Score-driven models for forecasting

F. Blasques S.J. Koopman A. Lucas

VU University Amsterdam, Tinbergen Institute, CREATES

Eighth ECB Workshop on Forecasting Techniques

ECB, Frankfurt am Mainz, 13-14 June 2014

We have an observed time series $y_1, \ldots, y_n$, collected in $\boldsymbol{y}$, and an unobserved path of equal length $f_1, \ldots, f_n$, collected in $\boldsymbol{f}$.

Assume that

- DGP is $\boldsymbol{y} \sim p(\boldsymbol{y}; \boldsymbol{f})$ where $\boldsymbol{f}$ represents a time-varying feature of the "true" model density.
- time series econometrician opts for a predictive model $\tilde{p}(y_t | y_1, \ldots, y_{t-1}; \boldsymbol{\theta})$, with parameter vector $\boldsymbol{\theta}$, and correctly considers the effect associated with $\boldsymbol{f}$ to be time-varying.

The basic framework

Assume that

- DGP is $\boldsymbol{y} \sim p(\boldsymbol{y}; \boldsymbol{f})$
- econometric model $\tilde{p}(y_t | y_1, \ldots, y_{t-1}; \boldsymbol{\theta})$ correctly considers the effect associated with $\boldsymbol{f}$ to be time-varying.

The basic framework

Assume that

- DGP is $\boldsymbol{y} \sim p(\boldsymbol{y}; \boldsymbol{f})$
- econometric model $\tilde{p}(y_t | y_1, \ldots, y_{t-1}; \boldsymbol{\theta})$ correctly considers the effect associated with $\boldsymbol{f}$ to be time-varying.

In an observation-driven model, the time-varying effect is extracted as a direct function of past data: $\tilde{f}_t = \tilde{f}_t(y_1, \ldots, y_{t-1}; \boldsymbol{\theta})$. We have $\tilde{p}(y_t | \tilde{f}_t, y_1, \ldots, y_{t-1}; \boldsymbol{\theta})$.

Assume that

- DGP is $\boldsymbol{y} \sim p(\boldsymbol{y}; \boldsymbol{f})$
- econometric model $\tilde{p}(y_t|y_1, \ldots, y_{t-1}; \boldsymbol{\theta})$ correctly considers the effect associated with $\boldsymbol{f}$ to be time-varying.

In an observation-driven model, the time-varying effect is extracted as a direct function of past data: $\tilde{f}_t = \tilde{f}_t(y_1, \ldots, y_{t-1}; \boldsymbol{\theta})$. We have $\tilde{p}(y_t|\tilde{f}_t, y_1, \ldots, y_{t-1}; \boldsymbol{\theta})$.

Consider GARCH model where $\boldsymbol{f}$ is the time-varying variance, then we have $\tilde{p}(y_t|\tilde{f}_t; \boldsymbol{\theta}) \equiv N(0, \tilde{f}_t)$ and

$$\tilde{f}_{t+1} = \omega + \beta \tilde{f}_t + \alpha(y_t^2 - \tilde{f}_t),$$

and with $\boldsymbol{\theta} = (\omega, \alpha, \beta)'$.

The basic framework

In a general setting (Creal, Koopman & Lucas 2008, 2011, 2013),

$$\tilde{p}(y_t|\tilde{f}_t, y_1, \ldots, y_{t-1}; \boldsymbol{\theta}) \neq N(0, \tilde{f}_t),$$

we propose for $\tilde{f}_t$ to adopt the updating scheme

$$\tilde{f}_{t+1} = \omega + \beta \tilde{f}_t + \alpha s_t,$$

where s_t is the scaled score function $s_t = S_t \nabla_t$ with

$$\begin{aligned} \nabla_t &= \frac{\partial \ln p(y_t|y_1, \ldots, y_{t-1}, \tilde{f}_t; \theta)}{\partial \tilde{f}_t}, \\ S_t &= \mathcal{I}_{t-1}^{-1} = -E_{t-1}\left[\frac{\partial^2 \ln p(y_t|Y_{t-1}, \tilde{f}_t; \theta)}{\partial f_t \partial \tilde{f}_t'}\right]^{-1}. \end{aligned}$$

For this updating equation

$$\tilde{f}_{t+1} = \omega + \beta \tilde{f}_t + \alpha s_t,$$

For this updating equation

$$\tilde{f}_{t+1} = \omega + \beta \tilde{f}_t + \alpha s_t,$$

- we can view updating as a steepest ascent or Newton step for $\tilde{f}_t$ using the log conditional density $\tilde{p}(y_t|\tilde{f}_t, y_1, \ldots, y_{t-1}; \boldsymbol{\theta})$ as criterion function;
- choice for S_t may be square root matrix of the inverse Fisher information matrix;
- this S_t accounts for curvature of density as function of $\tilde{f}_t$;
- also, under correct model specification, s_t has unit variance.

Generalized Autoregressive Score (GAS)

The general GAS framework is

$$\tilde{p}(y_t|\tilde{f}_t, y_1, \ldots, y_{t-1}; \boldsymbol{\theta}),$$

where for $\tilde{f}_t$ we adopt the updating scheme

$$\tilde{f}_{t+1} = \omega + \beta \tilde{f}_t + \alpha s_t,$$

with s_t as the scaled score function $s_t = S_t \nabla_t$ and

$$\begin{aligned} \nabla_t &= \frac{\partial \ln p(y_t|y_1, \ldots, y_{t-1}, \tilde{f}_t; \theta)}{\partial \tilde{f}_t}, \\ S_t &= \mathcal{I}_{t-1}^{-1/2} = -E_{t-1}\left[\frac{\partial^2 \ln p(y_t|Y_{t-1}, \tilde{f}_t; \theta)}{\partial \tilde{f}_t \partial \tilde{f}_t'}\right]^{-1/2}. \end{aligned}$$

A class of volatility models is given by

$$y_t = \mu + \sigma(\tilde{f}_t)u_t, \qquad u_t \sim p_u(u_t; \boldsymbol{\theta}),$$

$$\tilde{f}_{t+1} = \omega + \beta\tilde{f}_t + \alpha s_t,$$

where:

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

Some special cases

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

Special cases of GAS

GAS updating for specific observation densities and scaling choices reduces to well-known models.

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

New GAS applications : `http://gasmodel.com`

Why GAS ?

- In econometrics, score and Hessian are familiar entities;
- Using contribution of score at time t of updating seems not unreasonable (quasi-Newton connection);
- Many GARCH-type time series models are effectively constructed in this way.
- In case of GARCH, the score driver has an interpretation : $\mathbb{E}_{t-1}(y_t^2) = \sigma^2$.
- In other cases (incl. GARCH with t-densities), choice of driver mechanism for updating is not so clear.
- But is the score then really a good idea ?

Why GAS ?

On the outset:

1. GAS update is very intuitive...
2. Using the score seems optimal in some sense...
3. Likelihood and KL divergence are closely related...

Question: Is the GAS update optimal in some KL sense?

Challenge 1: GAS update is surely not always correct!

Challenge 2: Comparing different observation-driven models is only interesting under misspecification with very general DGP!

DGP and Observation-Driven Model

True sequence of conditional densities:

$$\Big\{\ p(y_t|f_t)\ \Big\}\ , \qquad \textit{true}\ \text{tv parameter}\ \{f_t\}$$

Conditional densities *postulated by probabilistic model:*

$$\Big\{\ \tilde{p}\big(y_t|\tilde{f}_t;\boldsymbol{\theta}\big)\ \Big\}\ , \qquad \textit{filtered}\ \text{tv parameter}\ \{\tilde{f}_t\}$$

$\tilde{p}\big(y_t|\tilde{f}_t;\boldsymbol{\theta}\big)$ is *implicitly by observation equation:*

$$y_t = g\big(\tilde{f}_t\,,\,u_t\,;\,\boldsymbol{\theta}\big), \qquad u_t \sim p_u(\boldsymbol{\theta}),$$

Observation-driven parameter update:

$$\tilde{f}_{t+1} = \phi\big(y_t\,,\,\tilde{f}_t\,;\,\boldsymbol{\theta}\big), \qquad \forall\ t \in \mathbb{N}, \qquad \tilde{f}_1 \in \mathcal{F} \subseteq \mathbb{R},$$

Optimal Observation-Driven Update

Key objective: Characterize $\phi(\cdot)$ that possess optimality properties from information theoretic point of view.

Main Question: Is there an optimal form for the update

$$\tilde{f}_{t+1} = \phi\big(y_t\,,\, \tilde{f}_t\,;\, \boldsymbol{\theta}\big), \qquad \forall\, t \in \mathbb{N}, \qquad \tilde{f}_1 \in \mathcal{F} \subseteq \mathbb{R},$$

Answer: This depends on the notion of optimality!

Result 1: Only parameter updates based on the score always reduce the local Kullback-Leibler divergence p and $\tilde{p}$.

Result 2: The use of the score leads to considerably smaller global KL divergence in empirically relevant settings.

Note: Results hold for any DGP (any p and $\{f_t\}$)

Definitions: Local GAS Updates

GAS-update:

$$\tilde{f}_{t+1} = \phi\big(y_t\,,\ \tilde{f}_t\,;\,\boldsymbol{\theta}\big) = \omega + \alpha s(y_t, \tilde{f}_t) + \beta \tilde{f}_t, \qquad \forall\ t \in \mathbb{N},$$

Newton-GAS update: $(\ \omega = 0,\ \ \alpha > 0,\ \ \beta = 1\ \ )$

$$\tilde{f}_{t+1} = \alpha s(y_t, \tilde{f}_t) + \tilde{f}_t, \qquad \forall\ t \in \mathbb{N},$$

Local update: $\tilde{f}_{t+1}$ in neighborhood of $\tilde{f}_t$

Local optimality: Refers to local updates

Definition I: Realized KL Divergence

KL divergence between $p(\cdot|f_t)$ and $\tilde{p}\big(\cdot\,|\tilde{f}_{t+1};\boldsymbol{\theta}\big)$ is given by

$$\mathcal{D}_{\mathrm{KL}}\Big(p(\cdot|f_t)\,,\,\tilde{p}\big(\cdot\,|\tilde{f}_{t+1};\boldsymbol{\theta}\big)\Big) = \int_{-\infty}^{\infty} p(y_t|f_t)\ln\frac{p(y_t|f_t)}{\tilde{p}\big(y_t|\tilde{f}_{t+1};\boldsymbol{\theta}\big)}\,\mathrm{d}y_t.$$

Definition I: Realized KL Divergence

KL divergence between $p(\cdot|f_t)$ and $\tilde{p}(\,\cdot\,|\tilde{f}_{t+1};\boldsymbol{\theta})$ is given by

$$\mathcal{D}_{\mathrm{KL}}\Big(p(\cdot|f_t)\,,\,\tilde{p}(\,\cdot\,|\tilde{f}_{t+1};\boldsymbol{\theta})\Big) = \int_{-\infty}^{\infty} p(y_t|f_t)\ln\frac{p(y_t|f_t)}{\tilde{p}(y_t|\tilde{f}_{t+1};\boldsymbol{\theta})}\,\mathrm{d}y_t.$$

The *realized KL variation* $\Delta_{\mathrm{RKL}}^{t-1}$ of a parameter update from $\tilde{f}_t$ to $\tilde{f}_{t+1}$ is defined as

$$\Delta_{\mathrm{RKL}}^{t-1} = \mathcal{D}_{\mathrm{KL}}\Big(p(\cdot|f_t)\,,\,\tilde{p}(\,\cdot\,|\tilde{f}_{t+1};\boldsymbol{\theta})\Big) - \mathcal{D}_{\mathrm{KL}}\Big(p(\cdot|f_t)\,,\,\tilde{p}(\,\cdot\,|\tilde{f}_{t};\boldsymbol{\theta})\Big)$$

Definition II: Conditionally Expected KL Divergence

An optimal updating scheme, while subject to randomness, should have tendency to move in *correct direction*:

On average, the KL divergence should reduce in expectation.

Definition II: Conditionally Expected KL Divergence

An optimal updating scheme, while subject to randomness, should have tendency to move in *correct direction*:

On average, the KL divergence should reduce in expectation.

The *conditionally expected KL (CKL)* variation of a parameter update from $\tilde{f}_t \in \tilde{\mathcal{F}}$ to $\tilde{f}_{t+1} \in \tilde{\mathcal{F}}$ is given by

$$\Delta_{\mathrm{CKL}}^{t-1} = \int_F q(\tilde{f}_{t+1}|\tilde{f}_t, f_t; \boldsymbol{\theta}) \left[\int_Y p(y|f_t) \ln \frac{\tilde{p}(y|\tilde{f}_t; \boldsymbol{\theta})}{\tilde{p}(y|\tilde{f}_{t+1}; \boldsymbol{\theta})} \mathrm{d}y \right] \mathrm{d}\tilde{f}_{t+1},$$

where $q(\tilde{f}_{t+1}|\tilde{f}_t, f_t; \boldsymbol{\theta})$ denotes the density of $\tilde{f}_{t+1}$ conditional on both $\tilde{f}_t$ and f_t. For a given p_t, an update is CKL optimal if and only if $\Delta_{\mathrm{CKL}}^{t-1} <= 0$.

Propositions 1 and 2

all subject to some regularity conditions

Proposition 1 :

Every Newton-GAS update ($\tilde{f}_{t+1} = \alpha s_t + \tilde{f}_t$) is locally *RKL optimal* and *CKL optimal* for any true density p_t.

Propositions 1 and 2

all subject to some regularity conditions

Proposition 1 :

Every Newton-GAS update ($\tilde{f}_{t+1} = \alpha s_t + \tilde{f}_t$) is locally *RKL optimal* and *CKL optimal* for any true density p_t.

These properties can be generalized to 'score-equivalent' updates, we require a fundamental sign condition.

Propositions 1 and 2

all subject to some regularity conditions

Proposition 1 :

Every Newton-GAS update ($\tilde{f}_{t+1} = \alpha s_t + \tilde{f}_t$) is locally *RKL optimal* and *CKL optimal* for any true density p_t.

These properties can be generalized to 'score-equivalent' updates, we require a fundamental sign condition.

Proposition 2 :

For any given true density p_t, a parameter update is locally RKL optimal and CKL optimal if and only if the parameter update is *score-equivalent*, that is

$$\text{sign}(\Delta\phi(f, y; \boldsymbol{\theta})) = \text{sign}(\tilde{\nabla}(f, y; \boldsymbol{\theta}))$$

Proposition 3

For the GAS updating

$$\tilde{f}_{t+1} = \omega + \beta \tilde{f}_t + \alpha s_t,$$

we require specific and different conditions for ω, β and α to state

Proposition 3 :

The GAS update is locally RKL optimal and CKL optimal, for every p_t.

Example RKL for GARCH

The GAS volatility model with a normal distribution reduces to the standard GARCH model.

Then the necessary condition is $\alpha\,|y_t^2 - \tilde{f}_t| > |\omega + (\beta - 1)\tilde{f}_t|$:

It follows that the update becomes convincingly more RKL optimal if the observed y_t^2 deviates considerably from the filtered volatility $\tilde{f}_t$.

Application: Volatility Model

Data Generating Process:

$$y_t = \sqrt{f_t} u_t, \qquad u_t \sim \tau(\lambda),$$

$$\log f_t = a + b \log f_{t-1} + \epsilon_t, \qquad \epsilon_t \sim \text{NID}(0, \sigma_\epsilon^2),$$

Parameter Values: $a = 0$, $b = 0.98$, $\sigma_\epsilon = 0.065$, $\lambda \in [2, 8]$.

Compared Models: GARCH, t-GARCH and t-GAS.

$$y_t = \sqrt{\tilde{f}_t} u_t$$

$$\text{(GARCH)} \quad \tilde{f}_{t+1} = \omega + \alpha y_t^2 + \beta \tilde{f}_t, \quad u_t \sim N(0, \sigma^2)$$

$$\text{(t-GARCH)} \quad \tilde{f}_{t+1} = \omega + \alpha y_t^2 + \beta \tilde{f}_t, \quad u_t \sim \tau(\nu)$$

$$\text{(t-GAS)} \quad \tilde{f}_{t+1} = \omega + \alpha s(y_t, \tilde{f}_t) + \beta \tilde{f}_t, \quad u_t \sim \tau(\nu)$$

Application: Volatility Model

Data Generating Process:

$$y_t = \sqrt{f_t} u_t, \qquad u_t \sim \tau(\lambda),$$
$$\log f_t = a + b \log f_{t-1} + \epsilon_t, \qquad \epsilon_t \sim \text{NID}(0, \sigma_\epsilon^2),$$

Parameter Values: $a = 0$, $b = 0.98$, $\sigma_\epsilon = 0.065$, $\lambda \in [2, 8]$.

Note: Comparison at pseudo-true parameters $\boldsymbol{\theta}^* = \arg\min KL$

Sample size T: Large enough for ML estimator to converge to pseudo-true parameter (at 3rd decimal place).

Asymptotic Theory: Convergence of ML estimator to pseudo-true parameter in misspecified GAS, see BKL (2013).

The pseudo-true parameters

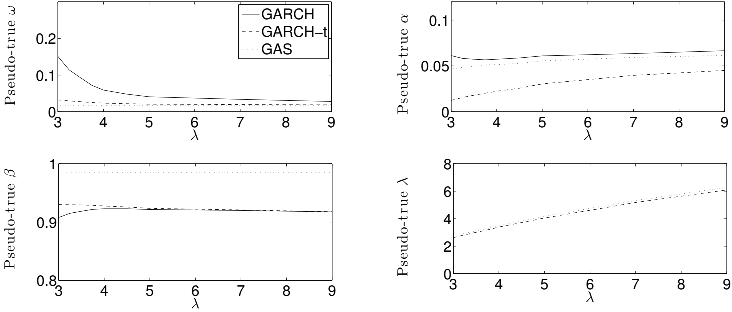

Figure: DGP with $a = 0$, $b = 0.98$, $\sigma_\epsilon = 0.065$ and $\lambda \in [3, 8]$. Each model estimated separately for true λ by MLE for simulated series, $T = 35,000$.

Relative KL divergence

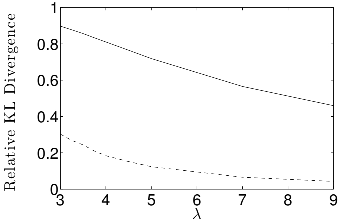

Figure: Relative KL divergence of GAS-t relative to GARCH (solid) and GARCH-t (dashed): 1-KL(GAS-t)/KL(GARCH)

Relative KL divergence

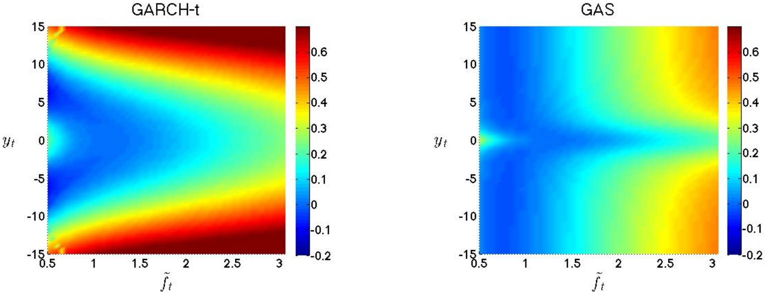

Figure: RKL optimality regions for GAS-t and GARCH-t: $\lambda = 3$ and true $f_t \approx 1.2$.

Conditional Expected KL divergence

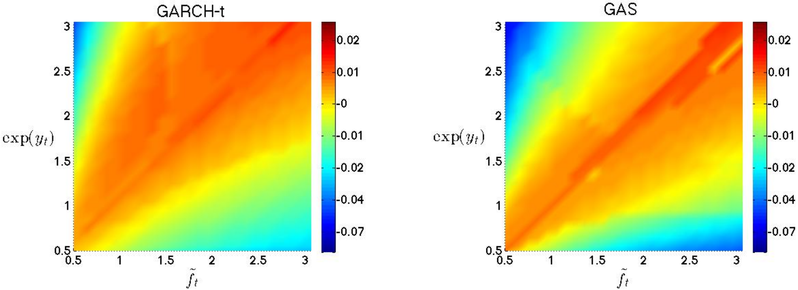

Figure: CKL variation for GAS-t and GARCH-t: $\lambda = 3$.

What about actual forecasting ?

Forecasting the time-varying feature in the model is of key importance for the forecasting of the time series y_t.

This is the study of Koopman, Lucas and Scharth (2014) and the results are presented next.

Monte Carlo study

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

Monte Carlo study

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	$\frac{(1+\rho^2)(\omega_t\hat{z}_{1,t}-\rho_t)-\rho_t(\omega_t\hat{z}_{2,t}-2)}{(1-\rho^2)^2}$	$\frac{\left(\nu+2+\nu\rho_t^2\right)}{(\nu+4)(1-\rho_t^2)^2}$	$z_{1,t}z_{2,t}$
	$\omega_t = \frac{\nu+2}{\nu+\frac{\hat{z}_{2,t}-2\rho_t\hat{z}_{1,t}}{1-\rho^2}}$		

Monte Carlo study

Model	Distribution	State Space, GAS			
Type		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.

Monte Carlo study

We compute one-step ahead predictions for the next 2,000 values of $\tilde{f}_t$ 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 $\tilde{f}_t$.

Simulation results

DGP by state space model

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

DGP by GAS

Distribution	Relative mean-square error			Mean-square error		
	St Sp	GAS	ACM	St Sp	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

Empirical Study

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, parameters are re-estimated every three months, expanding window, incl. 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 in empirical study

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

Empirical result

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
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
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