Score-driven time series models

Paolo Gorgi

Dept. of Statistical Sciences, University of Padua
gorgi@stat.unipd.it
October 1*st*, 2015

UNIVERSITÀ DEGLI STUDI DI PADOVA

Outline

Introduction (1)

- A general approach to specify a dynamic model for the observed data $\{y_1, \ldots y_n\}$ is to consider a parametric model and to allow some of the parameters to be time varying.

- The random elements y_t, conditionally on λ_t, are independently distributed as

$$y_t|\lambda_t \sim p(y_t|\lambda_t, \theta),$$

  where $p(\cdot|\lambda_t; \theta)$ is a density function.

- Depending on the specification of λ_t, two classes of models: **parameter driven** and **observation driven** models.

Introduction (2)

- **Parameter driven models:** λ_t is a stochastic process that is not perfectly predictable given the past information $\mathcal{F}_{t-1}$.
- **Observation driven models:** λ_t depends only on past random variables and it is perfectly predictable given $\mathcal{F}_{t-1}$.

Example: time varying variance

$$y_t = \sigma_t \varepsilon_t, \;\; \varepsilon_t \overset{\text{iid}}{\sim} N(0, 1),$$

Stochastic Volatility model:

$$\ln \sigma_{t+1} = \omega + \beta \ln \sigma_t + \sigma_\eta \eta_t, \;\; \eta_t \overset{\text{iid}}{\sim} N(0, 1).$$

GARCH model:

$$\sigma_{t+1}^2 = \omega + \beta \sigma_t^2 + \alpha (y_t^2 - \sigma_t^2).$$

Introduction (3)

- A great advantage of **observation driven models** is that the likelihood function is available in closed form. There is **no need of time-consuming simulation-based methods** for inference.
- For the GARCH model the use of the squared observations y_t^2 as driving mechanism for σ_t^2 is intuitive. However, in general it is not obvious what function should be used to update λ_t.

Question: How to specify λ_t in an observation-driven setting?

Answer: Generalized Autoregressive Score (GAS) models.

GAS framework

Università degli Studi di Padova

- The peculiarity of GAS models is that λ_t is driven by the score of the predictive log-likelihood, i.e.

$$u_{\lambda,t} := \frac{\partial \log p(y_t|\lambda_t;\theta)}{\partial \lambda_t}.$$

- the time varying λ_t is defined as

$$\lambda_{t+1} = \omega_\lambda + \beta_\lambda \lambda_t + \alpha_\lambda s_{\lambda,t}, \quad s_{\lambda,t} = S_{\lambda,t} u_{\lambda,t}$$

  where $S_{\lambda,t}$ is a positive scaling factor, function of λ_t.

- Typical choices of $S_{\lambda,t}$ are the identity, I_t^{-1} and $I_t^{-1/2}$, where $I_t = E[u_{\lambda,t}^2|\lambda_t]$.

Why GAS models?

Some reasons to use GAS models

- Main reason: they are very effective in practical applications.
- They are observation driven models and therefore they can be easily estimated by maximum likelihood, no need of simulation methods.
- The use of the score of the predictive likelihood as updating mechanism, besides being intuitive, is optimal from an information prospective, see Blasques *et al.* (2015).

Some properties of the GAS process

- The sequence $\{s_{\lambda,t}\}_{t\in\mathbb{Z}}$ is a martingale difference, i.e. $E[s_{\lambda,t}|\lambda_t] = 0$.
- The h-steps ahead conditional expectation for λ_t is

$$E[\lambda_{t+h}|\lambda_t] = \omega_\lambda \sum_{i=0}^{h-1} \beta_\lambda^i + \beta_\lambda^h \lambda_t.$$

- $E[\lambda_t] = \omega_\lambda/(1-\beta_\lambda)$, when $E|\lambda_t| < \infty$.
- $Var[u_t|\lambda_t] = I_t$.

Likelihood-based inference

UNIVERSITÀ DEGLI STUDI DI PADOVA

- GAS models can be easily estimated by maximum likelihood, the conditional log-likelihood function is written as

$$\hat{L}_n(\theta) = \sum_{i=1}^{n} \log p(y_t | \hat{\lambda}_t(\theta), \theta).$$

- The time varying parameter $\hat{\lambda}_t(\theta)$ is obtained recursively as

$$\hat{\lambda}_t(\theta) = \omega_\lambda + \beta_\lambda \hat{\lambda}_{t-1}(\theta) + \alpha \hat{s}_{\lambda,t}(\theta), \; t \in \mathbb{N}$$

for a given initialization $\hat{\lambda}_0(\theta)$.

- The initialization $\hat{\lambda}_0(\theta)$ is typically set equal to $\omega_\lambda/(1-\beta_\lambda)$. Alternatively, $\hat{\lambda}_0(\theta)$ can be considered a parameter to be estimated.

GAS models

- It turns out that many existing observation-driven models are **GAS** models; for instance, the well known Gaussian **GARCH** model is a Gaussian volatility GAS model with scaling factor $S_t = I_t^{-1}$.
- Other existing models that are GAS models: **EGARCH** of Nelson (1991), **ACD** of Engle and Russell (1998), **MEM** of Engle (2002), **ACM** of Rydberg and Shephard (2003) and the **Poisson model** of Davis *et al.* (2003).
- GAS approach leads also to many new models, see for instance Creat *et al.* (2013) and Harvey (2013).

Two examples of application

Two examples of GAS models for time series with fat tails:

- **Fat-tailed location model.** A student-t model with time varying mean, see Harvey and Luati (2013) and Harvey (2013).
- **Fat-tailed volatility model** A student-t model with time varying variability, see Creat *et al.* (2013) and Harvey (2013).

Example: t-GAS location model (1)

- Consider the model

$$y_t = \mu_t + \varepsilon_t, \qquad \varepsilon_t \overset{\text{iid}}{\sim} t_\nu,$$

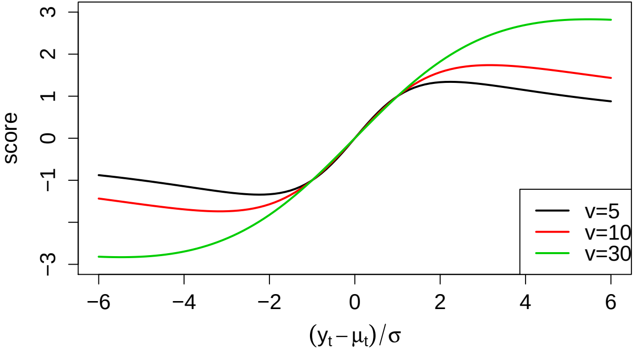
  where

$$\mu_{t+1} = \omega_\mu + \beta_\mu \mu_t + \alpha_\mu \frac{(\nu+1)(y_t - \mu_t)\sigma^{-1}}{\nu + (y_t - \mu_t)^2 \sigma^{-2}}.$$

- In the limit $\nu \to \infty$, the model distribution of ε_t is Gaussian and μ_t is a linear combination of past y_t as the score is proportional to $y_t - \mu_t$.

Example: t-GAS location model (2)

- Impact of the standardized observation $(y_t - \mu_t)/\sigma$ on the score for different values of ν



Example: t-GAS location model (3)

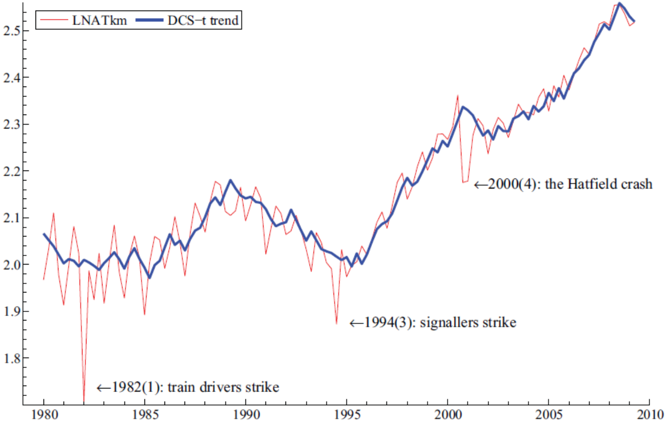


Figure: Quarterly UK National Rail Travel, Km traveled by UK passengers. This plot is from Harvey and Luati (2013).

Example: t-GAS volatility model (1)

- Fat tails are a well known feature observed in stock returns and student-t distribution is often considered

$$y_t = \sigma_t \varepsilon_t, \;\; \varepsilon_t \overset{\text{iid}}{\sim} t_\nu,$$

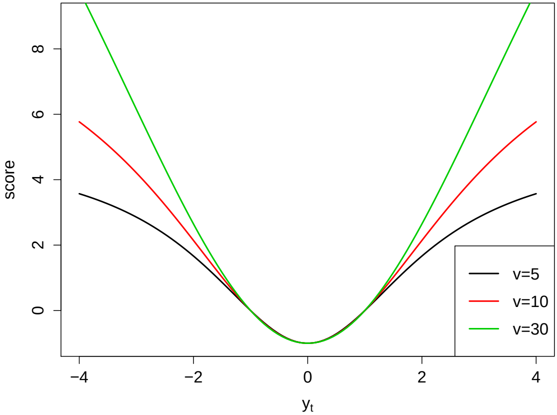
where

$$\sigma_{t+1}^2 = \omega_\sigma + \beta_\sigma \sigma_t^2 + \alpha_\sigma \left(\frac{(\nu+1) y_t^2}{\nu + \sigma_t^{-2} y_t^2} - \sigma_t^2 \right).$$

- In the limit $\nu \to \infty$, the t-GAS volatility model becomes a GARCH model.

Example: t-GAS volatility model (2)

- Impact of y_t on the score of the t-GAS volatility model.



Example: t-GAS volatility model (3)

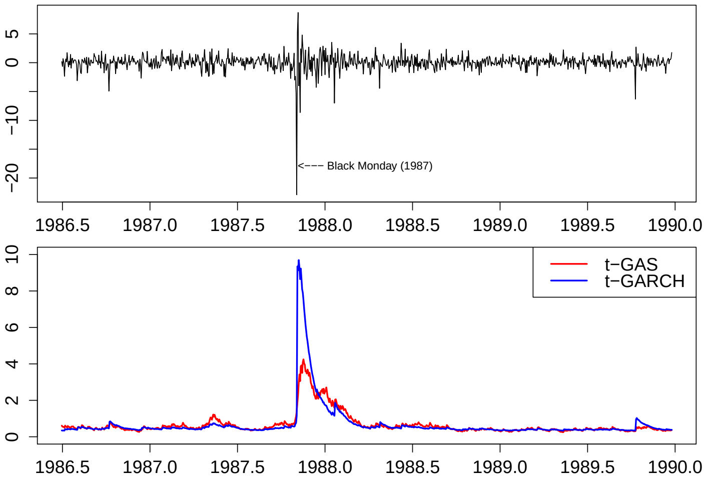


Figure: S&P 500 daily Index.

A flexible class of GAS models (1)

- Some time series exhibit complex dynamic behaviours and a GAS model may not be enough to properly model the data.
- The idea is to make GAS models more flexible allowing the parameter α_λ in the GAS recursion to be time varying.
- We consider the parameter α_λ because, in some sense, it determines the amount of information in the data.

A flexible class of GAS models (2)

- The GAS recursion becomes

$$\lambda_{t+1} = \omega_\lambda + \beta_\lambda \lambda_t + h(f_{t+1}) s_{\lambda,t}, \qquad s_{\lambda,t} = S_{\lambda,t} u_{\lambda,t}$$

  where $h(\cdot)$ is a link function and the time varying α_λ is $h(f_t)$.

- We specify f_t relying on the score of the predictive log-likelihood $\partial \log p(y_t|\lambda_t;\theta)/\partial f_t$

$$f_{t+1} = \omega_f + \beta_f f_t + \alpha_f s_{f,t}, \qquad s_{f,t} = S_{f,t} u_{\lambda,t} u_{\lambda,t-1}.$$

- The resulting specification of $s_{f,t}$ is very intuitive!

Simulation example (1)

- We model the data with a Gaussian model with time varying mean

$$y_t = \mu_t + \epsilon_t, \quad \epsilon_t \overset{\text{iid}}{\sim} N(0, \sigma^2).$$

- The time varying mean μ_t is given by

$$\begin{aligned} \mu_{t+1} &= \mu_t + \exp(f_{t+1}) s_{\mu,t} \\ f_{t+1} &= \omega_f + \beta_f f_t + \alpha_f s_{f,t}, \end{aligned}$$

where $s_{\mu,t} = (y_t - \mu_t)$ and $s_{f,t} = (y_t - \mu_t)(y_{t-1} - \mu_{t-1})$.

Simulation example (2)

- Consider the following data generating process

$$y_t = \mu_t^o + \epsilon_t, \quad \varepsilon_t \overset{\text{iid}}{\sim} N(0, \sigma^2),$$

  the time varying mean is given by

$$\mu_t^o = \begin{cases} 0 & \text{if} \quad \sin\left((\pi t - 1)/200\right) > 0 \\ 3 & \text{if} \quad \sin\left((\pi t - 1)/200\right) < 0. \end{cases}$$

- We generate time series from this process and show how this flexible GAS model is able to better approximate the true mean μ_t^o

Simulation example (3)

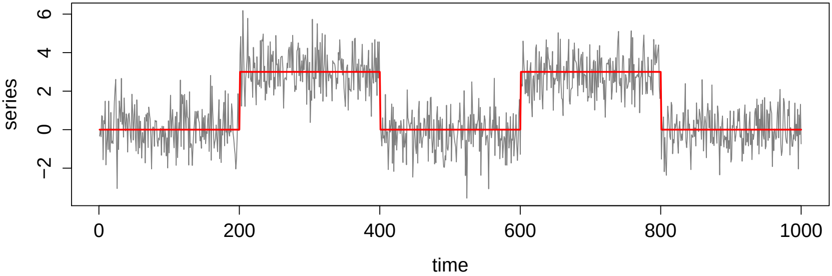


Figure: Time series generated from the data generating process.

Simulation example (4)

Università degli Studi di Padova

The average mean square error is 47% larger for the GAS model without the time varying f_t.

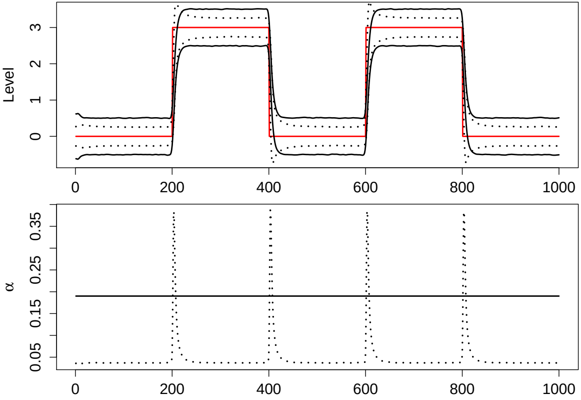


Figure: 1st plot: 90% confidence bounds filtered μ_t. 2nd plot: average estimate of $\exp(f_t)$. The dashed lines are for the model with time varying f_t and the continuous lines are for the fixed f_t.

Empirical example: the series

- Changing dynamics in the US inflation process are well documented in the literature.

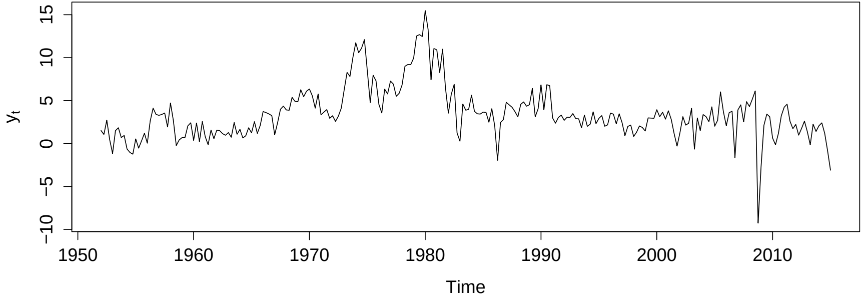


Figure: Quarterly US consumer price inflation series.

Empirical example: a fat-tailed model (1)

- Consider the following model

$$y_t = \mu_t + \sigma_t \epsilon_t, \quad \varepsilon_t \overset{\text{iid}}{\sim} t_\nu(0,1),$$

where

$$\begin{aligned} \mu_{t+1} &= \mu_t + \exp(f_{t+1}) s_{\mu,t}, \\ f_{t+1} &= \omega_f + \beta_f f_t + \alpha_f s_{f,t}, \\ \ln \sigma_{t+1} &= \omega_\sigma + \beta_\sigma \ln \sigma_t + \alpha_\sigma s_{\sigma,t}, \end{aligned}$$

- The innovations are defined as

$$s_{\mu,t} = \frac{(\nu+1)(y_t-\mu_t)\sigma_t^{-1}}{(\nu-2)+(y_t-\mu_t)^2\sigma_t^{-2}}, \quad s_{\sigma,t} = \frac{(\nu+1)(y_t-\mu_t)^2\sigma_t^{-2}}{(\nu-2)+(y_t-\mu_t)^2\sigma_t^{-2}} - 1$$

$$s_{f,t} = s_{\mu,t} s_{\mu,t-1}.$$

Empirical example: a fat-tailed model (2)

UNIVERSITÀ DEGLI STUDI DI PADOVA

Impact of the standardized observation $(y_t - \mu_t)\sigma_t^{-1}$ on the scores for $\nu = 5$

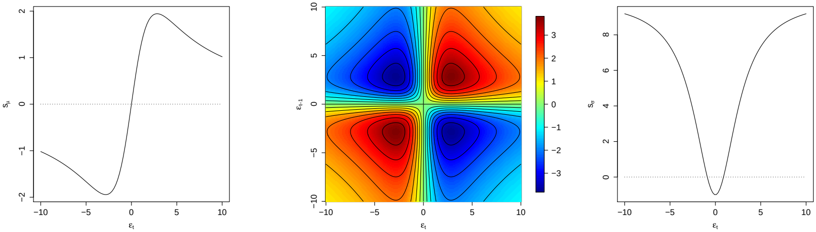


Figure: First plot: $s_{\mu,t}$, second plot: $s_{f,t}$, third plot: $s_{\sigma,t}$.

Empirical example: estimates

	δ_f	β_f	α_f	δ_σ	β_σ	α_σ	v	AICc
M1	-0.75 (0.40)	0.97 (0.03)	0.85 (0.42)	0.53 (0.12)	0.86 (0.09)	0.11 (0.04)	5.57 (1.57)	967
M2	-0.75 (0.20)	0.91 (0.03)	0.71 (0.24)	0.59 (0.09)	-	-	3.82 (0.55)	978
M3	-0.23 (0.14)	-	-	0.54 (0.10)	0.87 (0.12)	0.08 (0.05)	7.58 (2.40)	976
M4	-0.15 (0.11)	-	-	0.56 (0.07)	-	-	5.64 (1.43)	986

Table: The parameters δ_i, $i = 1, 2$ are a re-parametrization, i.e. $\delta_i = \omega_i/(1 - \beta_i)$. The last column contains the corrected Akaike Information Criterion.

Empirical example: filtered parameters

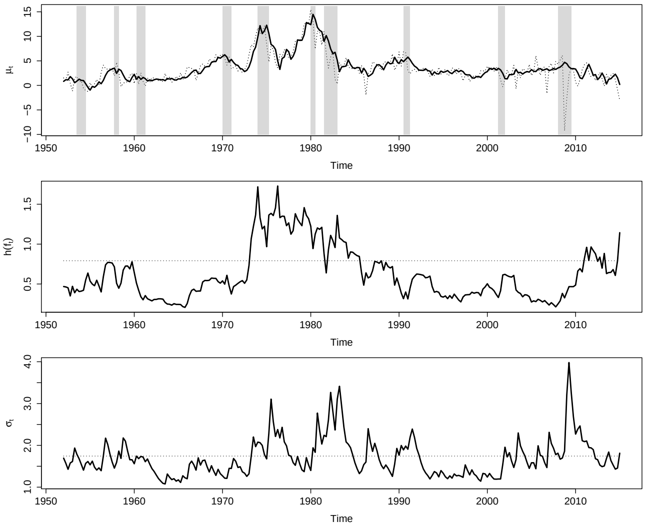


Figure: First plot: estimated μ_t, second plot: $\exp(f_t)$, third plot σ_t.

Empirical example: filtered mean

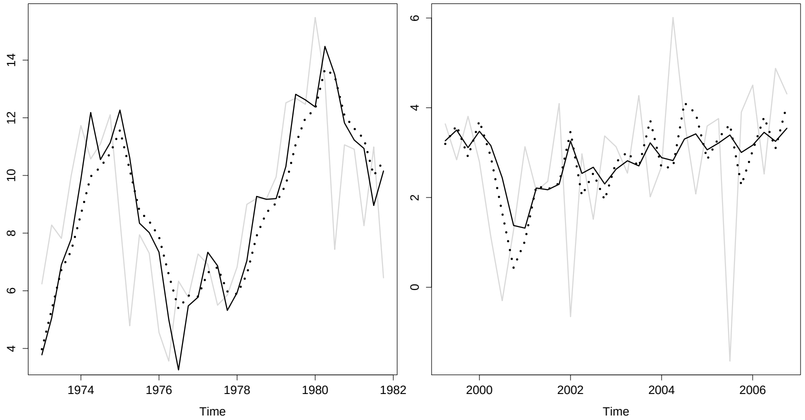


Figure: Time varying mean. Dashed lines: standard GAS model, continuous lines: flexible GAS model.

Forecasting performance of the model.

	$h = 1$	$h = 2$	$h = 3$	$h = 4$
M1	1.00	**1.00**	**1.00**	**1.00**
M2	**0.99**	1.01	1.04	1.02
M3	1.04	1.16	1.14	1.18*
M4	1.02	1.13	1.13	1.15*
Local level model	1.08	1.26	1.20	1.20*
ARIMA(4,1,0)	1.11	1.30	1.33	1.29*
ARIMA(1,1,1)	1.05	1.21	1.16	1.17*

Table: Out of sample FMSE ratio from the last 100 observations of the quarterly US consumer price inflation series. The model is estimated on rolling windows and the DM test is used. The benchmark is model 1. The FMSE of model 1 is at the denominator of the ratio.

Concluding comments

- GAS models provide a general framework to specify dynamic observation-driven models.
- They are very effective in practical applications and the parameters can be easily estimated.
- The flexible class of GAS models we propose allow to handle more complex situations.
- Something I did not mention: the score-driven update for the additional parameter f_t can be justified by a similar optimality argument as in Blasques *et al.* (2015).

Some references

Blasques, F., Koopman, S. J. and Lucas, A. (2015). Information-theoretic optimality of observation-driven time series models for continuous responses. *Biometrika* **102**, 325–343.

Creal, D., Koopman, S. J. & Lucas, A. (2013). generalized autoregressive score models with applications. *Journal of Applied Econometrics* **28**, 1099–1255.

Harvey, A. & Luati, A. (2014). Filtering With Heavy Tails. *Journal of the American Statistical Association* **109**, 1112–1122.

Harvey, A. (2013). Dynamic Models for Volatility and Heavy Tails, Econometric Society Monograph. *New York: Cambridge University Press.*