Generalized Autoregressive Score Models

Leopoldo Catania
Aarhus University and CREATES
leopoldo.catania@econ.au.dk

Time Varying Parameter models

Time varying parameter models are widely used for filtering and prediction of quantities of interest in finance such as:

- The volatility of a stock
- The number of trades in a given time period
- The waiting time before the next trade of a stock
- Changes in the risk level of a financial instrument

What we usually do is:

1) Formulate a parametric model which describes the evolution of the underlying stochastic process.
2) Estimate the model
3) Draw our conclusions

Today we focus on 1).

Types of Parametric Models

Sir David Cox in a very influential classical paper (Cox, 1981) classified dynamic models in two categories:

- Parameter driven models
- Observation driven models

Let ψ_t be the parameter of interest at time t, and let Y_t be an observed random variable at time t, define $\mathbf{Y}_{1:t-1} = (Y_{t-1}, Y_{t-2}, \dots)$. Even though Cox used a different formulation, nowadays we say that:

- When $\psi_t = \psi(g(\mathbf{Y}_{1:t-1})$ that is ψ_t is known with certainty given $\mathcal{F}_{t-1} = \sigma(\mathbf{Y}_{1:t-1})$, we have an observation driven model.

- When $\psi_t = \psi^\dagger(\psi_{t-1}, \eta_t^\dagger)$ and $\eta_t^\dagger$ is not known with certainty given $\mathcal{F}_{t-1}$ we have a parameter driven model.

Types of Parametric Models

- Example of observation driven models are:
  - GARCH models (Engle, 1982; Bollerslev, 1987)
  - ACD models (Engle and Russell, 1998)
  - Multiplicative Error Models (Engle and Gallo, 2006)
  - Poisson Autoregression (Fokianos et al., 2009)
  - Time Varying Copulas (Patton, 2006)
  - . . .

- Example of parameter driven models are:
  - SV models (Taylor, 1986)
  - DSGE models (DeJong and Dave, 2011)
  - Dynamic Linear Models (West and Harrison, 1997)
  - Hidden Markov Models (Frühwirth-Schnatter, 2006)
  - Stochastic Copula Models (Hafner and Manner, 2012)
  - General Non–linear Non–Gaussian State Space models (Durbin and Koopman, 2012).
  - . . .

Types of Parametric Models

Most of the times we end up with the following characteristics for the two class of models:

- Observation driven models:
  - The conditional density $p(y_t|\mathbf{y}_{1:t-1})$, is usually available in closed form:
    a) The likelihood is available via the prediction error decomposition.
    b) One step ahead predictions are easy to compute.
  - Model tractability is easy.
  - Frequentist inference.

- Parameter driven models:
  - The conditional density $p(y_t|\mathbf{y}_{1:t-1})$, is available after integration of the latent parameter $p(y_t|\mathbf{y}_{1:t-1}) = \int_\Psi p(y_t, \psi_t|\mathbf{y}_{1:t-1})\mathrm{d}\psi_t$:
    a) The likelihood is difficult to obtain and sometimes needs to be simulated.
    b) Predictions are generally available only via simulation.
  - Model tractability is difficult.
  - Bayesian inference.

Observation driven models

The usual critiques to observations driven models are:

- They are too simplistic.
- Do not adequately describe the real properties of the underlying true data generating process.
- Assume that parameters at time t are deterministic given $\mathcal{F}_{t-1}$.

The general problem is that the filter constructed for the parameter(s) of interest most of the times lack of any statistical reasoning.

Assume that we have a good reason to specify $\psi_t = c_t + g(y_{t-1})$, where c_t is $\mathcal{F}_{t-1}$–measurable. The econometrician faces the problem: How to select $g(\cdot)$?. The answer is generally: Lets try: $y_{t-1}, y_{t-1}^2, |y_{t-1}|, \ldots$, and pick the one which allows me to: a) derive some theory, b) find reasonable filtered values for ψ_t.

Observation driven models

Of course the econometrician does not only try different parameterization of $g(y_{t-1})$, sometimes she searches for a "correct" functional which satisfies some moment conditions, for example $\mathbb{E}[g(Y_t)] = \mathbb{E}[\psi_t]$.

Consider for example the GARCH(1,1) model:

$$y_t = \sigma_t \varepsilon_t, \quad \varepsilon_t \sim \mathcal{D}(0, 1, \boldsymbol{\theta})$$
$$\sigma_t^2 = \omega + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2.$$

In this case $\mathbb{E}[Y_t^2] = \mathbb{E}[\sigma_t^2] = \sigma^2$ (provided that it exists).

Observation driven models

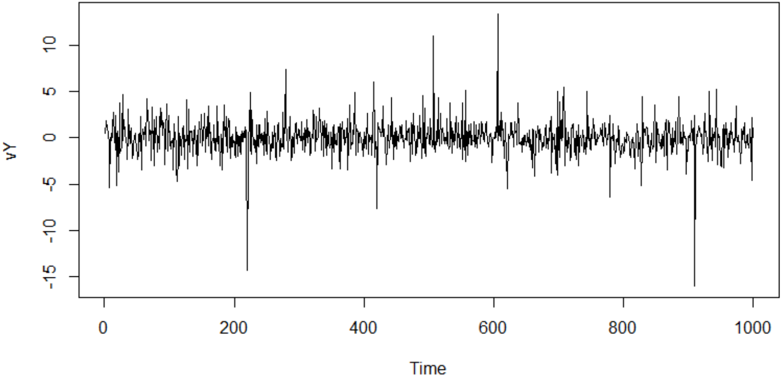
Few questions arises in this context:

- What if $\mathbb{E}[Y_t^2]$ is not defined?

- Is the *forcing variable* y_t^2 always adequate for σ_t^2? Is it always true that a large $|y_t|$ is signal of increased conditional variance?

- What if instead of the conditional variance σ_t^2 we have, say, the skewness parameter of a Skew–Student's t distribution? Which is in this case the "correct" forcing variable?

Observation driven models



One thousand iid draws from a Student's t distribution with $\nu = 3$

Generalized Autoregressive Score models

The class of Generalized Autoregressive Score (GAS) models provides coherent statistical answers to these questions.

GAS has been developed by Creal et al. (2013) and Harvey (2013). Harvey has named this class of models: Dynamic Conditional Score (DCS). He writes:

"Rather than the term dynamic conditional score (DCS) models, ..., Creal, Koopman and Lucas prefer the name generalized autoregressive score (GAS). However, despite the attraction of the acronym, the term "autoregresive" seems to me to convey a more limited dynamic structure than is actually the case.", (Harvey, 2013).

The first articles about GAS are: Creal et al. (2008) and Harvey and Chakravarty (2008). The first use of the score to robustify the Kalman filter was originally proposed by Masreliez (1975).

Intermezzo: The Newton-Raphson method for root finding

Suppose the function f is differentiable with continuous derivative f' and a root a.

Let $x_0 \in \Re$ our guess for a. Now the straight line through the point $(x_0, f(x_0))$ with slope $f'(x_0)$ is the best straight line approximation to the function $f(x)$ at the point x_0. The equation of this straight line is given by:

$$y = f(x_0) + f'(x_0)(x - x_0) \tag{1}$$

This line crosses the x–axis at a point x_1, which is a better approximation than x_0 to a. We find x_1 solving (1) for x with $y = 0$:

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)},$$

i.e., we update our guess x_0 by a factor $\frac{f(x_0)}{f'(x_0)}$ which gives the direction to the root of the function.

Intermezzo: The Newton-Raphson method for root finding

If we iterate this procedure we end up with the following algorithm:

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)},$$

Under conditions on f, f' and f'' it is possible to show that $x_{n+1} \to a$.

Since we are expecting $f(x_n) \to 0$, a good stopping condition for the Newton-Raphson algorithm is $|f(x_n)| \leq \varepsilon$ for some tolerance ε.

Intermezzo: The Newton-Raphson method for optimization

Since we deal with a problem of optimization we have $f'(x) = 0$. The Newton-Raphson method for optimization replaces $f(x)$ with $f'(x)$. Hence, the k-th Newton-Raphson step is

$$x_k = x_{k-1} - \frac{f'(x_{k-1})}{f''(x^{k-1})} \tag{2}$$

If $f(\cdot)$ is function of more than one variable, the step becomes

$$x_k = x_{k-1} - H^{-1}(x_{k-1})\nabla f(x_{k-1}) \tag{3}$$

where $H^{-1}(x_{k-1})$ is the inverse of the Hessian evaluated in x_{k-1} and $\nabla f(x_{k-1})$ is a column-vector.

The NR method is such that the convergence is *local* and *quadratic*. Near the solution, the convergence is very fast! The NR method is really suitable when the first and second-order information are readily and easily calculated.

Generalized Autoregressive Score Models

Assume to observe the time–series $\mathbf{y}_{1:T} = (y_1, \ldots, y_T)$, its joint density is given by:

$$p(\mathbf{y}_{1:T}; \psi) = p(y_1; \psi) \prod_{t=2}^{T} p(y_t | \mathbf{y}_{1:t-1}; \psi) \tag{4}$$

and the log likelihood for ψ reads:

$$\begin{aligned} \mathcal{L}(\psi; \mathbf{y}_{1:T}) &= \log p(y_1; \psi) + \sum_{t=2}^{T} \log p(y_t | \mathbf{y}_{1:t-1}; \psi) && (5) \\ &= \sum_{t=1}^{T} \ell_t, && (6) \end{aligned}$$

hence, the log likelihood contribution of observation y_t is:

$$\ell_t = \log p(y_t | \mathbf{y}_{1:t-1}; \psi)$$

Generalized Autoregressive Score Models

Assume to specify an observation driven model for the time–series $\mathbf{y}_{1:T}$ parameterized in terms of the dynamic parameter $\psi_t = \psi(\mathbf{y}_{1:t-1})$.

GAS implements this filter as:

$$\begin{aligned}\psi_t &= \psi(\mathbf{y}_{1:t-1}) \\ &= \omega + \alpha u_{t-1} + \beta\psi_{t-1},\end{aligned}$$

where $u_t = S_t \nabla_t$ and:

$$\begin{aligned}\nabla_t &= \frac{\partial \log p(y_t|\mathbf{y}_{1:t-1};\psi)}{\partial\psi_t} = \frac{\partial\ell_t}{\partial\psi_t} \\ S_t &= \mathcal{I}_t^{-1} = E_{t-1}[\nabla_t^2]^{-1}\end{aligned}$$

Generalized Autoregressive Score Models

The quantity $\nabla_t = \nabla(y_t; \psi_t)$ is the score of the conditional distribution $p(y_t|\mathbf{y}_{1:t-1}; \psi_t)$ and, as in the Newton–Raphson algorithm, gives the direction of the update.

The quantity $S_t = \mathcal{I}_t^{-1} = \mathcal{I}(\psi_t)^{-1}$ is the inverse of the Fisher information matrix of ψ_t evaluated with respect to the information set available at time $t-1$. This quantity scales the score in order to account for the curvature of the likelihood at time t.

Note that ∇_t depends on y_t and the parameter value ψ_t. However, $\psi_t = \psi(\mathbf{y}_{1:t-1})$ such that $\nabla_t = \nabla(\mathbf{y}_{1:t})$, *i.e.*, the forcing variable u_t is a (generally non–linear) function of the data $\mathbf{y}_{1:t}$. Differently, y_t has been integrated out in the computation of $\mathcal{I}_t$ since the expectation is taken with respect to the information up to time $t-1$.

Properties of the score

The score ∇_t has zero expectation:

$$\begin{aligned}
\mathbb{E}_{t-1}[\nabla_t] &= \int_{\mathcal{Y}} \frac{\partial \log p(y_t|\mathbf{y}_{1:t-1};\psi)}{\partial \psi_t} p(y_t|\mathbf{y}_{1:t-1};\psi)\mathrm{d}y \\
&= \int_{\mathcal{Y}} \frac{\partial p(y_t|\mathbf{y}_{1:t-1};\psi)}{\partial \psi_t} \frac{1}{p(y_t|\mathbf{y}_{1:t-1};\psi)} p(y_t|\mathbf{y}_{1:t-1};\psi)\mathrm{d}y \\
&= \int_{\mathcal{Y}} \frac{\partial p(y_t|\mathbf{y}_{1:t-1};\psi)}{\partial \psi_t}\mathrm{d}y \\
&= \frac{\partial}{\partial \psi_t} \int_{\mathcal{Y}} p(y_t|\mathbf{y}_{1:t-1};\psi)\mathrm{d}y \\
&= \frac{\partial}{\partial \psi_t} 1 = 0
\end{aligned}$$

Provided that the Leibniz integral rule can be applied. It follows that $\{\nabla_{t-s}, s > 0\}$ forms a Martingale Difference Sequence.

Properties of the score

Since $\mathbb{E}_{t-1}[\nabla_t] = 0$, we have $Var(\nabla_t) = \mathbb{E}_{t-1}[\nabla_t^2] = \mathcal{I}_t(\psi_t)$, *i.e.*, the Fisher information matrix $\mathcal{I}_t(\psi_t)$ is the variance of the score.

It follows that the forcing variable $u_t = \mathcal{I}_t^{-1}\nabla_t$ has zero expectation and variance $\mathcal{I}_t^{-1}$

Unfolding the process of ψ_t we obtain:

$$\psi_t = \frac{\omega}{1-\beta} + \alpha \sum_{s=0}^{\infty} \beta^s u_{t-s-1}$$

such that $\mathbb{E}[\psi_t] = \frac{\omega}{1-\beta}$ and

$$Var(\psi_t) = \alpha^2 \sum_{s=0}^{\infty} \beta^{2s} \mathcal{I}_{t-s-1}^{-1}$$

Since $\mathbb{E}[u_t u_{t-s}] = 0$ for all $s \neq 0$.

Generalizations of the scaling mechanism

The choice of $S_t = \mathcal{I}_t^{-1}$ is somehow arbitrary. More generally, it is possible to write $S_t = \mathcal{I}_t^{-d}$ for $d \in (0, 1/2, 1)$, that is:

- $d = 0$: No scaling $u_t = \nabla_t$
- $d = 1/2$: Inverse square root scaling $u_t = \mathcal{I}_t^{-1/2}\nabla_t$, such that $Var(u_t) = 1$.
- $d = 1$: Inverse scaling $u_t = \mathcal{I}_t^{-1}\nabla_t$

The parameter d can be selected using: i) likelihood criteria, ii) theory arguments, iii) computational reasons.

For example, $\mathcal{I}_t$ is not always available in closed form. In this case $d = 0$ is a good strategy.

Handling parameter constraints

Most of the times we are in the situation when $\psi_t \in \Psi \subset \Re$. Think for example at ψ_t being the volatility at time t.

In these cases, we often work with: $\psi_t = \lambda(\tilde{\psi}_t)$, where $\tilde{\psi}_t \in \Re$ and $\lambda : \Re \rightarrow \Psi$ is a differentiable mapping function which is measurable with respect to the filtration $\mathcal{F}_{t-1} = \sigma(\mathbf{y}_{1:t-1})$.

We then rewrite the filter as:

$$
\begin{aligned}
\psi_t &= \lambda(\tilde{\psi}_t) \\
\tilde{\psi}_t &= \omega + \alpha \tilde{u}_{t-1} + \beta \tilde{\psi}_{t-1}
\end{aligned}
$$

Handling parameter constraints

After model reparametrization, the forcing variable needs to be modified as $\tilde{u}_t = \tilde{S}_t \tilde{\nabla}_t$. Assume $d = 1$ and set $\tilde{S}_t = \tilde{\mathcal{I}}_t^{-1}$. We have that:

$$\tilde{\nabla}_t = \frac{\partial \log p(y_t | \mathbf{y}_{1:t-1}; \tilde{\psi}_t)}{\partial \tilde{\psi}_t}$$

$$\tilde{\mathcal{I}}_t = \mathbb{E}_{t-1}[\tilde{\nabla}_t^2]$$

It is evident that:

$$\tilde{\nabla}_t = \frac{\partial \psi_t}{\partial \tilde{\psi}_t} \nabla_t$$

$$\tilde{\mathcal{I}}_t = \frac{\partial \psi_t}{\partial \tilde{\psi}_t}^2 \mathcal{I}_t$$

Handling parameter constraints

We have that for different choices of d, the jacobian of the mapping function $\frac{\partial \psi_t}{\partial \tilde{\psi}_t}$ affects the updating mechanism in different ways:

- $d = 0$: $\tilde{u}_t = \frac{\partial \psi_t}{\partial \tilde{\psi}_t} \nabla_t$.
- $d = 1/2$: $\tilde{u}_t = u_t$.
- $d = 1$: $\tilde{u}_t = \frac{\partial \psi_t}{\partial \tilde{\psi}_t}^{-1} u_t$

It follows that, handling parameter constraints with GAS models is straightforward and poses very little additional difficulties.

Volatility example under Gaussianity

Assume:

$$Y_t | \mathbf{y}_{t-1} \sim \mathcal{N}(0, \sigma_t^2),$$

The log pdf is proportional to:

$$\log p(y_t | \mathbf{y}_{t-1}; \sigma_t^2) \propto -\frac{1}{2}\left(\log(\sigma_t^2) + \frac{y_t^2}{\sigma_t^2}\right),$$

The score of σ_t^2 is:

$$\nabla_t = -\frac{1}{2\sigma_t^2}\left(\frac{y_t^2}{\sigma_t^2} - 1\right)$$

The information matrix for σ_t^2 is:

$$\mathcal{I}_t = \frac{1}{2\sigma_t^4}$$

Volatility example under Gaussianity

If we set $d = 1$, in the case of the GAS model with Gaussian innovation and time–varying variance, we find $u_t = y_t^2 - \sigma_t^2$. The corresponding updating equation is:

$$\begin{aligned}\sigma_t^2 &= \omega + \alpha(y_{t-1}^2 - \sigma_{t-1}^2) + \beta\sigma_{t-1}^2 \\ &= \omega + Ay_{t-1}^2 + B\sigma_{t-1}^2,\end{aligned}$$

where $A = \alpha$ and $B = \beta - \alpha$.

That is, a GAS model with Gaussian innovation, no mapping function and $d = 1$ resemble the GARCH(1,1) model. When $d \neq 1$ or we employ an exponential link function such as $\sigma_t = \exp(\tilde{\sigma}_t)$, the resulting model is different.

More of such special cases

- Exponential distribution (ACD and ACI): Engle and Russell (1998) and Russell (1999), respectively.
- Gamma distribution (MEM):Engle (2000) and Engle and Gallo (2006).
- Poisson: Davis et al. (2005).
- Multinomial distribution (ACM): Russell and Engle (2005).
- Binomial distribution: Cox (1958) and Rydberg and Shephard (2003)

Volatility example under Student's t

Assume $Y_t|\mathbf{y}_{1:t-1} \sim \mathcal{T}(\psi_t, \nu)$, where $\mathcal{T}$ represents a Student's t distribution with scale $\psi_t > 0$ and $\nu > 0$ degrees of freedom.

The log density is proportional to:

$$\log p(y_t|\mathbf{y}_{1:t-1}; \psi_t, \nu) \propto -\log \psi_t + \frac{1}{\psi_t}\frac{(\nu+1)y_t^2}{\nu\psi_t^2 + y_t^2}.$$

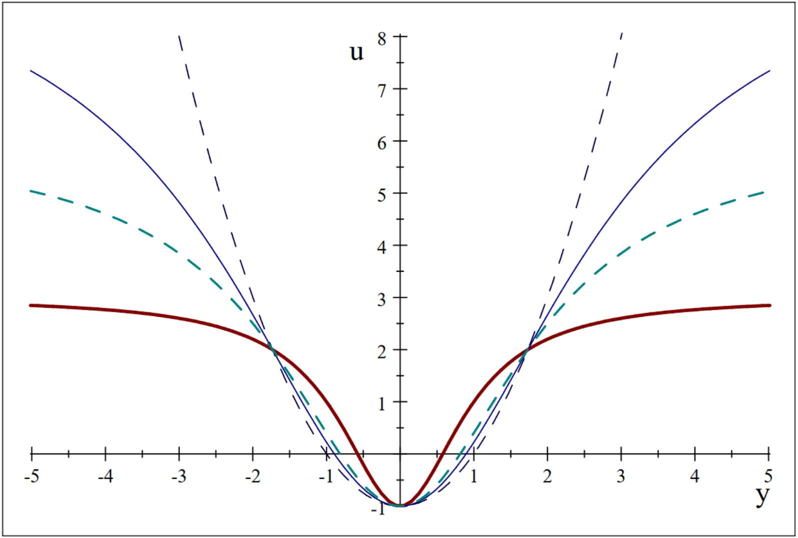
The score with respect to $\tilde{\psi}_t = \log(\psi_t)$ is:

$$\tilde{\nabla}_t = \frac{(\nu+1)y_t^2}{\nu\psi_t^2 + y_t^2} - 1$$

The Fisher information matrix for $\tilde{\psi}_t$ is:

$$\tilde{\mathcal{I}}_t = \frac{2\nu}{\nu+3}$$

Volatility example under Student's t



More about the volatility filter for the Student's t model

Taken from Ardia et al. (2017)

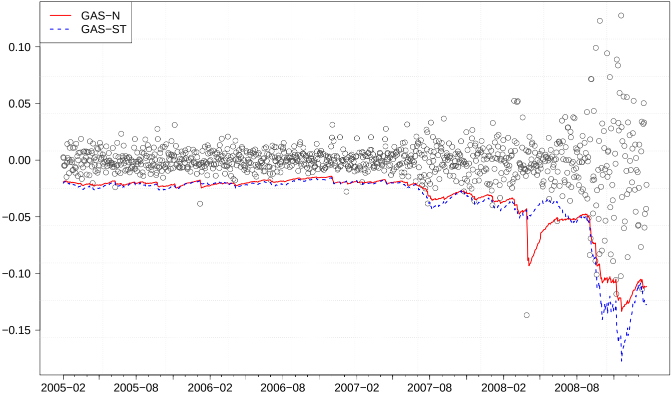


Figure: One–step ahead VaR forecasts for General Electric (GE) at the $\alpha = 1\%$ confidence level for the GAS–$\mathcal{N}$ (*solid*) and GAS–$\mathcal{ST}$ (*dotted*) models.

More about the volatility filter for the Student's t model

Taken from Harvey (2013)

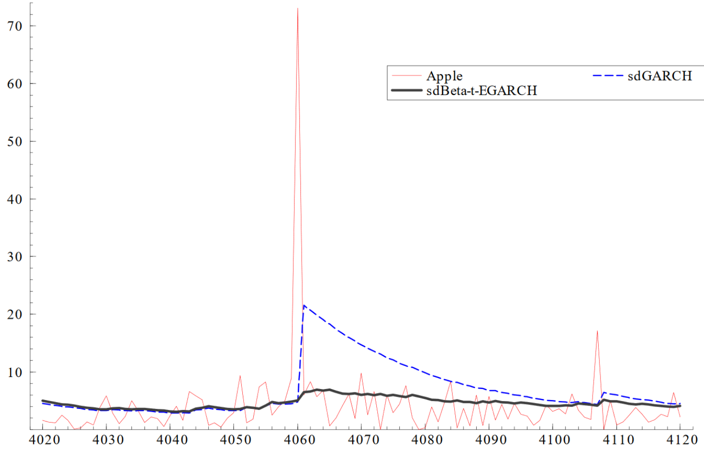


Figure: Absolute Apple returns and estimated volatility for GARCH and GAS with Student's t distribution.

More about the volatility filter for the Student's t model

Taken from www.gasmodel.com

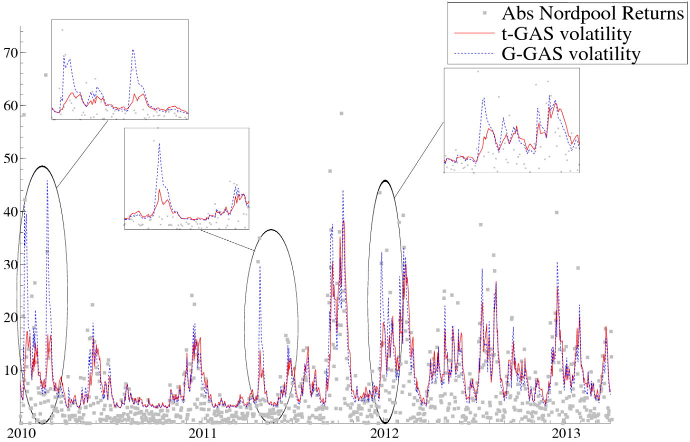


Figure: GAS estimated volatility for Nordpool electricity prices. Gaussian GAS is equivalent to Gaussian GARCH.

Estimation of GAS models

Estimation of GAS models can be done by Maximum Likelihood. For the Student's t volatility model let $\boldsymbol{\theta} = (\omega, \alpha, \beta, \nu)'$. The ML estimator for $\boldsymbol{\theta}$ is simply:

$$\hat{\boldsymbol{\theta}}^{ML} = \underset{\boldsymbol{\theta} \in \boldsymbol{\Theta}}{\arg\max} \sum_{t=1}^{T} \log p(y_t | \mathbf{y}_{1:t-1}; \psi_t(\boldsymbol{\theta}), \nu) \tag{7}$$

We generally impose: $|\beta| < 1$, $\alpha > 0$, $\omega \in \Re$ and $\nu > 2$. However, this may vary depending on the particular application.

Estimation of GAS models can be done in the R computational environment using the **GAS** package!

The **GAS** package for R

The **GAS** package for **R** permits to: i) simulate, ii) estimate, and iii) make predictions using GAS models.

- **GAS** can deal with univariate and multivariate models.
- Mostly written in C++.
- Works in parallel.
- It is available from CRAN and GitHub.

The two papers: Ardia et al. (2016) and Ardia et al. (2017) describe the main functionalities.

GAS models are not only for financial applications!!! Visit www.gasmodel.com.

GAS models are not only for financial applications

Label	Name	Type	Parameters	Scaling Type
`norm`	Gaussian	univariate	`location, scale`	`Identity, Inv, InvSqrt`
`snorm`	Skew–Gaussian	univariate	`location, scale, skewness`	`Identity`
`std`	Student–t	univariate	`location, scale, shape`	`Identity, Inv, InvSqrt`
`sstd`	Skew–Student–t	univariate	`location, scale, skewness, shape`	`Identity`
`ast`	Asymmetric Student–t with two tail decay parameters	univariate	`location, scale, skewness, shape, shape2`	`Identity, Inv, InvSqrt`
`ast1`	Asymmetric Student–t with one tail decay parameter	univariate	`location, scale, skewness, shape`	`Identity, Inv, InvSqrt`
`ald`	Asymmetric Laplace Distribution	univariate	`location, scale, skewness`	`Identity, Inv, InvSqrt`
`poi`	Poisson	univariate	`location`	`Identity, Inv, InvSqrt`
`negbin`	Negative Binomial	univariate	`location, scale`	`Identity, Inv, InvSqrt`
`ber`	Bernoulli	univariate	`location`	`Identity, Inv, InvSqrt`
`gamma`	Gamma	univariate	`scale, shape`	`Identity, Inv, InvSqrt`
`exp`	Exponential	univariate	`location`	`Identity, Inv, InvSqrt`
`beta`	Beta	univariate	`scale, shape`	`Identity, Inv, InvSqrt`
`mvnorm`	Multivariate Gaussian	multivariate	`location, scale, correlation`	`Identity`
`mvt`	Multivariate Student–t	multivariate	`location, scale, correlation, shape`	`Identity`

Table: Statistical distributions for which the `R` package **GAS** provides the functionality to simulate, estimate and forecast the time–variation in its parameters.

Examples of distributions

Distribution	Density	Link function
Poisson	$\frac{\lambda_t^{y_t}}{y_t!} e^{-\lambda_t}$	$\lambda_t = \exp(\alpha_t)$
Neg. Binomial	$\frac{\Gamma(k_1+y_t)}{\Gamma(k_1)\Gamma(y_t+1)} \left(\frac{k_1}{k_1+\lambda_t}\right)^{k_1} \left(\frac{\lambda_t}{k_1+\lambda_t}\right)^{y_t}$	$\lambda_t = \exp(\alpha_t)$
Exponential	$\lambda_t e^{-\lambda_t y_t}$	$\lambda_t = \exp(\alpha_t)$
Gamma	$\frac{1}{\Gamma(k_1)\beta_t^{k_1}} y_t^{k_1-1} e^{-y_t/\beta_t}$	$\beta_t = \exp(\alpha_t)$
Weibull	$\frac{k_1}{\beta_t} \left(\frac{y_t}{\beta_t}\right)^{k_1-1} e^{-(y_t/\beta_t)^{k_1}}$	$\beta_t = \exp(\alpha_t)$
Gaussian vol	$\frac{1}{\sqrt{2\pi}\sigma_t} e^{-y_t^2/2\sigma_t^2}$	$\sigma_t^2 = \exp(\alpha_t)$
Student's t vol	$\frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{(\nu-2)\pi}\Gamma(\frac{\nu}{2})\sigma_t} \left(1 + \frac{y_t^2}{(\nu-2)\sigma_t^2}\right)^{-\frac{\nu+1}{2}}$	$\sigma_t^2 = \exp(\alpha_t)$
Gaussian copula	$\frac{\frac{1}{2\pi\sqrt{1-\rho_t^2}} \exp\left[-\frac{z_{1t}^2+z_{2t}^2-2\rho_t z_{1t} z_{2t}}{2(1-\rho_t^2)}\right]}{\prod_{i=1}^2 \frac{1}{\sqrt{2\pi}} e^{-z_{it}^2/2}}$	$\rho_t = \frac{1-\exp(-\alpha_t)}{1+\exp(-\alpha_t)}$
Student's t copula	$\frac{\Gamma(\frac{\nu+2}{2})\Gamma(\frac{\nu}{2})}{\Gamma(\frac{\nu+1}{2})} \frac{\frac{1}{\sqrt{1-\rho_t^2}}\left[1+\frac{z_{1t}^2+z_{2t}^2-2\rho_t z_{1t} z_{2t}}{\nu(1-\rho_t^2)}\right]^{-\frac{\nu+2}{2}}}{\prod_{i=1}^2 (1+z_{it}/\nu)^{-\frac{\nu+1}{2}}}$	$\rho_t = \frac{1-\exp(-\alpha_t)}{1+\exp(-\alpha_t)}$

Examples of distributions

Distribution	GAS		ACM
	$\nabla_t(\theta_t)$	$\mathcal{I}_t(\theta_t)$	s_t
Poisson	$\frac{y_t}{\lambda_t} - 1$	$\frac{1}{\lambda_t}$	y_t
Neg. Binomial	$\frac{y_t}{\lambda_t} - \frac{k_1+y_t}{k_1+\lambda_t}$	$\frac{k_1}{\lambda_t(k_1+\lambda_t)}$	y_t
Exponential	$\frac{1}{\lambda_t} - y_t$	$\frac{1}{\lambda_t^2}$	y_t
Gamma	$\frac{y}{\theta_t^2} - \frac{k_1}{\beta_t}$	$\frac{k}{\beta_t^2}$	y_t/k_1
Weibull	$\frac{k_1}{\beta_t}\left[\left(\frac{y_t}{\beta_t}\right)^{k_1} - 1\right]$	$\left(\frac{k_1}{\beta_t}\right)^2$	$\frac{y_t}{\Gamma(1+k_1^{-1})}$
Gaussian vol	$\frac{1}{2\sigma_t^2}\left(\frac{y_t^2}{\sigma_t^2} - 1\right)$	$\frac{1}{2\sigma_t^4}$	y_t^2
Student's t vol	$\frac{1}{2\sigma_t^2}\left(\frac{\omega_t y_t^2}{\sigma_t^2} - 1\right)$	$\frac{\nu}{2(\nu+3)\sigma_t^4}$	y_t^2
	$\omega_t = \frac{\nu+1}{(\nu-2)+y_t^2/\sigma_t^2}$		
Gaussian cop	$\frac{(1+\rho^2)(\hat{z}_{1,t}-\rho_t)-\rho_t(\hat{z}_{2,t}-2)}{(1-\rho^2)^2}$	$\frac{1+\rho_t^2}{(1-\rho_t^2)^2}$	$z_{1,t}z_{2,t}$
Student's t cop	$\frac{(1+\rho^2)(\omega_t\hat{z}_{1,t}-\rho_t)-\rho_t(\omega_t\hat{z}_{2,t}-2)}{(1-\rho^2)^2}$	$\frac{(\nu+2+\nu\rho_t^2)}{(\nu+4)(1-\rho_t^2)^2}$	$z_{1,t}z_{2,t}$
	$\omega_t = \frac{\nu+2}{\nu+\frac{\hat{z}_{2,t}-2\rho_t\hat{z}_{1,t}}{1-\rho^2}}$		

References I

Ardia, D., Boudt, K., and Catania, L. (2016). Generalized Autoregressive Score Models in R: The **GAS** Package. *Available at SSRN: https://ssrn.com/abstract=2871444*.

Ardia, D., Boudt, K., and Catania, L. (2017). Value-at-Risk Prediction in R with the **GAS** Package. *Available at SSRN: https://ssrn.com/abstract=2871444*.

Bollerslev, T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. *The review of economics and statistics*, pages 542–547.

Cox, D. R. (1958). The regression analysis of binary sequences. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 215–242.

Cox, D. R. (1981). Statistical analysis of time series: Some recent developments [with discussion and reply]. *Scandinavian Journal of Statistics*, pages 93–115.

Creal, D., Koopman, S. J., and Lucas, A. (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics*, 28(5):777–795.

References II

Creal, D. D., Koopman, S. J., and Lucas, A. (2008). A general framework for observation driven time-varying parameter models.

Davis, R. A., Dunsmuir, W. T., and Streett, S. B. (2005). Maximum likelihood estimation for an observation driven model for poisson counts. *Methodology and Computing in Applied Probability*, 7(2):149–159.

DeJong, D. N. and Dave, C. (2011). *Structural macroeconometrics*. Princeton University Press.

Durbin, J. and Koopman, S. J. (2012). *Time series analysis by state space methods*, volume 38. OUP Oxford.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the Econometric Society*, pages 987–1007.

Engle, R. F. (2000). The econometrics of ultra-high-frequency data. *Econometrica*, 68(1):1–22.

References III

Engle, R. F. and Gallo, G. M. (2006). A multiple indicators model for volatility using intra-daily data. *Journal of Econometrics*, 131(1-2):3–27.

Engle, R. F. and Russell, J. R. (1998). Autoregressive conditional duration: a new model for irregularly spaced transaction data. *Econometrica*, pages 1127–1162.

Fokianos, K., Rahbek, A., and Tjostheim, D. (2009). Poisson autoregression. *Journal of the American Statistical Association*, 104(488):1430–1439.

Frühwirth-Schnatter, S. (2006). *Finite mixture and Markov switching models*. Springer Science & Business Media.

Hafner, C. M. and Manner, H. (2012). Dynamic stochastic copula models: Estimation, inference and applications. *Journal of Applied Econometrics*, 27(2):269–295.

Harvey, A. C. (2013). *Dynamic models for volatility and heavy tails: with applications to financial and economic time series*, volume 52. Cambridge University Press.

Harvey, A. C. and Chakravarty, T. (2008). Beta-t-(e) garch.

References IV

Masreliez, C. (1975). Approximate non-gaussian filtering with linear state and observation relations. *IEEE Transactions on Automatic Control*, 20(1):107–110.

Patton, A. J. (2006). Modelling asymmetric exchange rate dependence. *International economic review*, 47(2):527–556.

Russell, J. R. (1999). Econometric modeling of multivariate irregularly-spaced high-frequency data.

Russell, J. R. and Engle, R. F. (2005). A discrete-state continuous-time model of financial transactions prices and times: The autoregressive conditional multinomial–autoregressive conditional duration model. *Journal of Business & Economic Statistics*, 23(2):166–180.

Rydberg, T. H. and Shephard, N. (2003). Dynamics of trade-by-trade price movements: decomposition and models. *Journal of Financial Econometrics*, 1(1):2–25.

Taylor, S. J. (1986). *Modelling financial time series*. Wiley.

West, M. and Harrison, W. (1997). *Bayesian forecasting and Dynamic Models*. Springer.