Dynamic Models for Volatility and Heavy Tails
by Andrew Harvey

Discussion by Gabriele Fiorentini
University of Florence
and
Rimini Centre for Economic Analysis (RCEA)

Frankfurt, 4-5 May 2012

I enjoyed the papers and I am looking forward to get the first draft of the entire book.

Thera are a lot of contributions in several directions.

I will concentrate on DCS specfications which were new to me.

The score of dynamic regression models: general case

$$y_t = \mu_t(\boldsymbol{\theta}) + \sigma_t(\boldsymbol{\theta})\varepsilon_t, \qquad \varepsilon_t | I_{t-1}; \boldsymbol{\theta}, \boldsymbol{\eta} \sim i.i.d.\, D(0, 1, \boldsymbol{\eta})$$

$$\begin{aligned} \mathbf{s}_{\theta t}(\theta, \eta) &= -\frac{1}{\sigma_t(\theta)} \frac{\partial \mu_t(\theta)}{\partial \theta} \frac{\partial \ln f\left[\varepsilon_t(\theta), \eta\right]}{\partial \varepsilon} \\ &\quad -\frac{1}{2\sigma_t^2(\theta)} \frac{\partial \sigma_t^2(\theta)}{\partial \theta} \left\{ \frac{\partial \ln f\left[\varepsilon_t(\theta), \eta\right]}{\partial \varepsilon} \varepsilon_t(\theta) - 1 \right\} \end{aligned}$$

$$\begin{aligned} \mathbf{s}_{\theta t}(\theta, \eta) &= \frac{1}{\sigma_t(\theta)} \frac{\partial \mu_t(\theta)}{\partial \theta} \delta_t\left[\varepsilon_t(\theta), \eta\right] \varepsilon_t(\theta) \\ &\quad +\frac{1}{2\sigma_t^2(\theta)} \frac{\partial \sigma_t^2(\theta)}{\partial \theta} \left\{ \delta_t\left[\varepsilon_t(\theta), \eta\right] \varepsilon_t^2(\theta) - 1 \right\} \end{aligned}$$

$\delta_t\left[\varepsilon_t(\theta), \eta\right]$ depends on the assumed distribution

The score of dynamic regression models: general case

$$y_t = \mu_t(\boldsymbol{\theta}) + \sigma_t(\boldsymbol{\theta})\varepsilon_t, \qquad \varepsilon_t | I_{t-1}; \boldsymbol{\theta}, \boldsymbol{\eta} \sim i.i.d.\, D(0, 1, \boldsymbol{\eta})$$

$$\begin{aligned} \mathbf{s}_{\boldsymbol{\theta} t}(\boldsymbol{\theta}, \boldsymbol{\eta}) &= -\frac{1}{\sigma_t(\boldsymbol{\theta})} \frac{\partial \mu_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \ln f\left[\varepsilon_t(\boldsymbol{\theta}), \boldsymbol{\eta}\right]}{\partial \varepsilon} \\ &\quad - \frac{1}{2\sigma_t^2(\boldsymbol{\theta})} \frac{\partial \sigma_t^2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \left\{ \frac{\partial \ln f\left[\varepsilon_t(\boldsymbol{\theta}), \boldsymbol{\eta}\right]}{\partial \varepsilon} \varepsilon_t(\boldsymbol{\theta}) - 1 \right\} \end{aligned}$$

$$\begin{aligned} \mathbf{s}_{\theta t}(\theta, \eta) &= \frac{1}{\sigma_t(\theta)} \frac{\partial \mu_t(\theta)}{\partial \theta} \delta_t\left[\varepsilon_t(\theta), \eta\right] \varepsilon_t(\theta) \\ &\quad + \frac{1}{2\sigma_t^2(\theta)} \frac{\partial \sigma_t^2(\theta)}{\partial \theta} \left\{ \delta_t\left[\varepsilon_t(\theta), \eta\right] \varepsilon_t^2(\theta) - 1 \right\} \end{aligned}$$

$\delta_t\left[\varepsilon_t(\theta), \eta\right]$ depends on the assumed distribution

The score of dynamic regression models: general case

$$y_t = \mu_t(\boldsymbol{\theta}) + \sigma_t(\boldsymbol{\theta})\varepsilon_t, \qquad \varepsilon_t | I_{t-1}; \boldsymbol{\theta}, \boldsymbol{\eta} \sim i.i.d.\, D(0, 1, \boldsymbol{\eta})$$

$$\begin{aligned} \mathbf{s}_{\theta t}(\boldsymbol{\theta}, \boldsymbol{\eta}) &= -\frac{1}{\sigma_t(\boldsymbol{\theta})} \frac{\partial \mu_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \ln f\left[\varepsilon_t(\boldsymbol{\theta}), \boldsymbol{\eta}\right]}{\partial \varepsilon} \\ &\quad -\frac{1}{2\sigma_t^2(\boldsymbol{\theta})} \frac{\partial \sigma_t^2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \left\{ \frac{\partial \ln f\left[\varepsilon_t(\boldsymbol{\theta}), \boldsymbol{\eta}\right]}{\partial \varepsilon} \varepsilon_t(\boldsymbol{\theta}) - 1 \right\} \end{aligned}$$

$$\begin{aligned} \mathbf{s}_{\theta t}(\boldsymbol{\theta}, \boldsymbol{\eta}) &= \frac{1}{\sigma_t(\boldsymbol{\theta})} \frac{\partial \mu_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \delta_t\left[\varepsilon_t(\boldsymbol{\theta}), \boldsymbol{\eta}\right] \varepsilon_t(\boldsymbol{\theta}) \\ &\quad + \frac{1}{2\sigma_t^2(\boldsymbol{\theta})} \frac{\partial \sigma_t^2(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \left\{ \delta_t\left[\varepsilon_t(\boldsymbol{\theta}), \boldsymbol{\eta}\right] \varepsilon_t^2(\boldsymbol{\theta}) - 1 \right\} \end{aligned}$$

$\delta_t\left[\varepsilon_t(\boldsymbol{\theta}), \boldsymbol{\eta}\right]$ depends on the assumed distribution

The score of dynamic regression models: particular cases

Gaussian: $\delta_t\left[\varepsilon_t(\boldsymbol{\theta}), \boldsymbol{\eta}\right] = 1$

$$
\begin{aligned}
\delta_t\left[\varepsilon_t(\boldsymbol{\theta}), \boldsymbol{\eta}\right]\varepsilon_t(\boldsymbol{\theta}) &= \varepsilon_t(\boldsymbol{\theta}) \\
\delta_t\left[\varepsilon_t(\boldsymbol{\theta}), \boldsymbol{\eta}\right]\varepsilon_t^2(\boldsymbol{\theta}) - 1 &= \varepsilon_t^2(\boldsymbol{\theta}) - 1
\end{aligned}
$$

This expressions motivate the use of PML estimators which remain consistent *(un.su.re. conditions)* for $\boldsymbol{\theta}$ even if the true distribution is unknown.
PML forms the basis for semiparametric and sequential estimation procedures.

Standardized Laplace:

$$
\begin{aligned}
\delta_t\left[\varepsilon_t(\theta), \eta\right]\varepsilon_t(\theta) &= \sqrt{2} sign(\varepsilon_t(\theta)) \\
\delta_t\left[\varepsilon_t(\theta), \eta\right]\varepsilon_t^2(\theta) - 1 &= \sqrt{2}|\varepsilon_t(\theta)| - 1
\end{aligned}
$$

The score of dynamic regression models: particular cases

Gaussian: $\delta_t\left[\varepsilon_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]=1$

$$\begin{aligned}\delta_t\left[\varepsilon_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]\varepsilon_t(\boldsymbol{\theta}) &= \varepsilon_t(\boldsymbol{\theta})\\ \delta_t\left[\varepsilon_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]\varepsilon_t^2(\boldsymbol{\theta})-1 &= \varepsilon_t^2(\boldsymbol{\theta})-1\end{aligned}$$

This expressions motivate the use of PML estimators which remain consistent *(un.su.re. conditions)* for $\boldsymbol{\theta}$ even if the true distribution is unknown.
PML forms the basis for semiparametric and sequential estimation procedures.

Standardized Laplace:

$$\begin{aligned}\delta_t\left[\varepsilon_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]\varepsilon_t(\boldsymbol{\theta}) &= \sqrt{2}sign(\varepsilon_t(\boldsymbol{\theta}))\\ \delta_t\left[\varepsilon_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]\varepsilon_t^2(\boldsymbol{\theta})-1 &= \sqrt{2}|\varepsilon_t(\boldsymbol{\theta})|-1\end{aligned}$$

The score of dynamic regression models: particular cases

Standardized Student t with1$/\eta$ dgf: $\delta_t\left[\varepsilon_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right] = \frac{\eta+1}{1-2\eta+\eta\varepsilon^2(\boldsymbol{\theta})}$

$$\begin{aligned}
\delta_t\left[\varepsilon_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]\varepsilon_t(\boldsymbol{\theta}) &= \frac{\eta+1}{1-2\eta+\eta\varepsilon^2(\boldsymbol{\theta})}\varepsilon_t(\boldsymbol{\theta}) \\
\delta_t\left[\varepsilon_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]\varepsilon_t^2(\boldsymbol{\theta}) - 1 &= \frac{\eta+1}{1-2\eta+\eta\varepsilon^2(\boldsymbol{\theta})}\varepsilon_t^2(\boldsymbol{\theta}) - 1
\end{aligned}$$

Andrew's Beta u_t

Discrete Location Scale Mixture of Normals:
$\delta_t\left[\varepsilon_t(\theta),\eta\right] = \text{ugly}^*$

Gives an example of the score factors when the conditional distribution is asymmetric

$*$ see FS(2010) "New testing approaches for mean-variance predictability"

The score of dynamic regression models: particular cases

Standardized Student t with1/η dgf: $\delta_t\left[\varepsilon_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right] = \frac{\eta+1}{1-2\eta+\eta\varepsilon^2(\boldsymbol{\theta})}$

$$\begin{aligned}
\delta_t\left[\varepsilon_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]\varepsilon_t(\boldsymbol{\theta}) &= \frac{\eta+1}{1-2\eta+\eta\varepsilon^2(\boldsymbol{\theta})}\varepsilon_t(\boldsymbol{\theta}) \\
\delta_t\left[\varepsilon_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right]\varepsilon_t^2(\boldsymbol{\theta}) - 1 &= \frac{\eta+1}{1-2\eta+\eta\varepsilon^2(\boldsymbol{\theta})}\varepsilon_t^2(\boldsymbol{\theta}) - 1
\end{aligned}$$

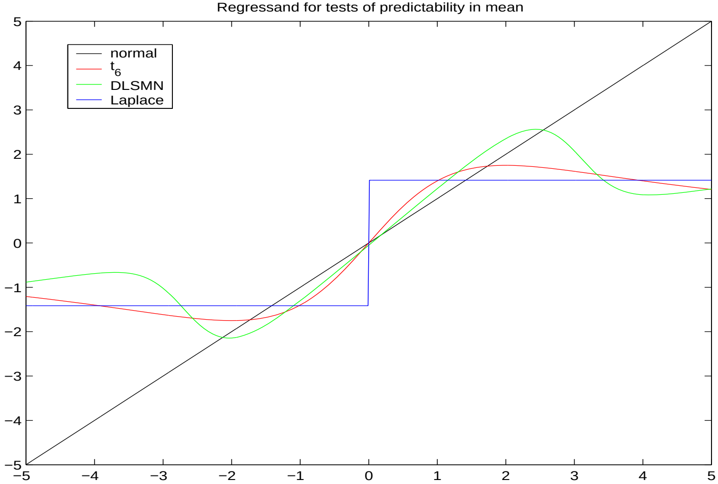
Andrew's Beta u_t

Discrete Location Scale Mixture of Normals:
$\delta_t\left[\varepsilon_t(\boldsymbol{\theta}),\boldsymbol{\eta}\right] = \text{ugly}^*$

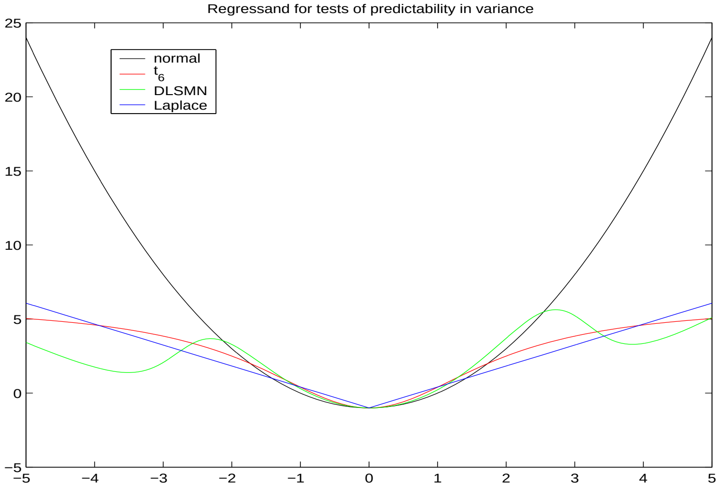
Gives an example of the score factors when the conditional distribution is asymmetric

$*$ see FS(2010) "New testing approaches for mean-variance predictability"

Source: FS(2010)



Source: FS(2010)



The damping factor $\delta_t[\varepsilon_t(\boldsymbol{\theta}), \boldsymbol{\eta}]$ and DCS models

$\delta_t[\varepsilon_t(\boldsymbol{\theta}), \boldsymbol{\eta}]$ can be regarded as a damping factor that accounts for the skewness and kurtosis of the conditional distribution, as in the robust estimation literature.

For example the Student t factor downweights big observations (outliers) when computing the average score.

DCS models take a step further and include the score factor in the filters of observation driven models. This choice is well motivated and it is argued that it has several advantages.

For example, in the DCS Garch model the conditional variance does not react abruptly to additive outliers.

However, in the DCS model we give up the possibility to partition model parameters into *dynamic* and *shape* parameters.

The damping factor $\delta_t\left[\varepsilon_t(\boldsymbol{\theta}), \boldsymbol{\eta}\right]$ and DCS models

$\delta_t\left[\varepsilon_t(\boldsymbol{\theta}), \boldsymbol{\eta}\right]$ can be regarded as a damping factor that accounts for the skewness and kurtosis of the conditional distribution, as in the robust estimation literature.

For example the Student t factor downweights big observations (outliers) when computing the average score.

DCS models take a step further and include the score factor in the filters of observation driven models. This choice is well motivated and it is argued that it has several advantages.

For example, in the DCS Garch model the conditional variance does not react abruptly to additive outliers.

However, in the DCS model we give up the possibility to partition model parameters into *dynamic* and *shape* parameters.

The damping factor $\delta_t\left[\varepsilon_t(\boldsymbol{\theta}), \boldsymbol{\eta}\right]$ and DCS models

$\delta_t\left[\varepsilon_t(\boldsymbol{\theta}), \boldsymbol{\eta}\right]$ can be regarded as a damping factor that accounts for the skewness and kurtosis of the conditional distribution, as in the robust estimation literature.

For example the Student t factor downweights big observations (outliers) when computing the average score.

DCS models take a step further and include the score factor in the filters of observation driven models. This choice is well motivated and it is argued that it has several advantages.

For example, in the DCS Garch model the conditional variance does not react abruptly to additive outliers.

However, in the DCS model we give up the possibility to partition model parameters into *dynamic* and *shape* parameters.

QUESTIONS

Isn't this close link between the shape of the conditional distribution and the dynamics of conditional moments too binding. Or, maybe, it is desirable to have it and offers clear advantages.

How could we define a PML estimator that remains CAN even if the conditional distribution is misspecified.

How could we define an efficient semiparametric estimator.

Could we estimate the model parameters sequentially. There are many case in which sequential estimation is a reasonable strategy.

What happens if we fix η in the filter. PML is certainly feasible and we still maintain a flexible nonlinear specification.

QUESTIONS

Isn't this close link between the shape of the conditional distribution and the dynamics of conditional moments too binding. Or, maybe, it is desirable to have it and offers clear advantages.

How could we define a PML estimator that remains CAN even if the conditional distribution is misspecified.

How could we define an efficient semiparametric estimator.

Could we estimate the model parameters sequentially. There are many case in which sequential estimation is a reasonable strategy.

What happens if we fix η in the filter. PML is certainly feasible and we still maintain a flexible nonlinear specification.

QUESTIONS

Isn't this close link between the shape of the conditional distribution and the dynamics of conditional moments too binding. Or, maybe, it is desirable to have it and offers clear advantages.

How could we define a PML estimator that remains CAN even if the conditional distribution is misspecified.

How could we define an efficient semiparametric estimator.

Could we estimate the model parameters sequentially. There are many case in which sequential estimation is a reasonable strategy.

What happens if we fix η in the filter. PML is certainly feasible and we still maintain a flexible nonlinear specification.

QUESTIONS

Isn't this close link between the shape of the conditional distribution and the dynamics of conditional moments too binding. Or, maybe, it is desirable to have it and offers clear advantages.

How could we define a PML estimator that remains CAN even if the conditional distribution is misspecified.

How could we define an efficient semiparametric estimator.

Could we estimate the model parameters sequentially. There are many case in which sequential estimation is a reasonable strategy.

What happens if we fix η in the filter. PML is certainly feasible and we still maintain a flexible nonlinear specification.

QUESTIONS

Isn't this close link between the shape of the conditional distribution and the dynamics of conditional moments too binding. Or, maybe, it is desirable to have it and offers clear advantages.

How could we define a PML estimator that remains CAN even if the conditional distribution is misspecified.

How could we define an efficient semiparametric estimator.

Could we estimate the model parameters sequentially. There are many case in which sequential estimation is a reasonable strategy.

What happens if we fix η in the filter. PML is certainly feasible and we still maintain a flexible nonlinear specification.

QUESTIONS

How do we test the null of Gaussian innovations against, say, Student t

$$E\left[\frac{3}{4}-\frac{3}{2}\varepsilon_t^2(\boldsymbol{\theta})+\frac{1}{4}\varepsilon_t^4(\boldsymbol{\theta})\right]=0$$

OUTLIERS: If an additive outlier is recognized and modelled properly (i.e. REGARIMA style) than it causes no problems. A nice example is Baille and Bollerslev (1989). If instead a "big observation" is a realization from the heavy tail conditional distribution than the GARCH variance may react too much and take a long time to go back the its average level. DCS should help by dumping the effect of big observations on the conditional variance.

(I couldn't find Harvey and Luati)

QUESTIONS

How do we test the null of Gaussian innovations against, say, Student t

$$E\left[\frac{3}{4} - \frac{3}{2}\varepsilon_t^2(\boldsymbol{\theta}) + \frac{1}{4}\varepsilon_t^4(\boldsymbol{\theta})\right] = 0$$

OUTLIERS: If an additive outlier is recognized and modelled properly (i.e. REGARIMA style) than it causes no problems. A nice example is Baille and Bollerslev (1989). If instead a "big observation" is a realization from the heavy tail conditional distribution than the GARCH variance may react too much and take a long time to go back the its average level. DCS should help by dumping the effect of big observations on the conditional variance.

(I couldn't find Harvey and Luati)

Source: FSC(2003)

Table 1. Estimates of a Conditionally Heteroscedastic Single-Factor Model for 26 U.K. Sectorial Indices. Monthly Excess Returns 1971:2–1990:10 (237 observations). Estimates of Dynamic Variance Parameters and Degrees of Freedom

$$\lambda_t = (1-\alpha-\beta)(1-\varrho^2) + \alpha[(f_{t-1|t-1} - \sqrt{(1-\alpha-\beta)/\alpha\varrho})^2 + \omega_{t-1|t-1}] + \beta\lambda_{t-1}, 0 \le \beta \le 1-\alpha \le 1, -1 \le \varrho \le 1$$

	Gaussian		Student t	
Parameter		SE		SE
α	.111	.075	.053	.026
β	.670	.258	.675	.120
ϱ	.951	.629	1.0	
η	0		.103	.012
Log-likelihood	−4,471.216		−4,221.162	

NOTE: $f_{t|t}$ denotes the Kalman filter estimate of the latent factor, and $\omega_{t|t}$ denotes the associated conditional mean squared error (Harvey et al. 1992). Standard errors (SE) are computed using analytical derivatives based on the expressions of Bollerslev and Wooldridge (1992) in the Gaussian case, and Proposition 1 in the case of the t.

term $\omega_{t-1|t-1}$ is included to reflect the uncertainty in the factor estimates (Harvey et al. 1992). We solved the usual scale indeterminacy of the factor by fixing $E(\lambda_t) = 1$. To do so, we set $\vartheta = (1-\alpha-\beta) - \alpha\upsilon$ and $\upsilon = \sqrt{(1-\alpha-\beta)/\alpha\varrho}$

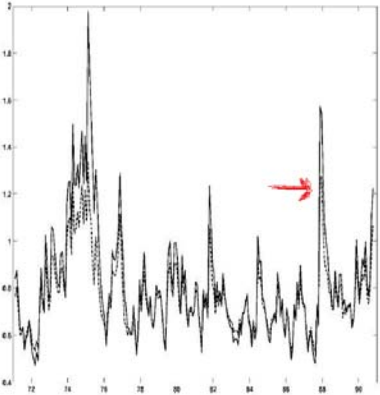


Figure 6. Estimated Conditional Standard Deviation of Equally Weighted Portfolio (——, Gaussian; - - -, Student t).

Testing for ARCH effects

In FS(2010) we develop LM predictability test based on more realistic distributions than the gaussian.

In a nutshell, we are testing whether past observations can predict the MD components of the score of the assumed conditional distribution.

This is the reason why we had figures similar to yours. We plot the regressands in the $T \times R^2$ form of the LM tests.

Importantly those tests have the correct size even under misspefication and aruguably more power than the normal based one.

Testing for ARCH effects

In FS(2010) we develop LM predictability test based on more realistic distributions than the gaussian.

In a nutshell, we are testing whether past observations can predict the MD components of the score of the assumed conditional distribution.

This is the reason why we had figures similar to yours. We plot the regressands in the $T \times R^2$ form of the LM tests.

Importantly those tests have the correct size even under misspefication and aruguably more power than the normal based one.

I considered an LM test for ARCH effects in the DCS specification

$$y_t = \pi_0 + \sigma_t(\boldsymbol{\theta})\varepsilon_t$$

$$\varepsilon_t | I_{t-1}; \pi, \alpha, \omega, \boldsymbol{\eta} \sim i.i.d.\ D(0, 1, \boldsymbol{\eta}) \text{ with density function } f(., \boldsymbol{\eta})$$

$$\sigma_t^2(\boldsymbol{\theta}) = \omega(1-\alpha) + \alpha\delta_t\left[\varepsilon_t(\boldsymbol{\theta}), \boldsymbol{\eta}\right](y_{t-1} - \pi)^2$$

The algebra of the LM test works out nicely because the link disappears under the null. The information matrix under the null is block diagonal with

$$I_{\alpha,\alpha}(\phi) = \frac{M_{ss}^2(\eta)}{4} \quad \text{for Student } t \ I_{\alpha,\alpha}(\phi) = \left(\frac{\nu}{\nu+3}\right)^2$$

We are actually testing if Andrew's u_{t-1} can predict Andrew's u_t

The test should have the correct size even under incorrect distributional specification.

I considered an LM test for ARCH effects in the DCS specification

$$y_t = \pi_0 + \sigma_t(\boldsymbol{\theta})\varepsilon_t$$

$$\varepsilon_t | I_{t-1}; \pi, \alpha, \omega, \boldsymbol{\eta} \sim i.i.d.\ D(0, 1, \boldsymbol{\eta}) \text{ with density function } f(., \boldsymbol{\eta})$$

$$\sigma_t^2(\boldsymbol{\theta}) = \omega(1-\alpha) + \alpha\delta_t\left[\varepsilon_t(\boldsymbol{\theta}), \boldsymbol{\eta}\right](y_{t-1} - \pi)^2$$

The algebra of the LM test works out nicely because the link disappears under the null. The information matrix under the null is block diagonal with

$$I_{\alpha,\alpha}(\phi) = \frac{M_{ss}^2(\eta)}{4} \quad \text{for Student } t \ I_{\alpha,\alpha}(\phi) = \left(\frac{\nu}{\nu+3}\right)^2$$

We are actually testing if Andrew's u_{t-1} can predict Andrew's u_t

The test should have the correct size even under incorrect distributional specification.

I considered an LM test for ARCH effects in the DCS specification

$$y_t = \pi_0 + \sigma_t(\boldsymbol{\theta})\varepsilon_t$$

$$\varepsilon_t | I_{t-1}; \pi, \alpha, \omega, \boldsymbol{\eta} \sim i.i.d.\ D(0, 1, \boldsymbol{\eta}) \text{ with density function } f(., \boldsymbol{\eta})$$

$$\sigma_t^2(\boldsymbol{\theta}) = \omega(1-\alpha) + \alpha\delta_t\left[\varepsilon_t(\boldsymbol{\theta}), \boldsymbol{\eta}\right](y_{t-1} - \pi)^2$$

The algebra of the LM test works out nicely because the link disappears under the null. The information matrix under the null is block diagonal with

$$I_{\alpha,\alpha}(\phi) = \frac{M_{ss}^2(\eta)}{4} \quad \text{for Student } t \ I_{\alpha,\alpha}(\phi) = \left(\frac{\nu}{\nu+3}\right)^2$$

We are actually testing if Andrew's u_{t-1} can predict Andrew's u_t

The test should have the correct size even under incorrect distributional specification.

Latent GARCH processes

I was surprised not to see any discussion about latent GARCH processes.

A proper latent GARCH is actually an SV model and thus not observation driven.

If suitably specified it becomes observation driven but the analysis is difficult (HRS, 1992).