Notes for Masters Thesis

Michael Kao

March 28, 2012

1 logd.mgarch

Old Stuff

1.1 log density

$$
\begin{aligned}
f(\mathbf{x};\theta_i,\sigma_t) &= \frac{1}{\sqrt{2\pi\theta_i^2\sigma_t^2}}\exp\left(-\frac{x^2}{2\theta_i^2\sigma_t^2}\right) \\
l(\mathbf{x};\theta_i,\sigma_t) = \log f(\mathbf{x};\theta_i,\sigma_t) &= -\frac{1}{2}\log 2\pi - \log\theta_i - \log\sigma_t - \frac{x_t^2}{2\theta_i^2\sigma_t^2} \\
\frac{\partial l}{\partial \sigma_t} &= \left(-\frac{1}{\sigma_t} + \frac{x_t^2}{\theta_i^2\sigma_t^3}\right)
\end{aligned}
$$

where

$$
\begin{aligned}
\sigma_t^2 &= \alpha_0 + \alpha_1\sigma_{t-1}^2 + \beta_1\sigma_{t-1}^2 \\
&= \alpha_0\sum_{i=0}^{t-2}\alpha_1^i + \alpha_1^{t-1}\sigma_1^2 + \beta_1\sum_{j=0}^{t-2}\alpha_1^j x_{t-j-1} \quad \text{for } t \geq 2 \\
\sigma_t &= \left(\alpha_0\sum_{i=0}^{t-2}\alpha_1^i + \alpha_1^{t-1}\sigma_1^2 + \beta_1\sum_{j=0}^{t-2}\alpha_1^j x_{t-j-1}\right)^{1/2}
\end{aligned}
$$

Then

$$\frac{\partial \sigma_t}{\partial \alpha_0} = \frac{1}{2\sigma_t}\left(\sum_{i=0}^{t-2}\alpha_1^i\right)$$
$$\frac{\partial \sigma_t}{\partial \alpha_1} = \frac{1}{2\sigma_t}\left(\alpha_0\sum_{i=1}^{t-2} i\alpha_1^{i-1} + (t-1)\alpha_1^{t-2}\sigma_1^2 + \beta_1\sum_{j=1}^{t-2} j\alpha_1^{j-1}x_{t-j-1}\right)$$
$$\frac{\partial \sigma_t}{\partial \beta_1} = \frac{1}{2\sigma_t}\left(\sum_{j=0}^{t-2}\alpha_1^j x_{t-j-1}\right)$$
$$\frac{\partial \sigma_t}{\partial \sigma_1} = \frac{1}{2\sigma_t}\left(2\alpha_1^{t-1}\sigma_1\right)$$

1.2 Partial Beta

$$\begin{aligned}
\frac{\partial l}{\partial \alpha_0} &= \frac{\partial l}{\partial \sigma} \times \frac{\partial \sigma}{\partial \alpha_0} \\
&= \left(\sum_{t=1}^{T}\sum_{i=1}^{M} -\frac{1}{\sigma_t} + \frac{x_t^2}{\theta_i^2\sigma_t^3}\right) \times \frac{1}{2\sigma_t}\left(\sum_{i=0}^{t-2}\alpha_1^i\right) \\
\frac{\partial l}{\partial \alpha_1} &= \frac{\partial l}{\partial \sigma} \times \frac{\partial \sigma}{\partial \alpha_1} \\
&= \left(\sum_{t=1}^{T}\sum_{i=1}^{M} -\frac{1}{\sigma_t} + \frac{x_t^2}{\theta_i^2\sigma_t^3}\right) \times \frac{1}{2\sigma_t}\left(\alpha_0\sum_{i=1}^{t-2} i\alpha_1^{i-1} + (t-1)\alpha_1^{t-2}\sigma_1^2 + \beta_1\sum_{j=1}^{t-2} j\alpha_1^{j-1}x_{t-j-1}\right) \\
\frac{\partial l}{\partial \beta_1} &= \frac{\partial l}{\partial \sigma} \times \frac{\partial \sigma}{\partial \beta_1} \\
&= \left(\sum_{t=1}^{T}\sum_{i=1}^{M} -\frac{1}{\sigma_t} + \frac{x_t^2}{\theta_i^2\sigma_t^3}\right) \times \frac{1}{2\sigma_t}\left(\sum_{j=0}^{t-2}\alpha_1^j x_{t-j-1}\right) \\
\frac{\partial l}{\partial \sigma_1} &= \frac{\partial l}{\partial \sigma} \times \frac{\partial \sigma}{\partial \sigma_1} \\
&= \left(\sum_{t=1}^{T}\sum_{i=1}^{M} -\frac{1}{\sigma_t} + \frac{x_t^2}{\theta_i^2\sigma_t^3}\right) \times \frac{\alpha^{t-1}\sigma_1}{\sigma_t}
\end{aligned}$$

1.3 Partial Theta

$$\frac{\partial l}{\partial \boldsymbol{\theta}} = \sum_{t=1}^{T}\sum_{i=1}^{M} -\frac{1}{\theta_i} + \frac{x_t^2}{\theta_i^3 \sigma_t^2}$$
$$\frac{\partial^2 l}{\partial \boldsymbol{\theta}^2} = \sum_{t=1}^{T}\sum_{i=1}^{M} \frac{1}{\theta_i^2} - \frac{3x_t^2}{\theta_i^4 \sigma_t^2}$$

2 Improvements and Further actions

1. Get the model (With scaling) CORRECT!!!!!!!
2. Include skewness, and mean estimation (This might be very interesting and useful as we have see that commodity skews different direction to the normal commodities). The skewness seems slightly difficult to implement as it may not have analytical solution, will have a look at the implementation in the fGarch package.
3. survey all the GARCH/financial package.
4. Work on all sorts of financial data.
5. Get Matthieu and others at FAO to review it.
6. Improve the efficiency of the logd.mgarch function.
7. Test the difference between real data and simulated data.
8. Have a look at the FAO book and talk to Matthieu about how the model can work. How would I assess the model other than using likelihood and information criteria based method? (e.g. prediction)

3 18th of December 2011

3.1 Skewed Normal Distribution

The skewness formulae adopted is based on the formulation of Fernandez and Stell [1998], the framework is quite general and can be used to introduce skewness to all unimodal, symmetric distributions. This form is used for two reasons, (1) It is also the same form used by fGarch package, (2) The more widely used form of $f(x) = 2\phi(x)\Phi(\alpha x)$ requires solving the derivative of log of the error function and thus can get quite complicated.

Any unimodal, symmetric density can be skewed by the following transformation:

$$f(z|\xi) = \frac{2}{\xi + \frac{1}{\xi}} \left[f(\xi z) H(-z) + f(\frac{z}{\xi}) H(z) \right]$$

Where $0 < \xi < \infty$ is the shape parameter which describes the degree of asymmetry. $H(z) = (1 + sign(z))/2$ is the heaviside unit step function.

The new log density is thus,

$$\begin{aligned} \log(f(x;\mu,\theta^2,\sigma^2,\xi)) = \log(2) + \log(\xi + 1/\xi) \\ - \frac{1}{2}\log(2\pi) - \log(\theta) - \log(\sigma) \\ - \frac{(x\xi - \mu)^2 H(-x) + (\frac{x}{\xi} - \mu)^2 H(x)}{2\theta^2\sigma^2} \end{aligned}$$

3.2 Thoughts and improvements

- Utilize github for facilitating code sharing between Yong and me.
- Now that most part of the model are in place, we can start carrying out the modelling and also moving back to testing of the model.
- Will try to get some non-stock market data and see how the model work. (e.g. Food/commodity price.)
- Application of model with VAR/ES type of risk measure. (e.g. If there is a food shock, how many people will be affected by this shock? Death, undernourishment, malnutrition ... etc)
- Maybe we can contact the author of the fGarch package and see if it is possible for them to incoporate our mixtrue density framework into their package?
- The tails of the distribution being influenced by the skewness.
- plot the skewness (curve) and the residual plot.
- include p, q terms.
- 10th of January.
- Incorporate the reparameterised skewness distribution, this possibiliy fixes the correlation and the NA generating problem we are having at the moment.

4 12th of January 2012

- Reformulate the skewness distribution. (Still trying to get this right). I think the distribution is not differentiable at some point, will need to check this.
- Removed the μ and σ_0 parameter.
- We know student-t is a special case of scale normal mixture, but under the discrete support point restriction the scale normal mixture is less parameter efficent.
- Should focus on showing mixture model works. Ignore items such as ARMA, and extending the GARCH(p, q).
- Can not install nspmix_1.0-1-2011-08-25.tgz from source. Gave me the following error.

```
mkao006@Asus-laptop:~/Downloads$ R CMD INSTALL nspmix_1.0-1-2011-08-25.tgz
* installing to library /home/mkao006/R/i686-pc-linux-gnu-library/2.14
* installing *source* package snpmle ...
** Creating default NAMESPACE file
** R
Error in .install_package_code_files(".", instdir) :
  unable to write code files
ERROR: unable to collate and parse R files for package snpmle
* removing /home/mkao006/R/i686-pc-linux-gnu-library/2.14/snpmle
* restoring previous /home/mkao006/R/i686-pc-linux-gnu-library/2.14/snpmle
```

4.1 Fixing the skewness distribution

The current formulation of the skewness distribution imposes a problem in estimation. The variance and the mean are both function of the skewness parameter (ξ) defined by the following function.

$$\mu_\xi = M_1(\xi - \frac{1}{\xi}),$$

$$\sigma_\xi^2 = (M_2 - M_1^2)(\xi^2 + \frac{1}{\xi^2}) + 2M_1^2 - M_2,$$

$$M_r = 2\int_0^\infty x^r f(x)dx,$$

Thus, we adopt the following formulation of a standardized skewed distribution used in fGarch and Lambert and Laurent [2001].

$$\begin{aligned} f^{\star}(z|\xi\theta) &= \frac{2\sigma}{\xi+\frac{1}{\xi}} f^{\star}(z_{\mu_\xi\sigma_\xi}|\theta) \\ z_{\mu_\xi\sigma_\xi} &= \xi^{\mathrm{sign}(\sigma_\xi z+\mu_\xi)}(\sigma_\xi z+\mu_\xi), \end{aligned}$$

4.2 The new distribution and its derivatives

Our mixture distribution has the following density function:

$$f(x_t;\boldsymbol{\theta},\boldsymbol{\pi},\boldsymbol{\beta}) = \sum_{i=1}^{m} \pi_i f(x_t;\theta_i,\boldsymbol{\beta})$$

However, I have removed σ_0^2 and μ from the β. The reason is that there is already an unbiased analytical solution ($\sigma_0^2 = \alpha_0/\beta_1$) already so it make no sense to estimate it. Furthermore, the mean is know to be stochastic and is usually estimated by a time series model and thus our previous formulation of estimating the mean with a constant is unsiotable.

The solution is to initialise σ_0^2 with the analytical solution and μ will be estimated seperate with a time-series model. GARCH model is a model for volatility estimation not for mean.

4.2.1 The density

We first calculate M_1 and M_2 for each component density that are required:

$$
\begin{aligned}
M_{1,i} &= 2\int_0^\infty x_t f(x_t;\theta_i,\boldsymbol{\beta})dx \\
&= 2\int_0^\infty \frac{x_t}{\sqrt{2\pi\theta_i^2\sigma_t^2}}\exp\frac{-x_t^2}{2\theta_i^2\sigma_t^2}dx \\
&= \sqrt{\frac{2}{\pi\theta_i^2\sigma_t^2}}\int_0^\infty x_t\exp\frac{-x_t^2}{2\theta_i^2\sigma_t^2}dx \\
&= \sqrt{\frac{2}{\pi\theta_i^2\sigma_t^2}}\left(\frac{\frac{1}{2}\Gamma(\frac{n+1}{2})}{a^{\frac{n+1}{2}}}\right) \\
&= \sqrt{\frac{2}{\pi}}\theta_i\sigma_t, \\
M_{1,i}^2 &= \frac{2}{\pi}\theta_i^2\sigma_t^2, \\
M_{2,i} &= 2\int_0^\infty x_t^2 f(x_t;\theta_i,\boldsymbol{\beta})dx \\
&= 2\int_0^\infty \frac{x_t^2}{\sqrt{2\pi\theta_i^2\sigma_t^2}}\exp\frac{-x_t^2}{2\theta_i^2\sigma_t^2}dx \\
&= \sqrt{\frac{2}{\pi\theta_i^2\sigma_t^2}}\int_0^\infty x_t^2\exp\frac{-x_t^2}{2\theta_i^2\sigma_t^2}dx \\
&= \sqrt{\frac{2}{\pi\theta_i^2\sigma_t^2}}\left(\frac{(2k-1)!!}{2^{k+1}a^k}\sqrt{\frac{\pi}{a}}\right) \\
&= \sqrt{\frac{2}{\pi\theta_i^2\sigma_t^2}}\left(\frac{\theta_i^2\sigma_t^2}{2}\sqrt{2\pi\theta_i^2\sigma_t^2}\right) \\
&= \theta_i^2\sigma_t^2
\end{aligned}
$$

Then we calculate the reparametised mean and variance:

$$
\begin{aligned}
\mu_{\xi_{i,t}} &= \sqrt{\frac{\pi}{2}}\theta_i\sigma_t(\xi - \xi^{-1}), \\
\sigma_{\xi_{i,t}} &= (\theta_i^2\sigma_t^2 - \frac{2}{\pi}\theta_i^2\sigma_t^2)(\xi^2 + \xi^{-2}) + \frac{4}{\pi}\theta_i^2\sigma_t^2 - \theta_i^2\sigma_t^2 \\
&= \theta_i^2\sigma_t^2[(1 - \frac{2}{\pi})(\xi^2 - \xi^{-2}) + (\frac{4}{\pi} - 1)] \\
\hat{z} &= (\sigma_{\xi_{i,t}} z_t + \mu_{\xi_{i,t}}) \\
z_{\mu_{\xi_{i,t}}\sigma_{\xi_{i,t}}} &= \hat{z}\xi^{-1}H(-\hat{z}) + \hat{z}\xi H(\hat{z})
\end{aligned}
$$

We then obtain the transformed density and log density as follow:

$$
\begin{aligned}
f(z_{\mu_{\xi_{i,t}}\sigma_{\xi_{i,t}}}|\xi) &= \frac{2\theta_i\sigma_t}{\xi + \xi^{-1}}\left(\frac{1}{\sqrt{2\pi\theta_i^2\sigma_t^2}}\exp\frac{-z^2_{\mu_{\xi_{i,t}}\sigma_{\xi_{i,t}}}}{2\theta_i^2\sigma_t^2}\right) \\
&= \frac{\sqrt{2}}{\sqrt{\pi}(\xi + \xi^{-1})}\exp\frac{-z^2_{\mu_{\xi_{i,t}}\sigma_{\xi_{i,t}}}}{2\theta_i^2\sigma_t^2} \\
\log f(z_{\mu_{\xi_{i,t}}\sigma_{\xi_{i,t}}}|\xi) &= \frac{1}{2}\log(2/\pi) - \log(\xi + \xi^{-1}) - \frac{z^2_{\mu_{\xi_{i,t}}\sigma_{\xi_{i,t}}}}{2\theta_i^2\sigma_t^2}
\end{aligned}
$$

4.2.2 The derivatives

Workings are not shown here but they should be correct, will check them with the numerical gradients later.

$$
\begin{aligned}
\frac{\partial \log f}{\partial \xi} &= \frac{(\xi^{-2}-1)}{(\xi+\xi^{-1})} \\
&+ \left\{ \theta_i^3\sigma_t^3[(1+\frac{2}{\pi})(1-\xi^{-4}) + (1-\frac{4}{\pi})\xi^{-2}]z_t + \sqrt{\frac{2^3}{\pi}}\theta_i^2\sigma_t^2\xi^{-3} \right\} H(-\hat{z}) \\
&+ \left\{ \theta_i^3\sigma_t^3[(1+\frac{2}{\pi})(\xi^2-\xi^{-2}) + (\frac{4}{\pi}-1)]z_t + \sqrt{\frac{2}{\pi}}\theta_i^2\sigma_t^2(\xi-\xi^{-1}) \right\} \xi^{-1}\delta(-\hat{z}) \\
&+ \left\{ \theta_i\sigma_t[(1+\frac{2}{\pi})(3\xi^2-\xi^{-2}) + (\frac{4}{\pi}-1)]z_t + \sqrt{\frac{2^3}{\pi}}\xi \right\} H(\hat{z}) \\
&+ \left\{ \theta_i\sigma_t[(1+\frac{2}{\pi})(\xi^2-\xi^{-2}) + (\frac{4}{\pi}-1)]z_t + \sqrt{\frac{2}{\pi}}(\xi-\xi^{-1}) \right\} \xi\delta(\hat{z}) \\
\frac{\partial \log f}{\partial \theta} &= \left\{ 3\theta_i^2\sigma_t^3[(1+\frac{2}{\pi})(\xi^2+\xi^{-2}) + (\frac{4}{\pi}-1)]z_t + \sqrt{\frac{2^3}{\pi}}\theta_i\sigma_t^2(\xi-\xi^{-1})) \right\} \xi^{-1}H(-\hat{z}) \\
&+ \left\{ \theta_i^3\sigma_t^3[(1+\frac{2}{\pi})(\xi^2-\xi^{-2}) + (\frac{4}{\pi}-1)]z_t + \sqrt{\frac{2}{\pi}}\theta_i^2\sigma_t^2(\xi-\xi^{-1}) \right\} \xi^{-1}\delta(-\hat{z}) \\
&+ \left\{ \sigma_t[(1+\frac{2}{\pi})(\xi^2-\xi^{-2}) + (\frac{4}{\pi}-1)]z_t \right\} \xi H(\hat{z}) \\
&+ \left\{ \theta_i\sigma_t[(1+\frac{2}{\pi})(\xi^2-\xi^{-2}) + (\frac{4}{\pi}-1)]z_t + \sqrt{\frac{2}{\pi}}(\xi-\xi^{-1}) \right\} \xi\delta(\hat{z}) \\
\frac{\partial \log f}{\partial \theta} &= \left\{ 3\theta_i^3\sigma_t^2[(1+\frac{2}{\pi})(\xi^2+\xi^{-2}) + (\frac{4}{\pi}-1)]z_t + \sqrt{\frac{2^3}{\pi}}\theta_i^2\sigma_t(\xi-\xi^{-1})) \right\} \xi^{-1}H(-\hat{z}) \\
&+ \left\{ \theta_i^3\sigma_t^3[(1+\frac{2}{\pi})(\xi^2-\xi^{-2}) + (\frac{4}{\pi}-1)]z_t + \sqrt{\frac{2}{\pi}}\theta_i^2\sigma_t^2(\xi-\xi^{-1}) \right\} \xi^{-1}\delta(-\hat{z}) \\
&+ \left\{ \theta_t[(1+\frac{2}{\pi})(\xi^2-\xi^{-2}) + (\frac{4}{\pi}-1)]z_t \right\} \xi H(\hat{z}) \\
&+ \left\{ \theta_i\sigma_t[(1+\frac{2}{\pi})(\xi^2-\xi^{-2}) + (\frac{4}{\pi}-1)]z_t + \sqrt{\frac{2}{\pi}}(\xi-\xi^{-1}) \right\} \xi\delta(\hat{z})
\end{aligned}
$$

After some thoughts, I have reformulate the problem and the temporary goal of the thesis. In order to demonstrate that the mixture normal distribution is a suitable model for the GARCH framework, we will short that under all circumstances and equal conditions that the mixture normal is a superior framework.

To achieve this we will first have the model to have estimation as close to the fGarch package, then we will show by adopting a mixture distribution, our fit was better and more flexible. We can already show that the two main distribution used in the fGarch package is simply a special case of the mixture normal, however the generalised error distribution can not be a special case of the mixture normal (or can it?)

What are are we going to do after this? Extend the model? What is our ultimate goal?

4.3 Thoughts

1. Check the relationship between different moments of a distribution. (e.g. the relationship of kurtosis and vairance as a function of skewness.)

2. Compare the shape and property of different skewness formulation.

3. Take the estimation of the mean out likelihood function, the reason being the estimation of the mean should be based on a stochastic process rather than a constant parameter. (We can try GARCH-in Mean). Review the literature on mean estimation of the GARCH model and also how the σ_0^2 is estimated.

4. Check how the fGarch package use ARIMA in estimating the mean, so that the same way can be applied to our formulation. This can also avoid possible breakdown as our timeseries is truely demeaned.

5. Think about how to integrate the cnm-ms frame work with the fGarch framework.

6. Talk to Adam about Cassava and other food prices.

7. Still not sure whether we should be estimating σ_0^2. (Do not estimate it, there is an unbiased estimate already based on the formula.)

8. Start writing out and draft the thesis, if is important at this stage to consider what is possible and what is not. Then formulate what we would like to achieve.

9. Review more literature on distribution and measure theory.

10. Should we consider restricting our model? So that we can use it for benchmark purposes.
11. We will use ARIMA to demean the series first for computation reasons. (But the problem is what order?)
12. The aim of the thesis is to explore the possibility and the potential of the mixture framework under the GARCH model. This was motivated by the large amount of literatures discussing the error distribution of the GARCH model.
13. Read through rugarch to give you some thoughts about what to do next.
14. Don't really want to reinvent the wheel. Ideally, I just want to develop an add-on so that people using the fGarch package can use this easily.
15. Check how the food price index is derived.
16. What is variance targeting mentioned in ugarch documentation.
17. rename the parameter, beta for the autoregressive coefficient, and alpha for the coefficient of the original time series.
18. We don't need to obtain an exact solution as in the fGarch or the rugarch package, since we are using different optimisatoin algorithm. However, the value obtained should be very close.
19. The Fernandez and Steel skewness parametisation is also used by the rugarch package.
20. Check the General Hyperbolic and the inverse normal gaussian distribution.
21. Check with Yong whether nspmix_1.0-1-2011-08-25.tgz is the newest version, and why I can't build the package when I can build the lsei package and all others from CRAN.
22. One reason to used the same skewness formulation is to enable us to benchmark the mixture distribution
23. Despite the fact that the student-t distribution is a special case of the scale normal mixture distribution, however, that is under the continuous case isn't it?

24. Sometimes the gradient is not zero and sometimes it is depending on the data. Is this a result of the constraint optimisation? But if we don't restrict it, the model explodes and creates error? what error?

5 17th January 2012

5.1 Thoughts

1. Giving up on assessing model based on log-likelihood and replicating everything from fGarch and rugarch. Will try to assess the model from a purely prediction perspective.

2. The log-likelihood is different even for the same model, have not thoroughly identify the reason but possibly due to the fact that we have included the first term.

3. Most of the run-time of the algorithem is actually spend on optimisation post scaling, rather than the logd.function or the cnm algorithm.

4. Reintroduced σ_0 into the parameter estimation solely for scaling reason.

5. Still not sure why there is difference between the numerical gradient and the analytical gradient (for the first and the fourth parameter). The size of deviation is actually inversly related to the size of the values of the data so this could be possibly related to numerical problem? This is also the problem which causes the algorithm to break down in all cases.

6. New codes implemented, what is the weight function?

7. Check the log-likelihood to be the same as fGarch

6 24th January 2012

1. The log-likelihood is the same whether or not the skewness is included.

2. Error in lsch(mix, beta, disc(mix$pt, sol), beta, x, which = c(1, 0, 0)) : Can not produce valid interior point in lsch()

3. Problem obtaining intra-day data for realised volatility.

4. Need to find a suitable minimum for the variance support space.

5. The numerical gradient is wrong I think because for some intitial values, it becomes zero.

6.1 Changes

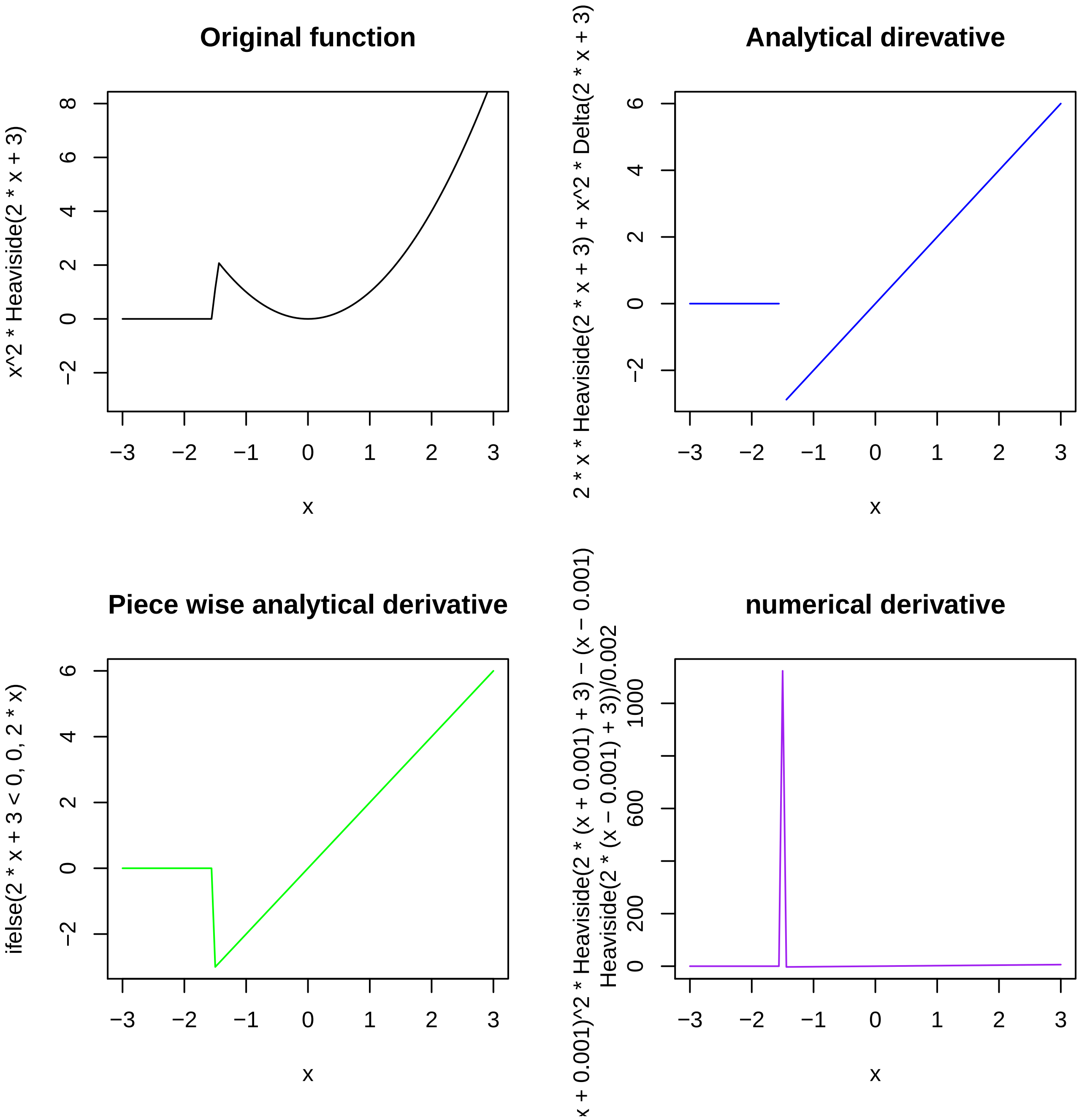
1. Check all three derivatives (numerical, symbolic, analytical)

2. Remove stationarity restriction on scaling

7 31st January 2012

The following graph shows that the numerical derivative is unstable close to the switching point. The piece wise and the analytical derivative are identical except the treatment at the switching point is different. I am still a bit skeptical about using piece wise analytic (this was what I actually used in the first and the current version). The reason is that the switching point is a function of the other betas, so that we have to take into account of that. Values that are close to the switching point will be incorrect if we don't take into account of this.

```
## Plot the Heaviside function and it's derivative
par(mfrow = c(2, 2))
## The original function
curve(x^2 * Heaviside(2 * x + 3), -3, 3, ylim = c(-3, 8),
      main = "Original function")
## The analytical derivative
curve(2 * x * Heaviside(2 * x + 3) + x^2 * Delta(2 * x + 3),  -3, 3,
      col = "blue", main = "Analytical direvative")
## The derivative if we treat it as piece-wise function
curve(ifelse(2 * x + 3 < 0, 0,  2 * x), -3, 3, col = "green",
      main = "Piece wise analytical derivative")
## The numerical derivative
curve(((x + 1e-3)^2 * Heaviside(2 * (x + 1e-3) + 3) -
      (x - 1e-3)^2 * Heaviside(2 * (x - 1e-3) + 3))/2e-3, -3, 3,
      col = "purple", main = "numerical derivative")
```



1.

The log-likelihood is defined as follow:

$$l = \log(\sigma_\xi) + \log(2) - \log(\xi + 1/\xi) - \log(\theta_i) - \log(\sigma_t) -$$
$$\frac{1}{2}\left((\sigma_\xi \sigma_t^{-1}\theta_i^{-1}x_t + \mu_\xi)(\xi^{-1}H(\sigma_\xi\sigma_t^{-1}\theta_i^{-1}x_t + \mu_\xi) + \xi H(-\sigma_\xi\sigma_t^{-1}\theta_i^{-1}x_t - \mu_\xi))\right)^2$$

And it's derivative with respect to σ_t is:

$$\begin{aligned}
\frac{\partial l}{\partial \sigma_t} = & -\frac{1}{\sigma_t} \\
& - \left((\sigma_\xi\sigma_t^{-1}\theta_i^{-1}x_t + \mu_\xi)(\xi^{-1} + \xi)H(\sigma_\xi\sigma_t^{-1}\theta_i^{-1}x_t + \mu_\xi)\right) \\
& \times \left([-(\sigma_\xi\sigma_t^{-2}\theta_i^{-1}x_t)\xi^{-1}H(\sigma_\xi\sigma_t^{-1}\theta_i^{-1}x_t + \mu_\xi)\right. \\
& + (\sigma_\xi\sigma_t^{-1}\theta_i^{-1}x_t + \mu_\xi)\xi^{-1}(\sigma_\xi\sigma_t^{-2}\theta_i^{-1}x_t)\delta(\sigma_\xi\sigma_t^{-1}\theta_i^{-1}x_t + \mu_\xi) \\
& + [-(\sigma_\xi\sigma_t^{-2}\theta_i^{-1}x_t)\xi H(-\sigma_\xi\sigma_t^{-1}\theta_i^{-1}x_t - \mu_\xi) \\
& + (\sigma_\xi\sigma_t^{-1}\theta_i^{-1}x_t + \mu_\xi)\xi(\sigma_\xi\sigma_t^{-2}\theta_i^{-1}x_t)\delta(-\sigma_\xi\sigma_t^{-1}\theta_i^{-1}x_t - \mu_\xi)] \qquad [\\
& = -\frac{1}{\sigma_t} + [z_\xi(\xi^{-1}H(z_\xi) + \xi H(-x_\xi))] \\
& \times [(\sigma_\xi\sigma_t^{-2}\theta_i^{-1}xt + \mu_\xi)(\xi^{-1}H(z_\xi) + \xi H(-x_\xi)) \\
& + (z_\xi\delta(z_\xi)(\xi - \xi^{-1})(\sigma_\xi\sigma_t^{-2}\theta_i^{-1}xt))] \\
& \approx -\frac{1}{\sigma_t} + [z_\xi(\xi^{-2}H(z_\xi) + \xi^2 H(-z_\xi))] \times (\sigma_\xi\sigma_t^{-2}\theta_i^{-1}xt)
\end{aligned}$$

which happens to be the same as the piece-wise derivative except at the switch point (Piece-wise is undefined at the switch point), the derivatives should be correct.

Given that high frequency data are no readily available, we will analyse time series like commodity price where high frequency data are not available and thus a different assessment approach will be required. Check how volatility of commodity prices are typically modelled/account for.

There is no doubt that the numerical gradient sufferes from cancellation error (in particular when both $beta_1$ and $alpha_1$ are small. Thus, can be severely inaccurate over certain intervals.

Despite the fact that the gradient are not zero, the algorithm has converged. The reason for this is due to the fact that the impact of the beta's are so small that it does not require the gradient to be zero for the log-likelihood to change. The requirement should be changed to sufficiently close to zero.

I think the model is working correctly. Things TODO before the next meeting:

- Write out the log-likelihood.

  - test the model with simulated data
  - plot the log-likelihood function w.r.t each β

- check the difference in the fit between include.mean = TRUE/FALSE within the garchFit function.

8 1st March 2012

Three things we have discussed and decide move towards.

- Check whether the solution obtain from increasing the line search iteration is different to the solution where it initially returned an error.
- Extend the GARCH(1, 1) model to GARCH(p, q)
- Run more simulation study to test the recovery rate/fit of the model.

In order to test whether the solution from the two setting is different, we will first find a time series in which the algorithm breaks (seed = 1), then obtain that solution to compare with the one that doesn't exit with an abnormally convergence code.

For the GARCH(p, q) model, it is simply having a σ_t which is a function of higher order term, check the fGarch documentation for reference of the filter representation. Check whether we can use the same representation as what's used in the fGarch package.

To assess how well the mixture fits the data, we will see how well the CDF of the mixture distribution follows the empirical distribution function of the data. (Really? is there a more objective way)?

9 8th March 2012

An Attempt was given to generalise the GARCH(1, 1) model to GARCH(p, q), however several obstacles was faced and would be extremely difficult to implement before the thesis deadline.

- Can not find an analytical expression for the general derivatives. Currently, there are only derivatives for the first order and they have been hard coded in the function. Unless we can find an expression/function which gives the derivatives of higher order we can only hard code maybe 20 of these derivatives in the function and use them when they are specified.

- Another problem is that when we increase the order of the conditional variance (p), we will also increase the number of initial variance estimated (i.e. σ_2, σ_3, etc ...). This also creates more problem when calculating the derivatives, since do we specify higher order terms as a function of σ_1 or σ_2? (Refering to 1.1 of page 1, we have currently wrote the function of σ_t as a function of α_0, α_1, β_1 and σ_1, how do we re-qrite this expression when we have σ_2, σ_3?)
- All the above obstacles can be potentially resolved if we use the numerical derivatives but the biggest problem is that the scaling would become extremely difficult if not impossible with several initial variance.

Given the time constraint, and the limited return my opinion is not to extend and carry on with just GARCH(1, 1).

The optimization adopted in the fGarch package does not require the derivative of the parameter and thus it does not face the same problem we have and therefore no precedent case for us to follow.

10 28th March 2012

Carrying out simulatino study based on different type of bell-shaped distribution with different type of sample size and show that the mixture normal distribution is a flexible frame work which is sufficiently capable of capturing most of the distribution regardless of the assumption.

The distributions in commonly used in financial time series:

- Normal distribution
- t-distribution
- Ged distribution
- Mixture normal distribution
- General hyperbolic distribution
- Cauchy distribution?
- Levy's alpha stable distribution

The deviation between the ECDF and the theoretical CDF and the Kolmogorov-Smirnov statistic is calculated to assess the fit.

simulation results

The sample size should be based on the typical length of certain time series. (e.g. Commodity 50 100, stock price 1000+ etc....)

50 samples

It appears that the t-distribution consistently outperforms all other distributions in terms of both mean and the maximum deviation regardless what the underlying distribution was.

100 samples

The same appears to be true for a sample size of 100.

500 samples

As the sample size increase to 500, the performance of the mixture normal distribution starts to catch up with the t-distribution. I suspect when the sample size are small, the characteristic of the underlying distribution can not be captured by the small sample and thus any distribution can have a descent fit but the reason why the t-distribution has a well fit is unknown.

1000 samples

As the sample size approaches 1000, the power of mixture normal becomes clear as there is sufficient data to capture the underlying distribution. Another supporting evidence is that the fit of the normal distribution deteriotes very rapidly. This can be supported by the Glivenko-cantelli theorem, in which the larger the sample size the ECDF converges almost surely to the theoretical CDF.

results

There seems to be strong evidence that the power of the mixture normal really depends on the sample size. The reason is not that the mixture normal is inefficient for small sample, rather it is the fact that small samples does not sufficiently capture the nature of the true underlying distribution.

Improved simulation study:

1. Obtain a large sample of price data (not restricted to stock market).
2. Fit Normal, t, ged, Levy, and other distribution to determine the likely range of the parameter space (e.g. df = 4   20 for t-distribution), and also the sample size.
3. Use these parametric distributions to generate samples.
4. Fit models to these samples and assess the fit.

Notes:

- Rewrite the mgarchSim function to include the other distributions and make sure they are comparable.
- Think how to initialise the simulated series.
- Need to defend on the mixture normal distribution is not over-fitting. Over-fitting when refers to a model capturing too much of the noise, however we want to capture as much of the noise as possible.