_ibex35	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
sq_ibex35						
L1.	.1197108	.0399601	3.00	0.003	.0413533	.1980683
L2.	.2084699	.0570815	3.65	0.000	.0965393	.3204006
L3.	.0651541	.0399854	1.63	0.103	-.0132529	.1435611
ribex35						
L1.	-.3796576	.114791	-3.31	0.001	-.6047502	-.1545649
_cons	1.170232	.1169486	10.01	0.000	.9409083	1.399555

Stylised facts

Financial returns are fat tailed and sometimes asymmetric

Ibex-35 kernel density estimate



- Volatility models are only able to capture part of this heteroskedasticity.

Outline

Volatility modelling

- Modelling volatility involves modelling the conditional means and variances simultaneously.
- For instance, in a Gaussian model we would have:

$$y_t = \mu_t + \sigma_t \varepsilon_t,$$

  where $\mu_t = E(y_t | I_{t-1})$, $\sigma_t = V(y_t | I_{t-1})$ and ε_t *iid* $N(0, 1)$.
- For stock returns, μ_t can many times be assumed to be a constant, due to the low persistence of this kind of data.
- We will consider conditional heteroskedastic models, in which σ_t is governed by a determistic function given past information.

The ARCH model

- The ARCH model is the natural extension of an autoregressive model to volatility.
- In an ARCH(m) model, the conditional variance is a quadratic function of past squared residuals:

$$\sigma_t^2 = \alpha_0 + \alpha_1 (y_{t-1} - \mu_{t-1})^2 + \alpha_2 (y_{t-2} - \mu_{t-2})^2 + \cdots + \alpha_m (y_{t-m} - \mu_{t-m})^2,$$

  where $\alpha_0 > 0$ and $\alpha_i \geq 0$, for $i > 0$ to ensure positivity.
- Intuitively, large past shocks imply a large variance at t, which in turn increases the probability of another large shock (i.e. volatility clustering).
- The coefficients must satisfy some restrictions to ensure that the unconditional variance σ_t^2 is finite.

Properties of the ARCH(1) model

- Consider the model

$$y_t = \sigma_t \varepsilon_t, \;\; \sigma_t^2 = \alpha_0 + \alpha_1 y_{t-1}^2,$$

  where the mean has been set to zero for simplicity.

- Unconditional variance:

$$\begin{aligned} V(y_t) &= E(y_t^2) = E[\sigma_t^2 E(\varepsilon_t^2 | I_{t-1})], \\ &= \alpha_0 + \alpha_1 E(\sigma_{t-1}^2) = \alpha_0/(1-\alpha_1). \end{aligned}$$

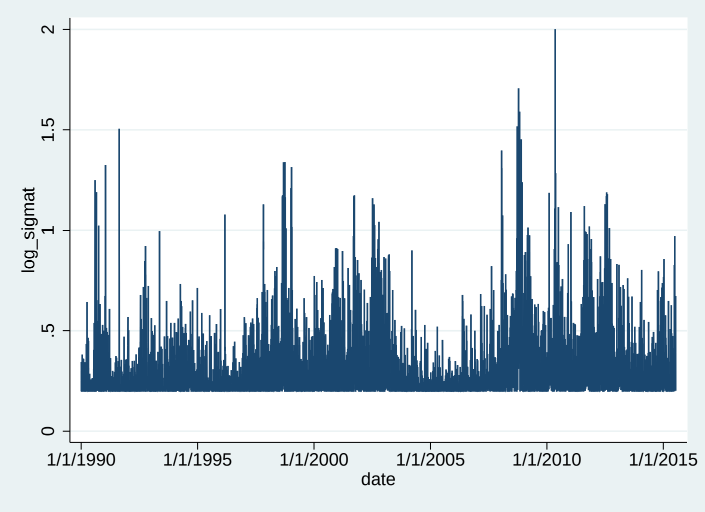
- Hence, $0 \leq \alpha_1 < 1$ so that the unconditional variance is finite.
- Additional restrictions may have to be imposed if we need higher order moments to be finite.

$$\text{Kurtosis} = 3\frac{1-\alpha_1^2}{1-3\alpha_1^2} > 3. \text{ (i.e. Non-Gaussian)}$$

```
. arch ribex35, arch(1) vce(robust);

ARCH family regression

Sample: 2 - 6588                                 Number of obs    =       6587
Distribution: Gaussian                           Wald chi2(.)     =          .
Log pseudolikelihood = -11349.53                 Prob > chi2      =          .

------------------------------------------------------------------------------
             |             Semirobust
     ribex35 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
ribex35      |
       _cons |   .0562341   .0173412     3.24   0.001      .022246    .0902222
-------------+----------------------------------------------------------------
ARCH         |
        arch |
         L1. |   .2575898   .0363288     7.09   0.000     .1863865     .328793
             |
       _cons |   1.481114   .0599323    24.71   0.000     1.363649    1.598579
------------------------------------------------------------------------------
```

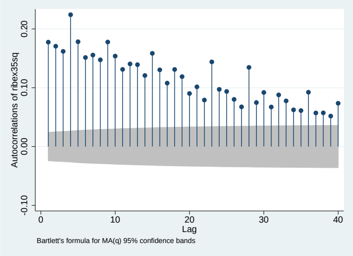
ARCH(1) example with Ibex-35 daily returns

One day-ahead volatility estimates (in logs)

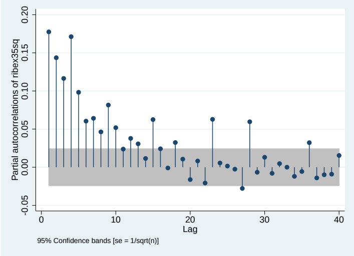


ARCH(1) example with Ibex-35 daily returns. Model fit

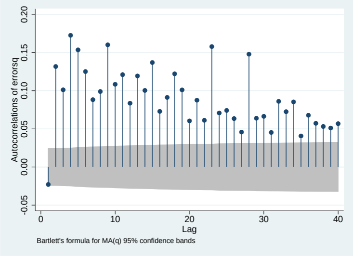
Correlogram: y_t^2



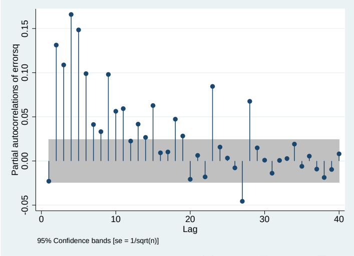
Partial correlogram: y_t^2



Correlogram: $(y_t/\sigma_t)^2$



Partial correlogram: $(y_t/\sigma_t)^2$



ARCH(10) example with Ibex-35 daily returns

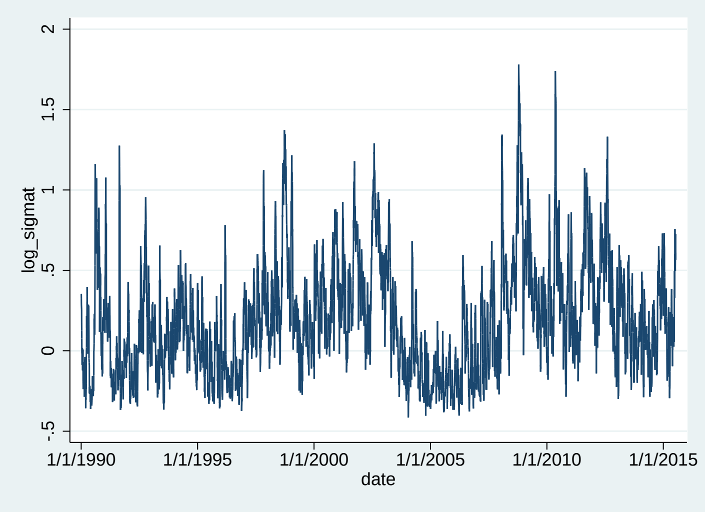
Estimates

```
. arch ribex35, arch(1/10) vce(robust);
```

ribex35	Coef.	Semirobust Std. Err.	z	P>|z|	[95% Conf.	Interval]
arch						
L1.	.0797631	.0181918	4.38	0.000	.0441079	.1154184
L2.	.0756889	.0175411	4.31	0.000	.0413091	.1100688
L3.	.1275963	.0223611	5.71	0.000	.0837693	.1714234
L4.	.1283109	.0255551	5.02	0.000	.0782239	.178398
L5.	.0801231	.0198484	4.04	0.000	.041221	.1190252
L6.	.0739575	.0176792	4.18	0.000	.0393068	.1086082
L7.	.0584332	.0199125	2.93	0.003	.0194053	.097461
L8.	.0636567	.0192321	3.31	0.001	.0259625	.101351
L9.	.0748387	.0277646	2.70	0.007	.0204212	.1292563
L10.	.0687725	.0178184	3.86	0.000	.033849	.103696
cons	.3894397	.0503322	7.74	0.000	.2907904	.488089

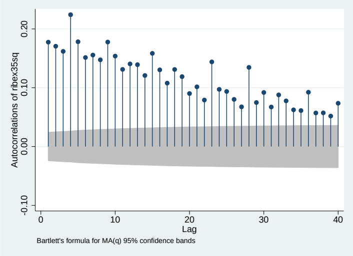
ARCH(10) example with Ibex-35 daily returns

One day-ahead volatility estimates (in logs)

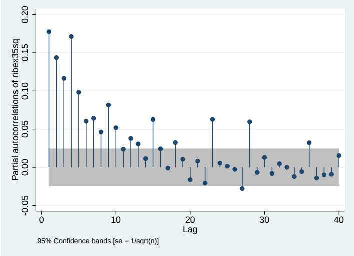


ARCH(10) example with Ibex-35 daily returns. Model fit

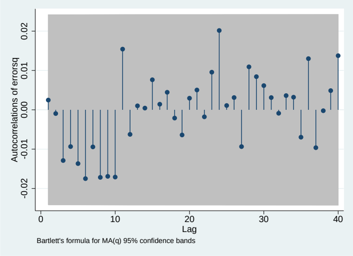
Correlogram: y_t^2



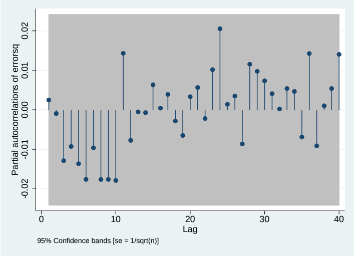
Partial correlogram: y_t^2



Correlogram: $(y_t/\sigma_t)^2$



Partial correlogram: $(y_t/\sigma_t)^2$



Weaknesses of the ARCH model

- The model assumes that positive and negative shocks have the same impact on volatility.
- In a Gaussian model with ARCH(1) effects, α must be smaller than $1/3$ so that the fourth moment is finite.
- An ARCH of a very high order is required to capture the very long term persistence that is typical in financial data, which is costly in terms of parameters.

The GARCH model

- The GARCH model avoids using a long series of ARCH effects by subsuming them in the lagged variance.
- Typically, a GARCH(1,1) suffices to provide an acceptable fit:

$$\sigma_t^2 = \alpha_0 + \alpha_1 (y_{t-1} - \mu_{t-1})^2 + \beta_1 \sigma_{t-1}^2.$$

- Intuition: this model is equivalent to an ARCH(∞) whose coefficients are functions of just three free parameters,

$$\sigma_t^2 = \frac{\alpha_0}{1-\beta_1} + \alpha_1 \epsilon_{t-1}^2 + \alpha_1 \beta_1 \epsilon_{t-2}^2 + \alpha_1 \beta_1^2 \epsilon_{t-3}^2 + \cdots ,$$

where $\epsilon_t = y_t - \mu_t$.

GARCH(1,1)

- The unconditional variance can be obtained as:

$$E[\sigma_t^2] = \alpha_0 + \alpha_1 E[(y_{t-1} - \mu_{t-1})^2] + \beta_1 E[\sigma_{t-1}^2] \Rightarrow E[\sigma_t^2] = \frac{\alpha_0}{1 - \alpha_1 - \alpha_2}.$$

- Positivity of σ_t^2 and the existence of the unconditional variance requires that $\alpha_0 > 0, \alpha_1, \beta_1 \geq 0$ and $\alpha_1 + \beta_1 < 1$.
- This model also yields unconditional positive excess kurtosis:

$$\text{Kurtosis} = \frac{3[1 - (\alpha_1 + \beta_1)^2]}{1 - (\alpha_1 + \beta_1)^2 - 2\alpha_1^2} > 3.$$

GARCH(1,1) example with Ibex-35 daily returns

Estimates

```
. arch ribex35, arch(1) garch(1) vce(robust);
```

ribex35	Coef.	Semirobust Std. Err.	z	P>|z|	[95% Conf.	Interval]
ribex35						
_cons	.0703088	.0139937	5.02	0.000	.0428817	.0977358
ARCH						
arch						
L1.	.093113	.0113171	8.23	0.000	.0709318	.1152941
garch						
L1.	.8911576	.0110973	80.30	0.000	.8694073	.9129079
_cons	.0328289	.0085832	3.82	0.000	.0160061	.0496516

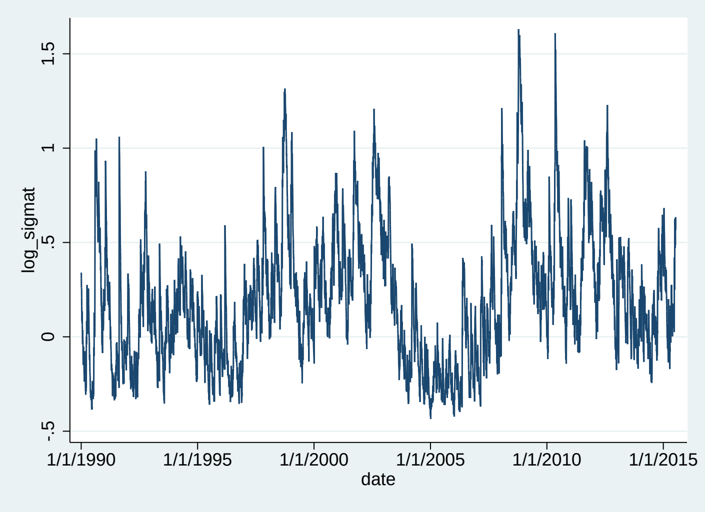
Interpretation of the GARCH(1,1) parameters

$$\sigma_t^2 = \alpha_0 + \alpha_1(y_{t-1} - \mu_{t-1})^2 + \beta_1\sigma_{t-1}^2.$$

- The error parameter α_1 measures the reaction of σ_t to market shocks. If α_1 is large, then σ_t is very sensitive to (recent) market events.
- The lag parameter β_1 measures the persistence irrespective of market events. If it is large, then volatility decreases very slowly after a large market shock.
- In most (if not all) financial applications, β_1 clearly dominates over α_1.
- $\alpha_1 + \beta_1$ determines the speed of convergence of σ_t to the long run level (slow for high values).
- Lastly, $\alpha_0/(1 - \alpha_1 - \alpha_2)$ is the long term average volatility.

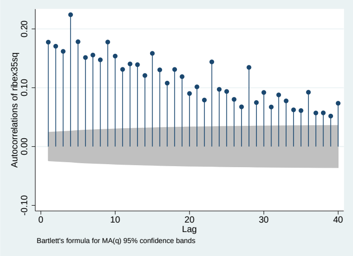
GARCH(1,1) example with Ibex-35 daily returns

One day-ahead volatility estimates (in logs)

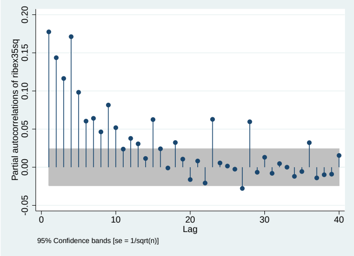


GARCH(1,1) example with Ibex-35 daily returns. Model fit

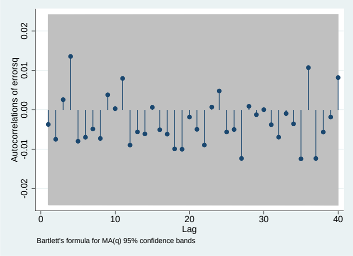
Correlogram: y_t^2



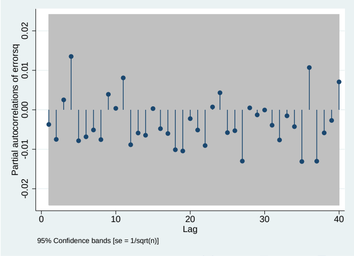
Partial correlogram: y_t^2



Correlogram: $(y_t/\sigma_t)^2$



Partial correlogram: $(y_t/\sigma_t)^2$



ARCH vs GARCH

Likelihood fit

	Log-likelihood	Parameters
ARCH(1)	-11349.5	3
ARCH(10)	-10673.3	12
GARCH(1,1)	-10654.3	4

Extensions of the GARCH(1,1) model

- The GARCH(1,1) is still extensively used in its original form.
- However, many extensions have been proposed to overcome its main limitation: absence of symmetry.
- One interesting example addressing this issue is the asymmetric GARCH model:

$$\sigma_t^2 = \alpha_0 + \alpha_1(|\epsilon_{t-1}| + \gamma_1\epsilon_{t-1})^2 + \beta_1\sigma_{t-1}^2,$$

  where $\epsilon_t = y_{t-1} - \mu_{t-1}$.
- If $\gamma_1 < 0$, then negative shocks increase volatility more than positive ones.
- In addition, this model introduces not only excess kurtosis but also skewness in the unconditional distribution.

Asymmetric GARCH(1,1) estimates

```
. arch ribex35, aarch(1) garch(1) vce(robust);
```

ribex35	Coef.	Semirobust Std. Err.	z	P>|z|	[95% Conf.	Interval]
ribex35						
_cons	.0378019	.0138386	2.73	0.006	.0106788	.064925
ARCH						
aarch						
L1.	.0646669	.0109868	5.89	0.000	.0431332	.0862007
aarch_e						
L1.	-.4307833	.0767934	-5.61	0.000	-.5812955	-.280271
garch						
L1.	.9052816	.0116056	78.00	0.000	.8825351	.9280281
_cons	.0325056	.0073692	4.41	0.000	.0180621	.046949

Likelihood fit

	Log-likelihood	Parameters
ARCH(1)	-11349.5	3
ARCH(10)	-10673.3	12
GARCH(1,1)	-10654.3	4
Asymmetric GARCH(1,1)	-10573.7	5

Estimation

- Volatility models are estimated by maximum likelihood.
- The parameter estimates are those that maximise the natural logarithm of the likelihood of the data.
- In the Gaussian case, this is:

$$L = -0.5 \sum_t log(2\pi\sigma_t^2) - 0.5 \sum_t \frac{(y_t - \mu_t)^2}{\sigma_t^2}$$

- Example: using solver in Excel.

Outline

1. Introduction
2. Volatility models: ARCH, GARCH
3. Volatility and non-normal distributions to model tail events

Consistency of Gaussian estimates

- Interestingly, the mean and variance estimates based on the Gaussian distribution can be consistent even if the true distribution is not Gaussian.
- The only requirement is that the models for the mean and variance are correctly specified.
- Hence, if we are only interested in the first two moments, the Gaussian distribution is a good choice (consistent and robust).
- However, if we want to compute higher order moments (skewness, kurtosis,...) or tail measures (e.g. VaR), then an incorrectly specified distribution may yield significant biases.

Non-Gaussian data

- Although it is true that the volatility model can generate unconditional skewness and kurtosis, it turns out that this is not enough for most financial applications.
- Even after accounting for time varying volatility, stock returns are typically characterised by further excess kurtosis and sometimes skewness.
- This departure from normality generally becomes more intense for increasing frequencies.
- There is a huge amount of research on extensions of the Gaussian density allowing for fat tails and asymmetry.
- We will focus on one of the simplest cases, the Student t distribution, which is fat tailed but symmetric.

- If the normality assumption is correct, then $\epsilon_t = (y_t - \mu_t)/\sigma_t$ should be Gaussian.
- We can easily test normality by considering the Jarque-Bera test:

$$JB = \frac{T}{6}sk^2 + \frac{T}{24}(ku - 3)^2 \to \chi^2_2$$

- In this case, normality can be easily rejected at all conventional levels:

	Coefficient	Jarque-Bera	P-value
Skewness	-0.35	137.26	0.00
Kurtosis	6.79	3935.14	0.00
		4072.40	0.00

The Student t distribution

- The density can be written as

$$f(x) = k\left(1 + \frac{x^2}{\nu - 2}\right)^{-\frac{\nu+1}{2}}$$

  where ν are the degrees of freedom.
- This distribution nests the standard normal density as a particular case when $\nu \to \infty$.
- The distribution is symmetric and leptokurtic, where

$$ku = 3\frac{\nu - 2}{\nu - 4} > 3.$$

- The variance is finite for $\nu > 2$, whereas kurtosis requires $\nu > 4$ to exist.

GARCH(1,1) for the Ibex-35 index returns, with Student t innovations

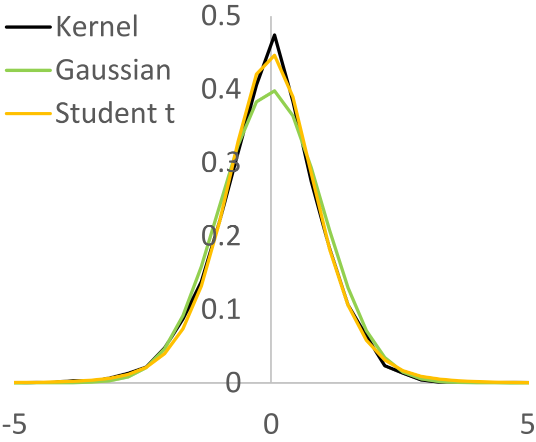
```
. arch ribex35, arch(1) garch(1) vce(robust) distribution(t);
```

ribex35	Coef.	Semirobust Std. Err.	z	P>|z|	[95% Conf.	Interval]
ribex35						
_cons	.0722811	.012381	5.84	0.000	.0480148	.0965473
ARCH						
arch L1.	.0902986	.0089906	10.04	0.000	.0726773	.10792
garch L1.	.9036045	.0091606	98.64	0.000	.8856501	.9215588
_cons	.0184234	.004176	4.41	0.000	.0102386	.0266082
/lndfm2	1.730824	.1396414	12.39	0.000	1.457132	2.004516
df	7.645305	.7883181			6.293629	9.422503

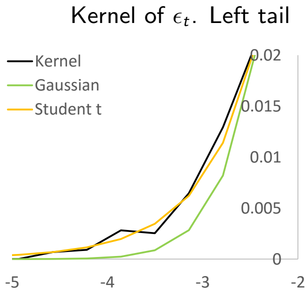
lndfm2$= \log(\nu - 2)$.

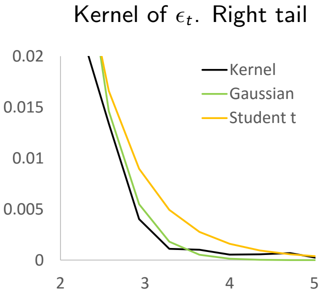
GARCH(1,1) for the Ibex-35 index returns, with Student t innovations

Kernel of ϵ_t.



GARCH(1,1) for the Ibex-35 index returns, with Student t innovations





Estimation

- The GARCH model with Student t innovations is also estimated y maximum likelihood.

$$
\begin{aligned}
L \quad = \quad & -0.5 \sum_t \log(\pi(\nu-2)\sigma_t^2) + \log\Gamma[0.5(\nu+1)] - \log\Gamma(0.5\nu) \\
& -0.5(\nu+1) \sum_t \log\left[1 + (\nu-2)^{-1}\frac{(y_t-\mu_t)^2}{\sigma_t^2}\right]
\end{aligned}
$$

- Example: using solver in Excel.

Imposing long term volatility

- The long term volatility forecast ω is sensitive to the choice of the historical data estimation period:

$$\sigma_t^2 = \omega(1 - \alpha_1 - \beta_1) + \alpha_1(y_{t-1} - \mu_{t-1})^2 + \beta_1\sigma_{t-1}^2.$$

- If there have been some extreme market movements during the sample, the estimated value of ω can be too high for prediction purposes.
- For instance, volatility reached unprecedented low levels between 2003 and 2007, biasing down volatility forecasts and tail risk measures.
- This is why sometimes the value of ω is imposed using a personal view about long term volatility.