Generalized Autoregressive Score Models for Time-varying Parameters: new models and applications

Francisco Blasques[a], Drew D. Creal[b], Pawel Janus[c],
Siem Jan Koopman[a,d], André Lucas[a],
Marcel Scharth[e], Bernd Schwaab[f]

[a]Vrije Universiteit Amsterdam and Tinbergen Institute
[b]Booth School of Business University of Chicago
[c]UBS Zurich
[d]CREATES, Aarhus University
[e]University of New South Wales, Australia
[f]European Central Bank, Frankfurt am Main

Workshop ISF 2014 Rotterdam

What next ?

- We present a short introduction & review on GAS models.
- Focus: time-varying parameter models.
- Score driven models reduce to many well-established models in financial econometrics.
- Here we show how interesting new model formulations can be derived.

Agenda

- Forecasting with GAS models and comparisons with State Space Models
- Dynamic Factor Models with Mixed Measurements and Mixed Frequencies
- Modelling Dynamic Volatilities and Correlations using GAS models

The basic framework

Consider model for the data y which we represent as $p(y;\psi)$.
Parameter vector is ψ.
In time series, we evaluate likelihood function via prediction errors

$$p(y;\psi) = p(y_1;\psi)\prod_{t=2}^{n} p(y_t|y_1,\dots,y_{t-1};\psi).$$

Assume that we want to consider a sub-set of ψ as time-varying :

$$\psi_t = (f_t;\theta),$$

where f_t represents the time-varying parameter and θ the remaining fixed coefficients.

The TV parameter f_t typically represents β_t and/or σ_t.
The TV parameter may be modelled in an autoregressive form

$$f_{t+1} = \omega + Bf_t + A \times \text{" some innovation "}.$$

Score driven models

The t-th contribution to the loglikelihood $\ell = \log p(y; \psi)$:

$$\ell_t = \log p(y_t | y_1, \ldots, y_{t-1}, f_1, \ldots, f_t; \theta),$$

where we assume that $f_1, \ldots, f_t$ are known (they are realized).

The parameter value for next period, f_{t+1}, is determined by an autoregressive updating function that has an innovation equal to the score of ℓ_t with respect to f_t.

By determining f_{t+1} in this way, we obtain a recursive algorithm for the estimation of time-varying parameters.

We have labelled this approach as the

generalized autoregressive score model,

or the GAS model. More details are given next.

Generalized autoregressive score model

For the observation equation,

$$y_t \sim p(y_t | Y_{t-1}, f_t; \theta), \qquad Y_t = \{y_1, \ldots, y_t\},$$

we propose a GAS updating scheme for f_t based on

$$f_{t+1} = \omega + Bf_t + As_t,$$

where the innovation or driving mechanism s_t is given by

$$s_t = S_t \cdot \nabla_t$$

where

$$\begin{aligned} \nabla_t &= \frac{\partial \ln p(y_t | Y_{t-1}, f_t; \theta)}{\partial f_t}, \\ S_t &= \mathcal{I}_{t-1}^{-1} = -E_{t-1}\left[\frac{\partial^2 \ln p(y_t | Y_{t-1}, f_t; \theta)}{\partial f_t \partial f_t'}\right]^{-1}. \end{aligned}$$

Volatility modelling

We have

$$y_t = \mu + \varepsilon_t, \qquad \varepsilon_t \sim \text{NID}(0, f_t).$$

The GAS model for f_t can be constructed by considering

$$\begin{aligned} y_t &\sim p(y_t | Y_{t-1}, f_t; \theta), \\ f_{t+1} &= \omega + B f_t + A s_t, \end{aligned}$$

with driving mechanism

$$s_t = S_t \cdot \nabla_t$$

where

$$\begin{aligned} \nabla_t &= \frac{\partial \ln p(y_t | Y_{t-1}, f_t; \theta)}{\partial f_t}, \\ S_t &= \mathcal{I}_{t-1}^{-1} = -E_{t-1} \left[\frac{\partial^2 \ln p(y_t | Y_{t-1}, f_t; \theta)}{\partial f_t \partial f_t'} \right]^{-1}. \end{aligned}$$

GAS variance updating reduces to GARCH

Assume $\mu = 0$, we have

$$y_t = \varepsilon_t, \qquad \varepsilon_t \sim \text{NID}(0, f_t),$$

with variance $f_t = \sigma_t^2$. Score and inverse information matrix are:

$$\begin{aligned}
\ln p(y_t | Y_{t-1}, f_t; \theta) &= -\frac{1}{2}\ln 2\pi - \frac{1}{2}\ln f_t - \frac{y_t^2}{2f_t}, \\
\nabla_t &= \frac{1}{2f_t^2} y_t^2 - \frac{1}{2f_t} \; = \; \frac{1}{2f_t^2}(y_t^2 - f_t), \\
E_{t-1}(\nabla_t) &= 0, \qquad -\mathcal{I}_{t-1} \; = \; -\frac{1}{2f_t^2}, \\
S_t \; = \; \mathcal{I}_{t-1}^{-1} &= 2f_t^2,
\end{aligned}$$

and we have $s_t = S_t \cdot \nabla_t = y_t^2 - f_t$ for the GAS updating

$$f_{t+1} = \omega + Bf_t + A(y_t^2 - f_t).$$

Hence, this GAS update scheme reduces to GARCH for $f_t = \sigma_t^2$:

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

Volatility modeling

A class of volatility models is given by

$$y_t = \mu + \sigma(f_t)u_t, \qquad u_t \sim p_u(u_t;\theta), \qquad t = 1, 2, \ldots, T,$$

$$f_{t+1} = \omega + \beta f_t + \alpha s_t,$$

where:

- $\sigma()$ is some continuous function;
- $p_u(u_t;\theta)$ is a standardized disturbance density;
- s_t is the scaled score based on $\partial \log p(y_t | Y_{t-1}, f_t;\theta) \,/\, \partial f_t$.

Some special cases

- $\sigma(f_t) = f_t$ and p_u is Gaussian : GAS $\Rightarrow$ GARCH;
- $\sigma(f_t) = \exp(f_t)$ and p_u is Gaussian : GAS $\Rightarrow$ EGARCH;
- $\sigma(f_t) = \exp(f_t)$ and p_u is Student's t : GAS $\Rightarrow$ t-GAS.

Another example: modelling durations

Consider an exponential ($\mathcal{E}$ is exponential density) model,

$$y_t = \lambda_t \varepsilon_t, \qquad \varepsilon_t \sim \mathcal{E}(1).$$

Let $f_t = \lambda_t$. The score and inverse of the information matrix are:

$$\nabla_t = \frac{y_t}{f_t^2} - \frac{1}{f_t},$$
$$S_t = \mathcal{I}_{t-1}^{-1} = f_t^2.$$

Here the GAS update scheme reduces to
the E-ACD model of Engle and Russell (1998):

$$f_{t+1} = \omega + A(y_t - f_t) + Bf_t$$

More of such special cases

GAS updating for appropriate observation densities and particular scaling choices reduces to well-known GARCH-type time series models.

- GARCH for $N(0, f_t)$: Engle (1982), Bollerslev (1986)
- EGARCH for $N(0, \exp f_t)$: Nelson (1991)
- Exponential distribution (ACD and ACI): Engle & Russell (1998) and Russell (2001), respectively
- Gamma distribution (MEM): Engle (2002), Engle & Gallo (2006)
- Poisson: Davis, Dunsmuir & Street (2003)
- Multinomial distribution (ACM): Russell & Engle (2005)
- Binomial distribution: Cox (1956), Rydberg & Shephard (2002)

We discuss this general GAS framework in
Creal, Koopman and Lucas (2013, JAE).

Discussion

- In econometrics, score and Hessian are familiar entities in estimation;
- Using contribution of score at time t only (wrt predictive density) and using it as an innovation in a time-varying parameter scheme is not unreasonable.
- It turns out that many GARCH-type time series models are effectively constructed in this way.
- In case of GARCH (Gaussian), innovation or driver mechanism has an interpretation : $\mathbb{E}(y_t^2) = \sigma^2$.
- In other cases (incl. GARCH with t-densities), choice of driver mechanism is not so clear.
- We then can rely on GAS and still get an appropriate updating scheme.

Statistical properties

The GAS(p, q) model is

$$
\begin{aligned}
y_t &\sim p(y_t | Y_{t-1}, f_t, f_{t-1}, \ldots, f_{t-q}; \theta), \\
f_{t+1} &= \omega + \sum_{i=0}^{p-1} A_i s_{t-i} + \sum_{j=0}^{q-1} B_j f_{t-j} \\
s_t &= S_t \cdot \nabla_t
\end{aligned}
$$

- The expectation of the score is zero: $E_{t-1}[\nabla_t] = 0$.
- As a result, s_t is a martingale difference sequence.
- If f_t is stationary, its unconditional expectation is $\mathrm{E}[f_t] = \omega \left(I - B(1)\right)^{-1}$.
- Conditions for stationarity and ergodicity of GAS process :
  Blasques, Koopman and Lucas (BKL, 2013).
- Asymptotic properties of MLE (Consistency, AN) : BKL 2014a.
- Optimality of score updating in Kullback-Leibler sense : BKL 2014b.

Different specifications

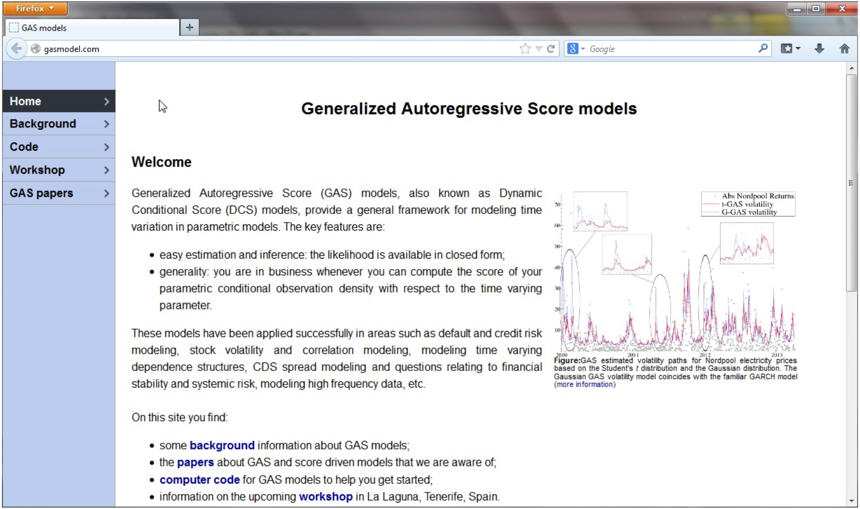
$$
\begin{aligned}
y_t &\sim p(y_t | Y_{t-1}, f_t, f_{t-1}, \ldots, f_{t-q}; \theta), \\
f_{t+1} &= \omega + \sum_{i=0}^{p-1} A_i s_{t-i} + \sum_{j=0}^{q-1} B_j f_{t-j} \\
s_t &= S_t \cdot \nabla_t
\end{aligned}
$$

- The default choice for scaling is $S_t = \mathcal{I}_{t-1}^{-1}$ or $S_t = \mathcal{I}_{t-1}^{-1/2}$.
- Alternative: $S_t = I$; "steepest descent" appears to be less stable...
- In case default choice is close to singular, we can do some mild smoothing of past $\mathcal{I}_t$'s using an EWMA scheme:

$$\mathcal{I}_{t-1}^c = \tilde{\alpha}\mathcal{I}_{t-2}^c + (1-\tilde{\alpha})\mathcal{I}_{t-1},$$

  and $S_t = (\mathcal{I}_{t-1}^c)^{-1}$. This appears to work very effectively.
- Extensions with long-memory: Janus, Koopman and Lucas (2012).

Recent developments: http://gasmodel.com



Firefox

GAS models

gasmodel.com

Home

Background

Code

Workshop

GAS papers

Generalized Autoregressive Score models

Welcome

Generalized Autoregressive Score (GAS) models, also known as Dynamic Conditional Score (DCS) models, provide a general framework for modeling time variation in parametric models. The key features are:

- easy estimation and inference: the likelihood is available in closed form;
- generality: you are in business whenever you can compute the score of your parametric conditional observation density with respect to the time varying parameter.

These models have been applied successfully in areas such as default and credit risk modeling, stock volatility and correlation modeling, modeling time varying dependence structures, CDS spread modeling and questions relating to financial stability and systemic risk, modeling high frequency data, etc.

Figure: GAS estimated volatility paths for Nordpool electricity prices based on the Student's *t* distribution and the Gaussian distribution. The Gaussian GAS volatility model coincides with the familiar GARCH model (more information)

On this site you find:

- some **background** information about GAS models;
- the **papers** about GAS and score driven models that we are aware of;
- **computer code** for GAS models to help you get started;
- information on the upcoming **workshop** in La Laguna, Tenerife, Spain.

Recent developments: `http://gasmodel.com`

- Harvey (2013) *Dynamic Models for Volatility and Heavy Tails: With Applications to Financial and Economic Time Series*
- De Lira Salvatierra and Patton (2013) *Dynamic Copula Models and High Frequency Data*
- Ito (2013) *Modeling Dynamic Diurnal Patterns in High Frequency Financial Data*
- Janus, Koopman and Lucas (2013) *Long memory GAS*
- Oh and Patton (2013) *Time-Varying Systemic Risk: Evidence from a Dynamic Copula Model of CDS Spreads*
- Lucas, Schwaab and Zhang (2013) *Measuring credit risk in a large banking system: econometric modeling and empirics*
- Boudt, Danielsson, Koopman and Lucas (2013) *Regime Switches in the Volatility and Correlation of Financial Institutions*

Prediction based on parameter-driven versus observation-driven models

S.J. Koopman, A. Lucas and M. Scharth

Focus is out-of-sample prediction of *parameters*

We consider dynamic models for count, intensity, duration, volatility and copula densities using three specifications that are popular in economic and financial time series:

1. nonlinear non-Gaussian state space model as formulated for example by Durbin and Koopman (2000); this is the class of parameter-driven models.
2. observation-driven models based on the score function of the predictive likelihood function as formulated by Creal, Koopman and Lucas (2011, 2013) and Harvey (2013).
3. observation-driven models based on the moment function of the time series; typical examples are GARCH model of Engle (1982) and Bollerslev (1987), autoregressive conditional duration model of Engle and Russell (1998), and multiplicative error models of Engle and Gallo (2006).

Are these models equally general ?

Parameter-driven models are flexible and can be easily adjusted in new settings.

Observation-driven models have so far lacked a similarly flexible unifying framework: for a new observation density and parametrisation.

- To update the time-varying parameter as a function of past and current data: what is the appropriate funtion ?
- In terms of volatility, y_t^2 can be argued as a "natural" driver.
- In many other settings it may not be evident...
- The score function provides a unified solution and can be easily applied.

Can we compare these classes of models ?

The predictive distribution of a parameter-driven model is a mixture of measurement densities for the stochastically time-varying parameter.

The predictive density of observation-driven models is simply the observation density given a perfectly predictable parameter.

- Parameter-driven models typically generate overdispersion related to mixtures: heavier tails and other features.
- We need to control for this difference.
- We need to accommodate similar degrees of overdispersion and fat tails as parameter-driven models.
- It requires models based on exponential-gamma, Weibull-gamma and double-gamma mixtures: intrinsically interesting duration and multiplicative error models.

Is it computationally feasible to do the comparisons ?

Parameter estimation for nonlinear non-Gaussian state space models is computationally intensive.

Large-scale comparative analyses such as Hansen and Lunde (2005) exclude parameter-driven models.

We now have numerically accelerated importance sampling method (NAIS) of Koopman, Lucas and Scharth (2013).

- NAIS is fast and numerically efficient parameter estimation for nonlinear non-Gaussian state space models: Koopman and Lit (2013)
- It requires no model-specific interventions other than the specification of the appropriate observation densities.
- NAIS can effectively be used in a Monte Carlo analysis.
- NAIS can also efficiently compute out-of-sample forecasts of time-varying parameters: the prime focus of our study.

Main findings

Findings I : based on nine model specifications and for the loss in mean square error. We also consider the model confidence sets (MCS) of Hansen, Lunde and Nason (2011).

- When the DGP is a state space model, the predictive accuracy of the misspecified GAS model is similar to that of the correctly specified state space model.
- Especially for GAS observation density that allows for heavy tails and overdispersion : the loss is smaller than 1% most of the time and never higher than 2.5%
- For the state space DGPs, the GAS model lies in the 90% model confidence set for at least 60% of the samples with as many as 2,000 observations.
- An observation-driven alternative to a parameter-driven model is available that is accurate in forecasting and easy to estimate.

Main findings

Findings II : based on nine model specifications and for the loss in mean square error. We also consider the model confidence sets (MCS) of Hansen, Lunde and Nason (2011).

- Score models outperform many of the familiar observation-driven models (ACM: GARCH, ACD, MEM)
- Score models capture additional information in the data that is not exploited by ACM models.
- Score models are therefore effective new tools for forecasting !

State space model:

We assume that y_t is generated by

$$y_t|\theta_t \sim p(y_t|\theta_t;\psi), \qquad \theta_t = \Lambda(\alpha_t), \qquad t = 1,\dots,n,$$

where θ_t is the time-varying parameter vector, $\Lambda(\cdot)$ is the link function, and scalar α_t has a linear dynamic specification.

The static parameter vector ψ incorporates additional fixed and unknown coefficients.

The state space model has updating equation

$$\alpha_{t+1} = \delta + \phi\alpha_t + \eta_t, \qquad \alpha_1 \sim \mathrm{N}(a_1, P_1), \qquad \eta_t \sim \mathrm{N}(0, \sigma_\eta^2),$$

where δ is a constant and ϕ is the autoregressive coefficient.

State space model:

Interesting examples of the state space model specifications include:

- *the stochastic volatility model*
  Tauchen and Pitts (1983), Taylor (1986), Melino and Turnbull (1990) and Ghysels, Harvey and Renault (1996),
- *the stochastic conditional duration model*
  Bauwens and Veredas (2004),
- *the stochastic conditional intensity model*
  Bauwens and Hautsch (2006),
- *the stochastic copula model*
  Hafner and Manner (2012),
- *the non-Gaussian unobserved components time series model*
  Durbin and Koopman (2000).

The GAS model

The observation-driven score model has updating equation

$$\alpha_{t+1} = d + a\, s_t + b\, \alpha_t,$$

where d, a and b are fixed coefficients and $s_t = s_t(\alpha_t, \mathcal{F}_t; \psi)$ is the driving mechanism with $\mathcal{F}_t$ being information set up to time t. The score is

$$s_t = S_t(\alpha_t) \cdot \nabla_t(\alpha_t), \quad \nabla_t(\alpha_t) = \frac{\partial \ln p(y_t \,|\, \alpha_t \,, \mathcal{F}_t \,; \psi)}{\partial \alpha_t},$$

where we take the scaling matrix as $S_t(\alpha_t) = \mathcal{I}_t(\alpha_t)^{-1/2}$

The parameter α_{t+1} is updated in the direction of steepest increase of the log-density at time t.

This update is a martingale difference under correct model specification.

Autoregressive conditional moment model

It is the same updating equation

$$\alpha_{t+1} = d + a\, s_t + b\, \alpha_t,$$

where d, a and b are fixed coefficients.

Here s_t is taken such that

$$E\left[s_t | \mathcal{F}_{t-1}\right] = \theta_t = \alpha_t.$$

We refer to this class as autoregressive conditional moment (ACM) models.

Intuitive notion : α_t should increase (or decrease) if the realised value for s_t is higher (or lower) than its conditional expectation.

It includes GARCH, ACD, MEM, etc.

Dynamic model specifications

Observation densities

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	$\frac{\Gamma(\frac{\nu+2}{2})\Gamma(\frac{\nu}{2})}{\Gamma(\frac{\nu+1}{2})} \frac{\frac{1}{\sqrt{1-\rho_t^2}}\left[1+\frac{z_{1t}^2+z_{2t}^2-2\rho_t z_{1t} z_{2t}}{\nu(1-\rho_t^2)}\right]^{-\frac{\nu+2}{2}}}{\prod_{i=1}^2 (1+z_{it}/\nu)^{-\frac{\nu+1}{2}}}$	$\rho_t = \frac{1-\exp(-\alpha_t)}{1+\exp(-\alpha_t)}$

Dynamic model specifications

Observation-driven model updates

Distribution	GAS $\nabla_t(\theta_t)$	GAS $\mathcal{I}_t(\theta_t)$	ACM 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	$\frac{1}{2\sigma_t^2}\left(\frac{\omega_t y_t^2}{\sigma_t^2} - 1\right)$, $\omega_t = \frac{\nu+1}{(\nu-2)+y_t^2/\sigma_t^2}$	$\frac{\nu}{2(\nu+3)\sigma_t^4}$	y_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	$\frac{(1+\rho^2)(\omega_t\hat{z}_{1,t}-\rho_t)-\rho_t(\omega_t\hat{z}_{2,t}-2)}{(1-\rho^2)^2}$, $\omega_t = \frac{\nu+2}{\nu+\frac{\hat{z}_{2,t}-2\rho_t\hat{z}_{1,t}}{1-\rho^2}}$	$\frac{(\nu+2+\nu\rho_t^2)}{(\nu+4)(1-\rho_t^2)^2}$	$z_{1,t}z_{2,t}$

Weibull-gamma model

Consider the Weibull distribution

$$p(y_t|\gamma_t; k_1) = \gamma_t\, k_1\, y_t^{k_1-1} \exp(-\gamma_t\, y_t^{k_1}),$$

where k_1 is a shape coefficient and γ_t is a time-varying scaling variable.

$$\mathbb{E}(y_t|\gamma_t, k) = \gamma_t^{-1/k_1}\, \Gamma(1/k_1 + 1).$$

Let $\gamma_t = \mu_t\, \nu_t$ where $\alpha_t = \log(\mu_t) \sim \text{GAS}$, $\nu_t \sim iid\, \Gamma(k_2^{-1}, k_2)$, with

$$p(\nu_t; k_2) = \frac{\nu_t^{k_2^{-1}-1} e^{-\nu_t/k_2}}{\Gamma(k_2^{-1}) k_2^{k_2^{-1}}}, \quad E(\nu_t) = 1, \quad Var(\nu_t) = k_2 < \infty.$$

The Weibull-gamma mixture or Burr density $p(y_t|\mu_t; k)$ is

$$\int_0^\infty p(y_t|\mu_t, \nu_t; k_1) p(\nu_t)\, \mathrm{d}\, \nu_t = \mu_t k_1 y_t^{k_1-1} \left(1 + k_2 \mu_t y_t^{k_1}\right)^{-(1+k_2^{-1})}$$

Also see Lancaster (1979), Grammig and Maurer (2000) and Andres and Harvey (2012).

Weibull-gamma model

We notice that $E(y_t|\mu_t; k_1, k_2)$ is

$$\mu_t^{-k_1^{-1}} \Gamma(k_1^{-1} + 1) \, \mathbb{E}\left(\nu_t^{-k_1^{-1}}\right) = (\mu_t k_2)^{-k_1^{-1}} \frac{\Gamma(k_2^{-1} - k_1^{-1})}{\Gamma(k_2^{-1})}.$$

We need to impose $0 < k_2 < k_1$ so that $\Gamma(k_2^{-1} - k_1^{-1})$ exists.

$$\nabla_t = \frac{1}{\mu_t} - (1 + k_2) \frac{y_t^{k_1}}{1 + k_2 \mu_t y_t^{k_1}}, \qquad \mathcal{I}_t^{-1} = \mu_t^2 (1 + 2k_2),$$

This update recovers the Weibull model when $k_2 \to 0$.

The scaled score is

$$s_t = \mathcal{I}_t^{-1/2} \nabla_t = \sqrt{1 + 2k_2} \left(1 - (1 + k_2) \frac{\mu_t y_t^{k_1}}{1 + k_2 \, \mu_t \, y_t^{k_1}} \right).$$

By setting $k_1 = 1$ above, the specification specialises to the exponential-gamma GAS model.

Weibull-gamma model

The effect of the mixture model

Next figures illustrate the probability density function and the GAS updates for the Weibull ($k_2 = 0$) and the Weibull-gamma mixture model ($k_2 = 0.5$) for $k_1 = 1.2$ and $\mu_t = 0.5$.

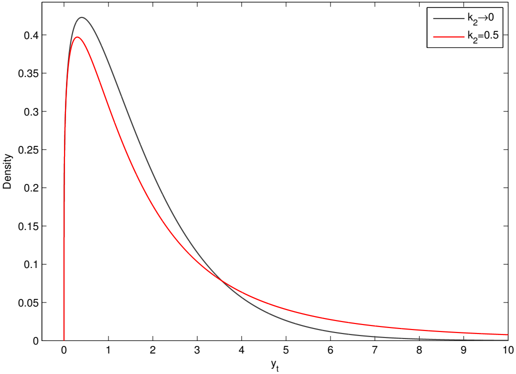
Panel (a) shows that the mixture density function significantly stretches the right tail of the distribution.
Panel (b) shows that realisations of y_t in the right tail of the distribution have limited additional impact on s_t in mixture model.

This property contrasts sharply to the corresponding ACM model where the update for the conditional mean is linear in y_t irrespective of the value of the mixture variance k_2.

We can do similar mixtures, eg for the Gamma :
the Gamma-Gamma mixture.

Weibull-gamma model

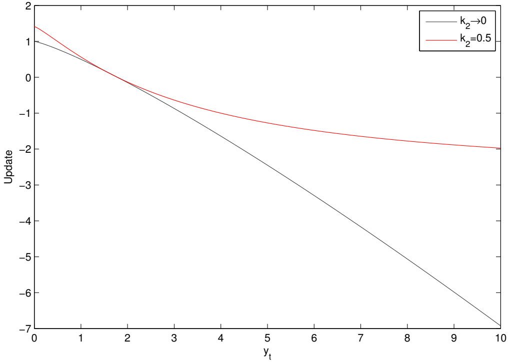
Observation-driven continuous mixture models



Panel (a) mixture density function it stretches the right tail

Weibull-gamma model

Observation-driven continuous mixture models



Panel (b) score function right tail has limited impact on y_t

Monte Carlo Study

Design : the GDPs

Model Type	Distribution	State Space, GAS δ, d	ϕ, b	σ_η, a	other
Count	Poisson	0.00	0.98	0.15	
Count	Neg. Binomial	0.00	0.98	0.15	$k_1 = 4$
Intensity	Exponential	0.00	0.98	0.15	
Duration	Gamma	0.00	0.98	0.15	$k_1 = 1.5$
Duration	Weibull	0.00	0.98	0.15	$k_1 = 1.2$
Volatility	Gaussian	0.00	0.98	0.15	
Volatility	Student's t	0.00	0.98	0.15	$\nu = 10$
Copula	Gaussian	0.02	0.98	0.10	
Copula	Student's t	0.02	0.98	0.10	$\nu = 10$

Monte Carlo Study

We consider these nine observation densities.
The autoregressive state equation completes the specifications of all parameter-driven models.

We draw $1,000$ time series realisations, $n = 4,000$ for each DGP.
In each simulation, we use the first $2,000$ observations to estimate the parameters for the following model specifications.

1. the correctly specified state space model;
2. the GAS model based on the same conditional observation density as the DGP
3. the ACM model for the corresponding specification;
4. in the case of the exponential, gamma, Weibull, and Gaussian models, a robust variant of the GAS and ACM specification.

We compute one-step ahead predictions for the next 2,000 values of θ_t given the parameter values estimated from the first 2,000 observations y_t.

We therefore consider two million $(2,000 \times 1,000)$ forecasts for each specification.

We measure the accuracy by means of the mean squared error (MSE), in levels and relative to the MSE of the state space model.

We compute the MSE across the two million forecasts of θ_t.

Monte Carlo Study

Results : state space is DGP

Relative mean-square error

Distribution	State Space		GAS		ACM	
	True	Estimated	(1)	(2)	(1)	(2)
Poisson	0.987	1.000	—	1.005	—	1.059
Neg. Binomial	0.982	1.000	—	1.008	—	1.030
Exponential	0.979	1.000	1.022	1.200	1.117	1.260
Gamma	0.985	1.000	1.004	1.050	1.033	1.032
Weibull	0.981	1.000	1.005	1.057	1.040	1.023
Gaussian	0.973	1.000	1.009	1.203	1.041	1.038
Student's t	0.968	1.000	—	1.004	—	1.145
Gaussian cop	0.957	1.000	—	1.014	—	1.312
Student's t cop	0.946	1.000	—	1.006	—	1.430

Monte Carlo Study

Results : GAS is DGP

Distribution	Relative mean-square error			Mean-square error		
	State Space	GAS	ACM	State Space	GAS	ACM
Poisson	2.888	1.000	9.187	0.012	0.004	0.038
Neg. Binomial	1.192	1.000	3.838	0.008	0.006	0.024
Exponential	5.849	1.000	4.959	0.048	0.008	0.041
Gamma	6.026	1.000	3.181	0.123	0.020	0.065
Weibull	7.614	1.000	5.217	0.050	0.007	0.034
Gaussian	8.039	1.000	6.253	0.180	0.022	0.140
Student's t	1.994	1.000	3.426	0.057	0.029	0.098
Gaussian cop	1.540	1.000	3.812	0.002	0.002	0.006
Student's t cop	1.175	1.000	5.490	0.002	0.002	0.010

Data

We have daily and high-frequency prices for twenty stocks from the Dow Jones index (January 1993 – June 2012) and five major stock indices between (January 1996 – October 2012).

Parameter estimation for all eight models is based on daily close-to-close returns.

We compute one-step ahead forecasts starting in 2001 and 2004 for the stocks and indices.

For each model the parameters are re-estimated every three months, in an expanding window including all previous daily returns.

The precision of the forecasts from a model is evaluated by comparing the volatility forecasts with the daily realised volatilities as measured from high-frequency data.

Models

1. SV
2. GAS
3. GARCH
4. EGARCH
5. SV with leverage
6. GAS with leverage
7. GJR : GARCH with leverage
8. EGARCH with leverage

Relative variance of the residuals of
Mincer-Zarnowitz regressions of the realised volatilities

Stock/index	No leverage			Leverage			
	SV	GAS	GARCH	SV	GAS	GJR	EGARCH
Am Exp	1.08	1.08	1.09	1.00	0.99	1.02	0.99
Boeing	1.07	1.06	1.13	1.00	0.99	1.04	1.00
Chevron	1.12	1.13	1.21	1.00	1.00	1.20	1.00
Disney	1.13	1.19	1.18	1.00	1.05	1.09	1.10
GE	1.06	1.04	1.06	1.00	0.99	1.01	1.01
IBM	1.12	1.11	1.23	1.00	0.98	1.11	1.00
JPMorgan	1.07	1.09	1.07	1.00	1.02	1.09	1.02
Coca-Cola	1.07	1.06	1.13	1.00	0.99	1.09	1.02
P & G	1.06	1.07	1.06	1.00	0.99	1.04	0.99
AT&T	1.06	1.06	1.11	1.00	1.03	1.08	1.04
DAX 30	1.27	1.26	1.27	1.00	1.01	1.14	0.99
FTSE 100	1.20	1.16	1.22	1.00	1.06	1.16	1.08
NASDAQ	1.20	1.20	1.21	1.00	0.99	1.01	1.00
S&P 500	1.28	1.30	1.35	1.00	1.04	1.22	1.05
Best model	0.00	0.00	0.00	0.48	0.36	0.00	0.16

Conclusions

GAS prediction for time-varying parameters is effective and convincing compared to parameter driven alternatives

Methodology for Credit and Systemic Risk Detection

D. Creal, B. Schwaab, S.J. Koopman & A. Lucas

- Economic time series often share common features, e.g. business cycle dynamics.
- Economic time series may be continuous and/or discrete and be observed at different frequencies.
- We introduce
  observation-driven mixed measurement panel data models
- The approach allows for non-linear, non-Gaussian models with common factor across different distributions.
- Application: we develop a models for credit ratings transitions and loss-given-default (LGDs) with macro factors.
- The models include:
  1. Time-varying Gaussian model
  2. Time-varying ordered logit
  3. Time-varying beta distribution

Non-Gaussian Dynamic Factors: Credit Risk Application

- Economic time series often share common features, e.g. business cycle dynamics.
- Economic time series may be continuous and/or discrete and be observed at different frequencies.
- We introduce
  observation-driven mixed measurement panel data models
- The approach allows for non-linear, non-Gaussian models with common factor across different distributions.
- Application: we develop a models for credit ratings transitions and loss-given-default (LGDs) with macro factors.
- The models include:
  1. Time-varying Gaussian model
  2. Time-varying ordered logit
  3. Time-varying beta distribution

Mixed measurement panel data models

We introduce mixed measurement observation driven models

$$
\begin{aligned}
y_{it} &\sim p_i(y_{it}|f_t, Y_{t-1}; \psi), \quad i = 1, \dots, N, \\
f_{t+1} &= \omega + B f_t + A s_t
\end{aligned}
$$

The score function is

$$
\begin{aligned}
s_t &= S_t \nabla_t \\
\nabla_t &= \sum_{i=1}^{N} \delta_{it} \nabla_{i,t} = \sum_{i=1}^{N} \delta_{it} \frac{\partial \log p_i(y_{it}|f_t, Y_{t-1}; \psi)}{\partial f_t},
\end{aligned}
$$

- The observations y_{it} may come from different distributions.
- The factors f_t may be common across distributions.
- KEY: The score function allows us to pool information from different observations to estimate the common factor f_t.
- δ_{it} is an indicator function equal to 1 if y_{it} is observed and zero otherwise. Missing values are naturally taken into account.

Scaling matrix

Consider the eigenvalue-eigenvector decomposition of Fisher's (conditional) information matrix

$$\mathcal{I}_t = \mathbb{E}_{t-1}[\nabla_t \nabla_t'] = U_t \Sigma_t U_t',$$

The scaling matrix is then defined as

$$S_t = U_t \Sigma_t^{-1/2} U_t'$$

- S_t is then the "square root" of a generalized inverse.
- The innovations s_t driving f_t have an identity covariance matrix, when the info. matrix is non-singular.
- The conditional information matrix is additive for our models:

$$\mathcal{I}_t = \mathbb{E}_{t-1}[\nabla_t \nabla_t'] = \sum_{i=1}^{N} \delta_{it} \, \mathbb{E}_{i,t-1}[\nabla_{it} \nabla_{it}'].$$

Log-likehood function and ML estimation

- The log-likelihood function for an observation-driven model can easily be computed.
- The ML estimator is

$$\hat{\psi} \quad = \quad \arg\max_{\psi} \sum_{t=1}^{T} \sum_{i=1}^{N} \delta_{it} \log p_i(y_{it} | f_t, Y_{t-1}; \psi),$$

- Estimation is similar to a GARCH model.

Credit risk

- Growing econometrics literature on models for credit risk: McNeil et al. (2005), Bauwens and Hautsch (JFEct, 2006), Gagliardini and Gourieroux (JFEct, 2005), Koopman Lucas and Monteiro (JEct, 2008), Duffie et al. (JFE, JoF 2008).
- Basic observations:
  1. Probability of default varies over time with the business cycle.
  2. Conditional on default, the loss (recovery rate) varies with the business cycle.
  3. We observe excess clustering of defaults and ratings transitions beyond what can be explained by simply adding covariates.
  4. The literature focuses on a credit risk or frailty factor.
- Industry standard models are too simple to capture these features.
- New models in the literature are parameter driven models requiring simulation methods for estimation.
- We provide observation driven alternatives.

Data: Moody's and FRED

- We observe data from Jan. 1980 to March 2010.
- $7,505$ companies are rated by Moody's.
- We pool these into 5 ratings categories (IG, BB, B, C, D).
- We observe transitions, e.g. IG $\rightarrow$ BB or C $\rightarrow$ D
- There are $J = 16$ total types of transitions.
- 19,450 total credit rating transitions.
- 1,342 transitions are defaults.
- 1,125 measurements of loss-given default (LGD).
- LGD is the fraction of principal an investor loses when a firm defaults.
- We also observe six macroeconomic variables: industrial production growth, credit spread, unemployment, annual S&P500 returns, realized volatility, real GDP growth (qtly).

Models

- Credit ratings can be modeled using the (static) ordered probit model of CreditMetrics; one of the current industry standards, see Gupton Stein (2005).
- LGD's are often modeled by (static) beta distributions.
- GOAL: Build models that improve on current industry standards and are (relatively) easy to implement and estimate.
    1. Time-varying Gaussian model
    2. Time-varying ordered logit
    3. Time-varying beta distribution
- Forecasting credit risk.
- Simulation of loss distributions and scenario analysis.
- Bank executives and regulators and can use them for "stress testing."

Mixed measurement model for credit risk

$$
\begin{aligned}
y_t^m &\sim \mathrm{N}\left(\mu_t, \Sigma_m\right) \\
y_{i,t}^c &\sim \text{Ordered Logit}\left(\pi_{ijt}, j \in \{\text{IG, BB, B, C, D}\}\right), \\
y_{k,t}^r &\sim \text{Beta}\left(a_{kt}, b_{kt}\right), \qquad k = 1, \ldots, K_t,
\end{aligned}
$$

- y_t^m are the macro variables.
- $y_{i,t}^c$ are indicator variables for each credit rating j for firm i.
- $y_{k,t}^r$ are the LGDs for the k-th default.
- K_t are the number of defaults in period t.
- μ_t, π_{ijt}, and (a_{kt}, b_{kt}) are functions of an $M \times 1$ vector of factors f_t.

Time varying Gaussian model for macro data

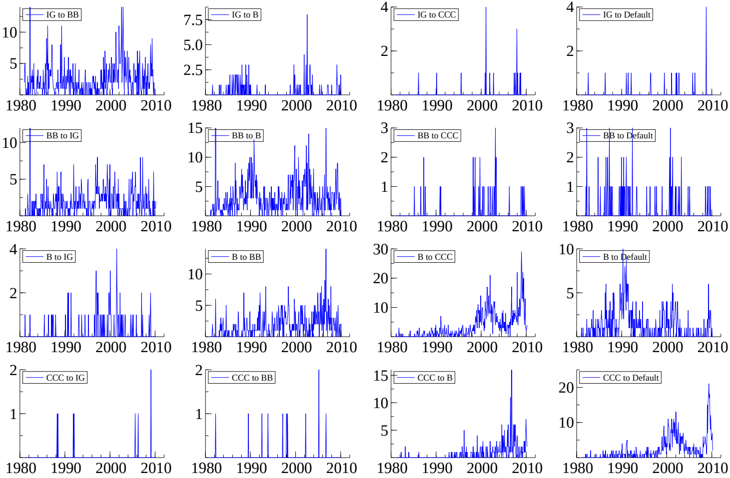
$$
\begin{aligned}
y_t^m &\sim \mathrm{N}\left(\mu_t, \Sigma_m\right), \\
\mu_t &= Z^m f_t.
\end{aligned}
$$

- Z^m is a $(6 \times M)$ matrix of factor loadings.
- Σ_m is a (6×6) diagonal covariance matrix.
- $\tilde{S}_t$ is a selection matrix indicating which macro variables are observed at time t.

$$
\begin{aligned}
\nabla_t^m &= \left(\tilde{S}_t Z^m\right)' \left(\tilde{S}_t \Sigma_m \tilde{S}_t'\right)^{-1} \tilde{S}_t \left(y_t^m - \mu_t\right), \\
\mathcal{I}_t^m &= \left(\tilde{S}_t Z^m\right)' \left(\tilde{S}_t \Sigma_m \tilde{S}_t'\right)^{-1} \tilde{S}_t Z^m.
\end{aligned}
$$

Moody's monthly credit ratings transitions

The data have been pooled together each month.



Time-varying ordered logit

$$
\begin{aligned}
y_{i,t}^c &\sim \text{Ordered Logit}\left(\pi_{ijt}, j \in \{\text{IG, BB, B, C, D}\}\right), \\
\pi_{ijt} &= \mathrm{P}\left[R_{i,t+1} = j\right] = \tilde{\pi}_{ijt} - \tilde{\pi}_{i,j-1,t}, \\
\tilde{\pi}_{ijt} &= \mathrm{P}\left[R_{i,t+1} \leq j\right] = \frac{\exp(\theta_{ijt})}{1 + \exp(\theta_{ijt})}, \\
\theta_{ijt} &= z_{ijt}^c - Z_{it}^{c\prime} f_t.
\end{aligned}
$$

- $J^c = 5$ categories $j \in \{\text{IG, BB, B, C, D}\}$.
- R_{it} is the rating for firm i at the start of month t.
- y_{it}^c is an indicator variable for each rating type.
- π_{ijt} is the probability that firm i is in category j.
- $\tilde{\pi}_{i,\text{D},t} = 0$ and $\tilde{\pi}_{i,\text{IG},t} = 1$.
- A time-varying ordered logit model is a new concept.

Time-varying ordered logit

The contribution to the log-likelihood at time t is

$$\ln p_i(y_{it}^c | f_t, Y_{t-1}; \psi) = \sum_{i=1}^{N_t} \sum_{j=1}^{J^c} y_{ijt}^c \log(\pi_{ijt})$$

The score and information matrices are

$$\begin{aligned} \nabla_t^c &= -\sum_{i=1}^{N_t} \sum_{j=1}^{J^c} \frac{y_{ijt}^c}{\pi_{ijt}} \cdot \dot{\pi}_{ijt} \cdot Z_{it}^c, \\ \mathcal{I}_t^c &= \sum_{i=1}^{N_t} n_{it} \left(\sum_j \frac{\dot{\pi}_{ij,t}^2}{\pi_{ij,t}} \right) Z_{it}^c Z_{it}^{c\prime} \end{aligned}$$

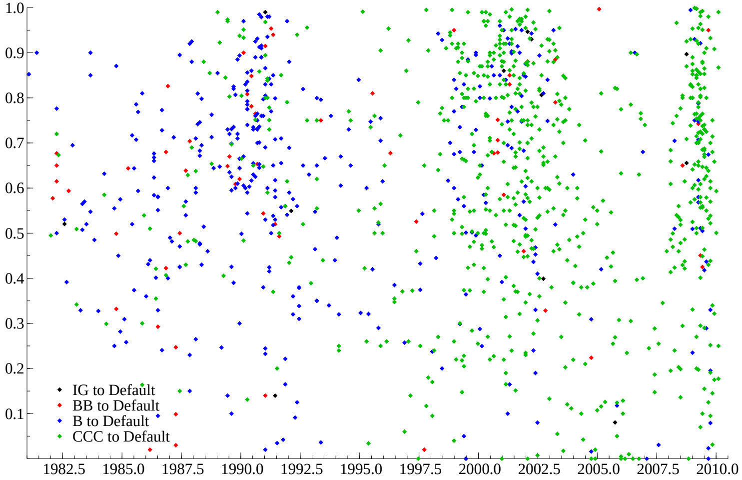
where

$$\dot{\pi}_{ijt} = \tilde{\pi}_{ijt}(1 - \tilde{\pi}_{ijt}) - \tilde{\pi}_{i,j-1,t}(1 - \tilde{\pi}_{i,j-1,t}).$$

Loss given default

- When a firm defaults, investors typically lose a fraction of their investment (alternatively, they recover a fraction of their investment).
- The fraction of losses experienced by investors also varies with the business cycle.
- A model for a time-varying beta distribution is developed.
- See McNeil and Wendin (2007 JEmpFin) for Bayesian inference in a state space model.

Loss given default by transition type



Time-varying beta distribution

$$y^r_{k,t} \sim \text{Beta}\left(a_{kt}, b_{kt}\right), \qquad k = 1, \ldots, K_t,$$

$$\begin{aligned} a_{kt} &= \beta_r \cdot \mu^r_{kt} \\ b_{kt} &= \beta_r \cdot (1 - \mu^r_{kt}) \end{aligned}$$

$$\log\left(\mu^r_{kt} / \left(1 - \mu^r_{kt}\right)\right) = z^r + Z^r f_t.$$

- We observe $K_t \geq 0$ defaults at time t.
- $0 < y^r_{k,t} < 1$ is the amount lost conditional on the k-th default.
- μ^r_{kt} is the mean of the beta distribution.
- z^r is the unconditional level of LGDs.
- Z^r is a $(1 \times M)$ vector of factor loadings.
- β_r is a scalar parameter

Time-varying beta distribution

The contribution to the log-likelihood at time t is

$$\begin{aligned}\ln p_i(y_{kt}^r | f_t, Y_{t-1}; \psi) &= \sum_{k=1}^{K_t} (a_{kt} - 1) \log\left(y_{kt}^r\right) + (b_{kt} - 1) \log\left(1 - y_{kt}^r\right) \\ &\quad - \log\left[B\left(a_{kt}, b_{kt}\right)\right]\end{aligned}$$

The score and information matrices are

$$\begin{aligned}\nabla_t^r &= \beta_r \sum_{k=1}^{K_t} \mu_{kt}^r (1 - \mu_{kt}^r) \left(Z^r\right)' (1, -1) \Big(\left(\log(y_{kt}^r), \log(1 - y_{kt}^r)\right)' - \dot{B}\left(a_{kt}, b_{kt}\right) \Big) \\ \mathcal{I}_t^r &= \beta_r \sum_{k=1}^{K_t} \left(\mu_{kt}^r (1 - \mu_{kt}^r)\right)^2 \left(Z^r\right)' (1, -1) \Big(\ddot{B}\left(a_{kt}, b_{kt}\right) \Big) (1, -1)' Z^r\end{aligned}$$

where

$$\sigma_{kt}^2 = \mu_{kt}^r \cdot (1 - \mu_{kt}^r)/(1 + \beta_r).$$

Estimation details

- The macro data y_t^m has been standardized.
- We consider models with $p = 1$ and $q = 1$ factor dynamics.
- For identification of the level parameters, we set $\omega = 0$ in the factor recursion:

$$f_{t+1} = A_1 s_t + B_1 f_t$$

- For identification of the factors, we also impose restrictions on Z^m, Z^c, and Z^r.
- Some parameters have been pooled for "rare" transitions; e.g., IG $\rightarrow$ D and BB $\rightarrow$ D.
- Moody's re-defined several categories in April 1982 and Oct. 1999 causing incidental re-ratings (outliers), which we handle via dummy variables for these dates.

AIC, BIC, and log-likelihoods for different models

	(2,0,0)	(2,1,0)	(2,2,0)	(3,0,0)
log-Like	-40447.9	-40199.1	-40162.8	-40056.2
AIC	81005.9	80520.1	80457.0	80242.4
BIC	81640.0	81223.0	81218.0	80991.0
	(3,1,0)	(3,2,0)	(3,1,1)	(3,2,1)
log-Like	-39817.1	-39780.8	-39812.6	**-39780.0**
AIC	79776.2	**79713.6**	79771.2	79716.0
BIC	80594.0	**80589.0**	80612.0	80615.0

The number of factors for each data type are represented by (m, c, r).

Parameter estimates for the (3,2,0) model

Macro loadings Z^m

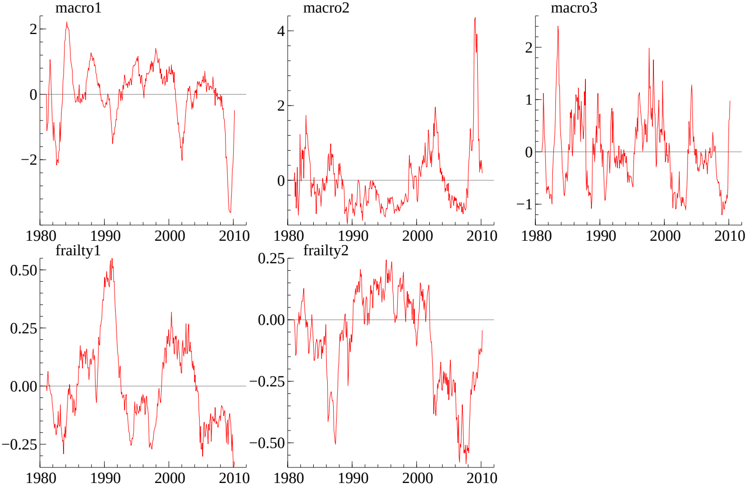
	macro_1	macro_2	macro_3	frailty_1	frailty_2
IP	1.000	0.000	0.000	0.000	0.000
	—	—	—	—	—
UR	-0.892***	0.122***	-0.062*	0.000	0.000
	(0.037)	(0.041)	(0.040)	—	—
RGDP	0.811***	0.072	0.336***	0.000	0.000
	(0.066)	(0.079)	(0.074)	—	—
Cr.Spr.	-0.169**	1.000	0.000	0.000	0.000
	(0.085)	—	—	—	—
$r_{S\&P}$	0.049	-0.268***	1.223***	0.000	0.000
	(0.093)	(0.081)	(0.093)	—	—
$\sigma_{S\&P}$	-0.007	0.648***	1.000	0.000	0.000
	(0.107)	(0.084)	—	—	—

Parameter estimates for the (3,2,0) model

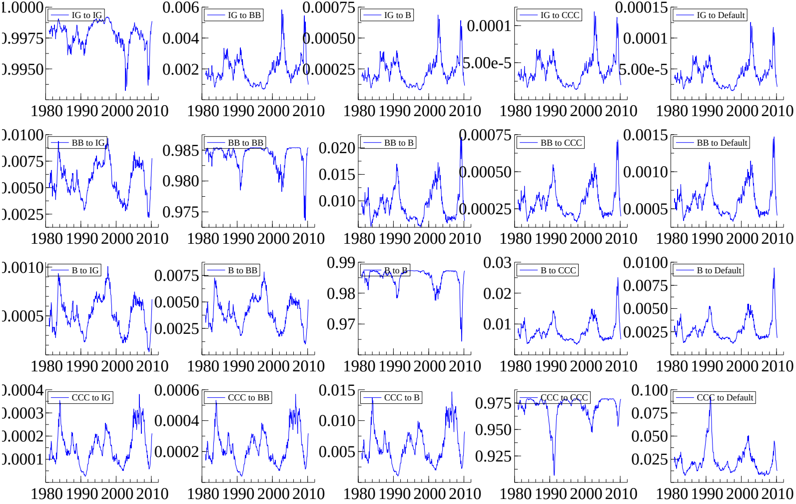
Credit rating and LGD loadings Z^c and Z^r

	macro_1	macro_2	macro_3	frailty_1	frailty_2
Z^c					
IG	-0.052	0.202***	-0.123**	1.475***	-1.165**
	(0.059)	(0.055)	(0.069)	(0.371)	(0.555)
BB	-0.078**	0.172***	-0.102***	1.000	0.000
	(0.037)	(0.037)	(0.040)	—	—
B	-0.184***	0.162***	-0.142***	0.970***	-0.016
	(0.035)	(0.031)	(0.040)	(0.156)	(0.158)
CCC	-0.262***	0.073*	-0.018	1.936***	1.000
	(0.057)	(0.050)	(0.075)	(0.465)	—
Z^r	0.018	0.276***	-0.082*	1.212***	1.065***
	(0.049)	(0.046)	(0.062)	(0.376)	(0.301)

Estimated factors for the (3,2,0) model



Time-varying transition probabilities



Volatility GAS models

- A class of volatility models is given by

$$y_t = \mu + \sigma(f_t)u_t, \qquad u_t \sim p_u(u_t;\theta), \qquad t = 1, 2, \ldots, T, \qquad (1)$$

$$f_{t+1} = \omega + \beta f_t + \alpha s_t, \qquad (2)$$

where:

- $\sigma()$ is some continuous function;
- $p_u(u_t;\theta)$ is a standardized disturbance density;
- s_t is the scaled score based on $\partial \log p(y_t | Y_{t-1}, f_t; \theta) \, / \, \partial f_t$.

- Some special cases
  - $\sigma(f_t) = f_t$ and p_u is Gaussian : GAS $\Rightarrow$ GARCH;
  - $\sigma(f_t) = \exp(f_t)$ and p_u is Gaussian : GAS $\Rightarrow$ EGARCH;
  - $\sigma(f_t) = \exp(f_t)$ and p_u is Student's t : GAS $\Rightarrow$ Beta-t-EGARCH.

Introducing FIGAS

The Fractionally Integrated Generalized Autoregressive Score (FIGAS) model is given by

$$y_t \sim p(y_t|Y_{t-1}, f_t; \theta), \qquad t = 1, 2, \dots, T, \tag{3}$$

$$f_t^* = (1-L)^d f_t, \qquad f_{t+1}^* = \omega + \beta f_t^* + \alpha s_t, \tag{4}$$

where:

- y_t denotes dependent variable; $Y_t = [y_1, \dots, y_t]'$;
- f_t is the time-varying parameter of interest;
- θ collects static parameters;
- d is the fractional integration order;
- $(1-L)^d = 1 - dL + \frac{d(d-1)}{2!}L^2 - \frac{d(d-1)(d-2)}{3!}L^3 + \dots$
- s_t is the scaled score based on $\partial \log p(y_t|Y_{t-1}, f_t; \theta) \,/\, \partial f_t$.

Some background on long memory / fractional integration

The FIGAS specification with a long memory process for $\{f_t\}$ is analogous to the ARFIMA model as in Granger & Joyeaux (1980) & Hosking (1981) for the conditional mean.

- FIGAS nests GAS for $d \equiv 0$ and Integrated GAS, or IGAS, for $d \equiv 1$;
- the $\{f_t\}$ process is stationary and invertible when $1 - \beta z \neq 0$, for $|z| < 1$ and when $-1 < d < 1/2$; see Palma (2007, Section 3.2);
- for $1/2 \leq d < 1$, the $\{f_t\}$ process is not stationary but mean-reverting;
  for $d = 1$, the $\{f_t\}$ process is not stationary and is not mean-reverting; see Baillie (1996, p.22).

General FIGAS specification

■ We introduce time-varying parameters with long memory properties in a bivariate heavy-tailed distribution for a set of stock equity returns.

- heavy-tails in returns with different tail properties;
- outliers for marginal and/or joint densities should not dillute volatility and/or correlation processes; especially relevant for long memory features;
- tail dependence is modeled explicitly.

Our approach :

■ We model marginal series by means of conditonal Student's t densities and we model dependence by means of a t copula.

■ The score function in the Student's t class of distributions depends on conditional weights that downweight extreme observations.

■ The degrees of freedom parameter for the Student's t distribution handles the level of robustness for statistical inference.

FIGAS for conditional variance

- Let y_t denote (demeaned) log-return of some asset, assume

$$y_t = \sigma_t \varepsilon_t, \qquad \varepsilon_t \sim \text{Student's } t_\nu(0,1),$$

with loglikelihood function given by

$$\ell_t = c(\nu) - \frac{1}{2}\log(\pi) - \frac{1}{2}\log(\sigma_t^2) - \frac{\nu+1}{2}\log\left(1 + \frac{y_t^2}{(\nu-2)\sigma_t^2}\right),$$

where $c(\nu) = \log\left\{\Gamma\left(\frac{\nu+1}{2}\right)/\Gamma\left(\frac{\nu}{2}\right)\right\} - \frac{1}{2}\log(\nu-2)$ and $\nu > 2$.

- Let $f_t = \log(\sigma_t^2)$, we have

$$\nabla_t = \frac{1}{2\,\sigma_t^2}\left[\omega_t\, y_t^2 - \sigma_t^2\right] \qquad \text{and} \qquad \mathcal{I}_t = \frac{1}{2}\frac{\nu}{\nu+3},$$

where

$$\omega_t = \frac{\nu+1}{\nu-2+y_t^2/\sigma_t^2} \in [0, (\nu+1)/(\nu-2)].$$

Time t weight ω_t attains zero if y_t^2 too large relative to current level of volatility.

FIGAS for conditional variance : the resulting model

The FIGAS model is then given by :

- demeaned log-return of some asset :

$$y_t = \sigma_t \varepsilon_t, \qquad \varepsilon_t \sim \text{Student's } t_\nu(0,1),$$

with loglikelihood function given by

$$\ell_t = c(\nu) - \frac{1}{2}\log(\pi) - \frac{1}{2}\log(\sigma_t^2) - \frac{\nu+1}{2}\log\left(1 + \frac{y_t^2}{(\nu-2)\sigma_t^2}\right),$$

where $\sigma_t^2 = \exp(f_t)$.

- log-variance is updated :

$$f_{t+1}^* = \omega + \beta f_t^* + \alpha s_t, \qquad f_t^* = (1-L)^d f_t,$$

where the scaled score is given by

$$s_t = \mathcal{I}_t^{-\frac{1}{2}} \nabla_t, \qquad \nabla_t = \frac{1}{2\,\sigma_t^2}\left[\omega_t\, y_t^2 - \sigma_t^2\right] \qquad \text{and} \qquad \mathcal{I}_t = \frac{1}{2}\frac{\nu}{\nu+3}.$$

- FIGAS with leverage (FIGASL) : $\alpha \Rightarrow \alpha + \gamma 1_{(y_t<0)}$.

Conditional volatility

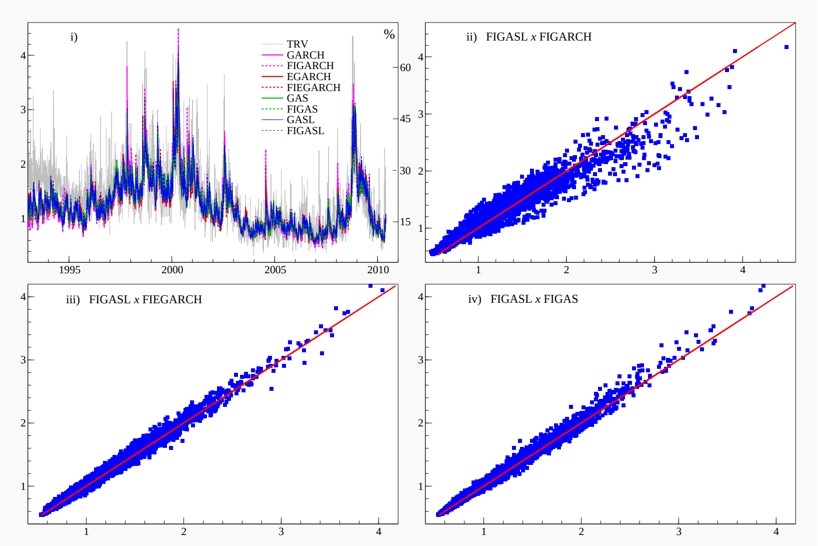


Figure 1: Estimated vol for P&G daily returns over January 4, 1993 to May 28, 2010

Robust filtering of volatility: the role of weight ω_t

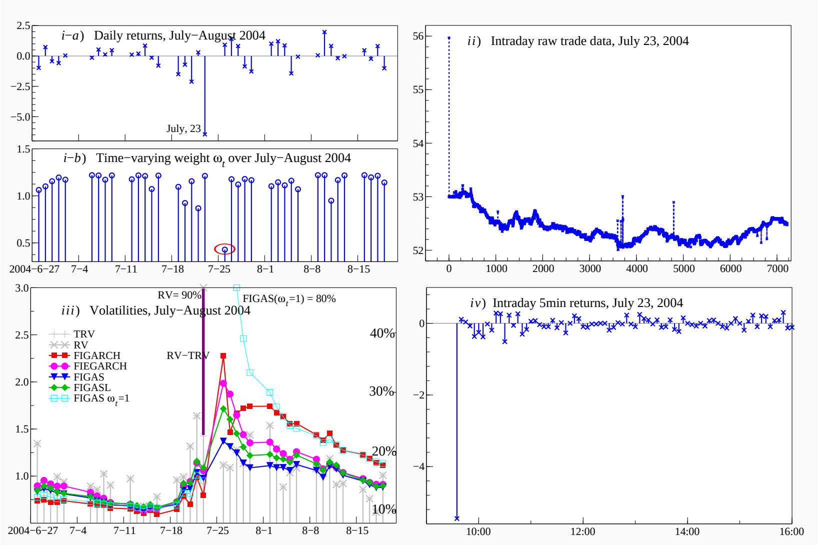


Figure 2: P&G case study

FIGAS for bivariate conditional dependence

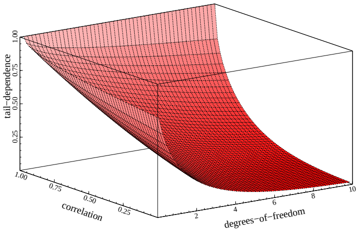
- for dependence between two marginal series : bivariate t copula

$$(1-\rho_t^2)^{-\frac{1}{2}} \frac{\Gamma(\frac{\eta+2}{2})\Gamma(\frac{\eta}{2})}{\left[\Gamma(\frac{\eta+1}{2})\right]^2} \frac{\left(1+\frac{1}{\eta(1-\rho_t^2)}\left(x_{1t}^2+x_{2t}^2-2\rho_t x_{1t}x_{2t}\right)\right)^{-\frac{\eta+2}{2}}}{\prod_{i=1}^2 \left(1+x_{it}^2/\eta\right)^{-\frac{\eta+1}{2}}},$$

where $x_{it} = t_\eta^{-1}(u_{it})\,, i=1,2,\ u_{it} \in (0,1),\ \rho_t \in (-1,1)$ and $\eta > 0$.

- t copula captures tail dependence which is governed by ρ_t and η

- extreme occurences of x_{1t} and/or x_{2t} can be due to heavy-tail nature (low η) of the t copula, not neccessarily due to high ρ_t:



FIGAS for bivariate conditional dependence

Define $f_t = \log(1+\rho_t \,/\, 1-\rho_t) \in \mathbb{R}$, we have

$$\begin{aligned}
\nabla_t &= \frac{\dot{\rho}_t}{(1-\rho_t^2)^2}\left[(1+\rho_t^2)\big(\pi_t x_{1t} x_{2t} - \rho_t\big) - \rho_t\big(\pi_t x_{1t}^2 + \pi_t x_{2t}^2 - 2\big)\right], \\
\mathcal{I}_t &= \frac{\dot{\rho}_t^2}{(1-\rho_t^2)^2}\left(1+\rho_t^2 - \frac{2\rho_t^2}{\eta+2}\right)\frac{\eta+2}{\eta+4},
\end{aligned}$$

where $\dot{\rho}_t$ is derivative of ρ_t wrt f_t, with time-dependent weight defined as

$$\pi_t = \frac{\eta+2}{\eta+m_t} \in [0, (\eta+2)/\eta],$$

where

$$m_t = x_t{}' R_t^{-1} x_t \geq 0, \quad \text{with} \quad x_t = [x_{1t}\ x_{2t}]' \quad \text{and} \quad R_t = \begin{pmatrix} 1 & \rho_t \\ \rho_t & 1 \end{pmatrix}.$$

For a finite η, extreme observations x_{1t} and/or x_{2t} leading to a large Mahalanobis distance m_t will, as the result of downweighting via π_t, have limited impact on the correlation dynamics.

FIGAS for bivariate conditional dependence

The FIGAS model for dependence is then given by :

- The t-copula is given as above with

$$\rho_t = \frac{1 - \exp f_t}{1 + \exp f_t},$$

- logit-dependence is updated :

$$f^*_{t+1} = \omega + \beta f^*_t + \alpha s_t, \qquad f^*_t = (1-L)^d f_t,$$

where the scaled score is given by

$$s_t = \mathcal{I}_t^{-\frac{1}{2}} \nabla_t,$$

where

$$\begin{aligned} \nabla_t &= \frac{\dot{\rho}_t}{(1-\rho_t^2)^2}\left[(1+\rho_t^2)\left(\pi_t x_{1t} x_{2t} - \rho_t\right) - \rho_t\left(\pi_t x_{1t}^2 + \pi_t x_{2t}^2 - 2\right)\right], \\ \mathcal{I}_t &= \frac{\dot{\rho}_t^2}{(1-\rho_t^2)^2}\left(1 + \rho_t^2 - \frac{2\rho_t^2}{\eta+2}\right)\frac{\eta+2}{\eta+4}. \end{aligned}$$

Conditional dependence

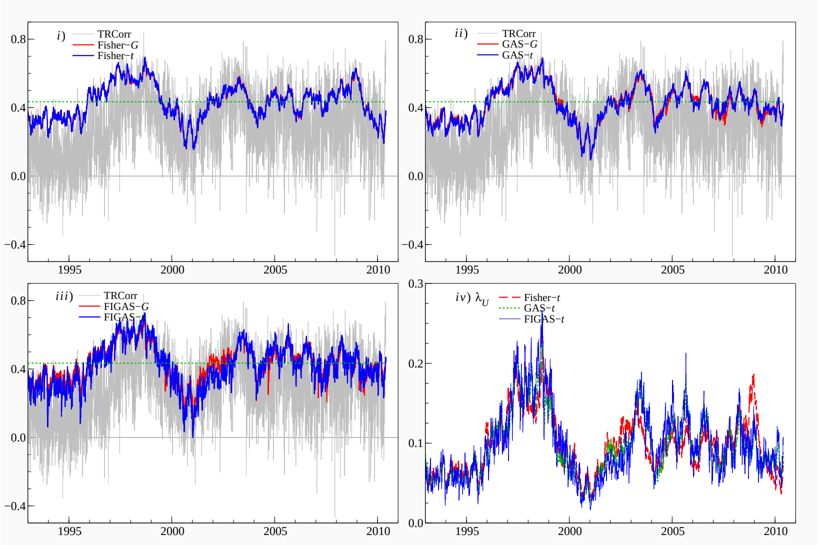


Figure 3: Estimated correlation for GE/KO daily returns over January 4, 1993 to May 28, 2010

Robust filtering of correlation: the role of π_t

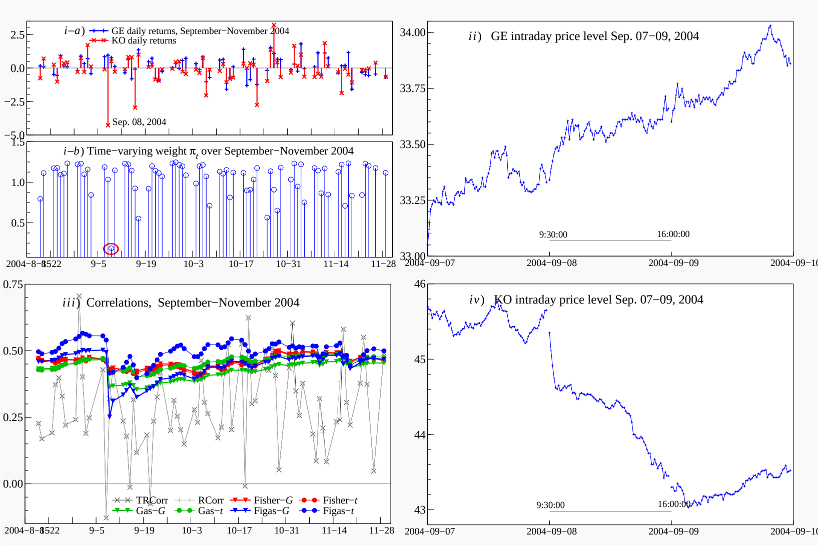


Figure 4: GE/KO case study

What have we done ?

- We have reviewed GAS models.
- Focus: modelling time-varying parameters in observation-driven approach.
- In particular we have shown that score driven models reduce to many well-established models in financial econometrics.
- Today we have shown how interesting new model formulations can be derived.

Examples:

- Forecasting with GAS models and comparisons with State Space Models
- Dynamic Factor Models with Mixed Measurements and Mixed Frequencies
- Modelling Dynamic Volatilities and Correlations using GAS models

Much more work to do !!

Generalized Autoregressive Score Models for Time-varying Parameters: new models and applications

Francisco Blasques[a], Drew D. Creal[b], Pawel Janus[c], Siem Jan Koopman[a,d], André Lucas[a], Marcel Scharth[e], Bernd Schwaab[f]

[a]Vrije Universiteit Amsterdam and Tinbergen Institute
[b]Booth School of Business University of Chicago
[c]UBS Zurich
[d]CREATES, Aarhus University
[e]University of New South Wales, Australia
[f]European Central Bank, Frankfurt am Main

Workshop ISF 2014 Rotterdam