

Generalized autoregressive score models with applications *

Drew Creal^a, Siem Jan Koopman^{b,d}, André Lucas^{c,d}

^(a) *University of Chicago, Booth School of Business*

^(b) *Department of Econometrics, VU University Amsterdam*

^(c) *Department of Finance, VU University Amsterdam, and Duisenberg school of finance*

^(d) *Tinbergen Institute, Amsterdam*

August 10, 2010

Abstract

We propose a class of observation driven time series models called generalized autoregressive score models for which the scaled score of the log-density is the driving mechanism. This approach provides a unified framework for the introduction of time-varying parameters in a wide class of non-linear models. Generalized autoregressive score models encompass well-known models such as the generalized autoregressive conditional heteroskedasticity, the autoregressive conditional duration and the time-varying Poisson count models. In addition, it allows for a wide range of new observation driven models. We illustrate our approach for two classes of models: multivariate point process models with time-varying parameters and models with time-varying copula functions. We develop the dynamic model specifications for these models in detail and provide simulation and empirical evidence.

Keywords: dynamic models, time-varying parameters, non-linearity, exponential family, marked point processes, copulas.

JEL classification codes: C10, C22, C32, C51.

*We would like to thank Karim Abadir, Konrad Banachewicz, Charles Bos, Jianqing Fan, Clive Granger †, Andrew Harvey, Marius Ooms, Neil Shephard and Michel van der Wel for their comments on an earlier draft of the paper. We have benefited from the comments of participants at the conference “High-Frequency Data Analysis in Financial Markets” of Hitotsubashi University, Tokyo, at a meeting of Society for Financial Econometrics (SoFiE) in Geneva, and at seminar presentations at the University of Alicante, University of Chicago, NAKÉ 2008 Utrecht research day, Center for Operations Research and Econometrics at Université Catholique de Louvain, Imperial College London, Oxford-Man Institute, Princeton University, Erasmus University Rotterdam, Tinbergen Institute and VU University Amsterdam. We thank Moody’s for providing the credit rating transition data for one of our applications. Correspondence: Drew D. Creal, University of Chicago Booth School of Business, 5807 South Woodlawn Avenue, Chicago, IL 60637. Tel: 773.834.5249; Email: dcreal@chicagobooth.edu

1 Introduction

In many settings of empirical interest, time variation in a selection of model parameters is important for capturing the dynamic behavior of univariate and multivariate time series processes. Time series models with time-varying parameters have been categorized by Cox (1981) into two classes of models: observation driven models and parameter driven models. In the observation driven approach, time variation of the parameters is introduced by letting parameters be functions of lagged dependent variables as well as contemporaneous and lagged exogenous variables. Although the parameters are stochastic, they are perfectly predictable given the past information. This approach simplifies likelihood evaluation and explains why observation driven models have become popular in the applied statistics and econometrics literature. Typical examples of these models are the generalized autoregressive conditional heteroskedasticity (GARCH) model of Engle (1982), Bollerslev (1986) and Engle and Bollerslev (1986), the autoregressive conditional duration and intensity (ACD and ACI, respectively) models of Engle and Russell (1998) and Russell (2001), the dynamic conditional correlation (DCC) model of Engle (2002a), the Poisson count models discussed by Davis, Dunsmuir, and Streett (2003), the dynamic copula models of Patton (2006), and the time-varying quantile model of Engle and Manganelli (2004). In our modeling framework for time-varying parameters, many of the existing observation driven models are encompassed as mentioned above. In addition, new models can be formulated and investigated.

In parameter driven models, the parameters are stochastic processes with their own source of error. Given past and concurrent observations, the parameters are not perfectly predictable. Typical examples are the stochastic volatility (SV) model, see Shephard (2005) for a detailed discussion, and the stochastic intensity models of Bauwens and Hautsch (2006) and Koopman, Lucas, and Monteiro (2008). Estimation is usually more involved for these models because the associated likelihood functions are not available in closed-form. Exceptions include linear Gaussian state space models and discrete-state hidden Markov models, see Harvey (1989) and Hamilton (1989), respectively. In most other cases, computing the likelihood function requires the evaluation of a high-dimensional integral based on simulation methods such as importance sampling and Markov chain Monte Carlo; for example, see Shephard and Pitt (1997).

The main contribution of this paper is the development of a framework for time-varying

parameters within the class of observation driven models. The primary difficulty in formulating a unified framework lies in the choice of a function that links the past observations to future parameter values. Such a function should be applicable to a wide class of non-linear and non-Gaussian models. In this paper, we argue that the scaled score function of the model density at time t is an effective choice for the driving mechanism of the time-varying parameters. By choosing the scaling appropriately, standard observation driven models such as the GARCH, ACD, and ACI models can be recovered. The scaled score, however, is equally applicable to non-standard multivariate models that have not been explored before and leads to the formulation of new observation driven models.

We refer to our observation driven model based on the scaled score function as the generalized autoregressive score (GAS) model. The GAS model has similar advantages as the GARCH model. Likelihood evaluation is straightforward. Extensions to asymmetric, long memory, and other more complicated dynamics can be considered without introducing further complexities. As it is based on the score, the GAS model exploits the complete density structure rather than only means and higher moments. This further differentiates the GAS model from alternative observation driven models in the literature, such as (within the exponential family of distributions) the generalized linear autoregressive (GLAR) models of Shephard (1995), the generalized autoregressive moving average (GARMA) models of Benjamin, Rigby, and Stanispoulos (2003), and the vector multiplicative error models (MEM) of Cipollini, Engle, and Gallo (2006).

Two empirical studies are presented to illustrate how the GAS modeling framework can be implemented as part of a statistical or econometric analysis. In financial econometric work, credit risk models are important as financial regulators are pressured to control credit risk and defaults. Such models are based on marked point-processes for different levels of risk and with time-varying intensities. Parameter estimation typically relies on computationally demanding methods, see for example, Duffie, Eckner, Horel, and Saita (2006). One of the main intricacies in this modeling area is that credit events are sparse for each individual company. Consequently, pooling restrictions have to be imposed to extract the information about time-variation in the parameters. This stands in sharp contrast to multivariate point-process models for, e.g., trade intensities. There, we typically observe many events (trades or quotes) for the same stock. We show how a multi-state model for pooled marked point-processes subject to pooling restrictions follows naturally within the GAS framework. The resulting observation driven approach for

time-varying intensities is practical and its implementation is straightforward. We analyze an extensive data set of Moody's ratings history of more than 8,000 U.S. corporates over a time span of almost thirty years.

In our second illustration we investigate whether returns from different financial markets exhibit dependence. Dependence in a multivariate setting can be conveniently captured using copula methods. These have been popularized in financial risk management by for example McNeil, Frey, and Embrechts (2005). Several authors have shown that the dependence structure may not be constant over time. The introduction of time-variation in a copula is considered by Patton (2006). We show that our GAS framework can be more successful in capturing time-varying dependencies as it accounts for more characteristics of the copula function via the score function. Evidence is given for simulated data and for empirical data. In the latter case, daily returns from different currency exchange rates are analyzed.

The remainder of the paper is organized as follows. In Section 2 we provide the basic GAS specification together with a set of motivating examples. Section 3 includes a discussion of the statistical properties of GAS models. Section 4 presents the two examples: a model for marked point-processes with time-varying parameters and a model with time-varying copula functions. Section 5 concludes.

2 Model specification

In this section we formulate a general class of observation driven time-varying parameter models. The basic specification is introduced and a set of examples is provided for illustrative purposes. We also discuss some further possible variations in the model specification.

2.1 Basic model specification

Let $N \times 1$ vector y_t denote a dependent variable of interest, f_t a time-varying parameter vector, x_t a vector of exogenous variables (covariates), all at time t , and θ a vector of static parameters. Define $Y_1^t = \{y_1, \dots, y_t\}$, $F_1^t = \{f_1, \dots, f_t\}$, and $X_1^t = \{x_1, \dots, x_t\}$. The available information set at time t consists of $\{f_t, \mathcal{F}_t\}$ where

$$\mathcal{F}_t = \{Y_1^{t-1}, F_1^{t-1}, X_1^t\}, \quad \text{for } t = 1, \dots, n.$$

We assume that y_t is generated by the observation density

$$p(y_t | f_t, \mathcal{F}_t; \theta). \quad (1)$$

Furthermore, we assume that the mechanism for updating the time-varying parameter f_t is given by the familiar autoregressive updating equation

$$f_{t+1} = \omega + \sum_{i=1}^p A_i s_{t-i+1} + \sum_{j=1}^q B_j f_{t-j+1}, \quad (2)$$

where ω is a vector of constants, coefficient matrices A_i and B_j have appropriate dimensions for $i = 1, \dots, p$ and $j = 1, \dots, q$, while s_t is an appropriate function of past data, $s_t = s_t(y_t, f_t, \mathcal{F}_t; \theta)$. The unknown coefficients in (2) are functions of θ , that is $\omega = \omega(\theta)$, $A_i = A_i(\theta)$, and $B_j = B_j(\theta)$ for $i = 1, \dots, p$ and $j = 1, \dots, q$. The main contribution of this paper is the particular choice for the driving mechanism s_t that is applicable uniformly over a wide class of observation densities and non-linear models.

Our approach is based on the observation density (1) for a given parameter f_t . When an observation y_t is realized, we update the time-varying parameter f_t to the next period $t + 1$ using (2) with

$$s_t = S_t \cdot \nabla_t, \quad \nabla_t = \partial \ln p(y_t | f_t, \mathcal{F}_t; \theta) / \partial f_t, \quad S_t = S(t, f_t, \mathcal{F}_t; \theta), \quad (3)$$

where $S(\cdot)$ is a matrix function. Given the dependence of the driving mechanism in (2) on the scaled score vector (3), we let the equations (1) – (3) define the generalized autoregressive score model with orders p and q . We abbreviate the resulting model by GAS(p, q).

The use of the score for updating f_t is very intuitive. It defines a steepest ascent direction for improving the model's local fit at time t given the current position of the parameter f_t . This provides the natural direction for updating the parameter. In addition, the score depends on the complete density, and not only on the first or second order moments of the observations y_t . This distinguishes the GAS framework from most other observation driven approaches in the literature. By exploiting the full density structure, the GAS model introduces new transformations of the data that can be used to update the time-varying parameter f_t .

Via its choice of the scaling matrix S_t , the GAS model allows for additional flexibility in how the score is used for updating f_t . It is important to note that each different choice for the scaling matrix S_t results in a different GAS model. The statistical and empirical properties of each of these models can be different and warrant separate inspection.

In many situations, it is natural to consider a form of scaling that depends on the variance of the score, i.e., on

$$\mathcal{I}_{t|t-1} = \mathbb{E}_{t-1} [\nabla_t \nabla_t'], \quad (4)$$

where \mathbb{E}_{t-1} is expectation with respect to density $p(y_t|f_t, \mathcal{F}_t; \theta)$. Our preferred choice in the current paper is the GAS model with

$$S_t = \mathcal{J}_{t|t-1}, \quad \mathcal{J}'_{t|t-1} \mathcal{J}_{t|t-1} = \mathcal{I}_{t|t-1}^{-1}, \quad (5)$$

such that S_t is the square root matrix of the (pseudo)-inverse information matrix for (1) with respect to f_t . The main advantage of this specific choice for S_t is that it renders the statistical properties of the corresponding GAS model more tractable. This follows from the fact that for $S_t = \mathcal{J}_{t|t-1}$ the GAS step s_t has constant unit variance. We discuss this in more detail in Section 3.

Other convenient choices are $S_t = \mathcal{I}_{t|t-1}^{-1}$ and $S_t = \mathbf{I}$. For $S_t = \mathcal{I}_{t|t-1}^{-1}$, the GAS model encompasses the well-known observation driven GARCH, ACD, and ACI models as well as most of the Poisson count models considered by Davis et al. (2003). For $S_t = \mathbf{I}$, the GAS model captures models such as the autoregressive conditional multinomial (ACM) model of Russell and Engle (2005) or the GARMA models of Benjamin, Rigby, and Stanispoulos (2003). In the context of a fully generic observation density $p(y_t|f_t, \mathcal{F}_t; \theta)$, however, the statistical properties of the GAS model for these alternative choices of S_t are typically much more complicated.

We can further generalize the GAS updating equation (2) in various directions. For example, it may be interesting to include exogenous variables in (2), or to generalize the evolution of f_t by including other non-linear effects such as regime-switching. In addition, it may be more appropriate in some applications to consider long-memory versions of (2), for example

$$f_{t+1} = \omega + \sum_{i=1}^{\infty} \frac{(i+d-1)!}{i!(d-1)!} s_{t-i+1},$$

for a scalar f_t and a fractional integration parameter $d < 1/2$. We obtain a fractionally integrated GAS (FIGAS) model specification in the same vein as the well-known ARFIMA and FIGARCH models, see the contributions of Hosking (1981) and Baillie, Bollerslev, and Mikkelsen (1996), respectively.

2.2 Examples of GAS models

In this section we provide a number of simple examples that show how to operationalize the GAS framework. The examples also reveal that the GAS framework encompasses a large number of available observation driven models presented in the literature for an appropriate choice of the scaling matrix S_t .

Example 1 (old and new GARCH models): Consider the basic model $y_t = \sigma_t \varepsilon_t$ where the Gaussian disturbance ε_t has zero mean and unit variance while σ_t is a time-varying standard deviation. It is a basic exercise to show that the GAS(1,1) model with $S_t = \mathcal{I}_{t|t-1}^{-1}$ and for $f_t = \sigma_t^2$ reduces to

$$f_{t+1} = \omega + A_1 (y_t^2 - f_t) + B_1 f_t, \quad (6)$$

which is equivalent to the standard GARCH(1,1) model as given by

$$f_{t+1} = \alpha_0 + \alpha_1 y_t^2 + \beta_1 f_t, \quad f_t = \sigma_t^2, \quad (7)$$

where coefficients $\alpha_0 = \omega$, $\alpha_1 = A_1$ and $\beta_1 = B_1 - A_1$ are unknown and require certain conditions for stationarity, see Bollerslev (1986). However, if we assume that ε_t follows a Student's t distribution with ν degrees of freedom and unit variance, the GAS(1,1) specification for the conditional variance leads to the updating equation

$$f_{t+1} = \omega + A_1 \cdot (1 + 3\nu^{-1}) \cdot \left(\frac{(1 + \nu^{-1})}{(1 - 2\nu^{-1})(1 + \nu^{-1}y_t^2 / (1 - 2\nu^{-1}) f_t)} y_t^2 - f_t \right) + B_1 f_t, \quad (8)$$

see also Creal, Koopman, and Lucas (2010) for the full multivariate case and further generalizations. In case $\nu^{-1} = 0$, the Student's t distribution reduces to the Gaussian distribution and the factor recursion (8) collapses to (6) as required. The recursion in (8), however, has an important difference with the standard t-GARCH(1,1) model of Bollerslev (1987) which has the Student's t density in (1) with the updating equation (6). The denominator of the second

term in the right-hand side of (8) causes a more moderate increase in the variance for a large realization of $|y_t|$ as long as ν is finite. The intuition is clear: if the errors are modeled by a fat-tailed distribution, a large absolute realization of y_t does not necessitate a substantial increase in the variance. The GAS updating mechanism for the model with Student's t errors therefore is substantially different from its familiar GARCH counterpart. In independent work, a similar variance updating equation as (8) for the univariate Student's t distribution is proposed by Harvey and Chakravarty (2008); they also discuss the properties of the model in more detail.

The GAS framework also provides a range of alternative time-varying variance equations for other heavy-tailed distributions. For example, consider the asymmetric Laplace distribution obtained by $y_t = w_t \cdot \tilde{y}_t^L + (1 - w_t) \cdot \tilde{y}_t^R$, where w_t is a Bernoulli random variable with $\Pr[w_t = 0] = (1 + \vartheta^2)^{-1}$ for coefficient $\vartheta > 0$ and where $-\tilde{y}_t^L$ and \tilde{y}_t^R are exponentially distributed random variables with means $\vartheta\sigma/2^{1/2}$ and $\sigma/(2^{1/2}\vartheta)$, respectively. The random variables w_t , \tilde{y}_t^L and \tilde{y}_t^R are assumed to be independent. The mean and variance of y_t are 0 and σ^2 , respectively. If we let $f_t = \log(\sigma_t^2)$, the GAS step takes the form

$$s_t = 2 \left(\frac{2^{1/2}(-y_t)}{\vartheta\sigma} - 1 \right) \cdot 1_{\{y_t|y_t \leq 0\}}(y_t) + 2 \left(\frac{2^{1/2}\vartheta y_t}{\sigma} - 1 \right) \cdot 1_{\{y_t|y_t > 0\}}(y_t), \quad (9)$$

where $1_A(x)$ is the indicator function for the set A , that is $1_A(x) = 1$ if $x \in A$, and zero otherwise. The GAS driving mechanism (9) is composed of linear segments with unequal absolute slopes. We can rewrite this as

$$s_t = \tilde{\vartheta}_1 \frac{2^{1/2}y_t}{\sigma} + \tilde{\vartheta}_2 \left(\frac{2^{1/2}|y_t|}{\sigma} - 2\tilde{\vartheta}_2^{-1} \right), \quad (10)$$

where $\tilde{\vartheta}_1 = (\vartheta^2 - 1)/\vartheta$ and $\tilde{\vartheta}_2 = (\vartheta^2 + 1)/\vartheta$. The specification (10) is equivalent to the driving mechanism as adopted for the so-called EGARCH model of Nelson (1991).

Example 2 (MEM, ACD and ACI models): Consider the model $y_t = \mu_t \varepsilon_t$ where ε_t has a gamma distribution with density $p(\varepsilon_t; \alpha) = \Gamma(\alpha)^{-1} \varepsilon_t^{\alpha-1} \alpha^\alpha \exp(-\alpha \varepsilon_t)$, coefficient α and mean μ_t as the mean of ε_t . Using a change of variable, we obtain the model density

$$p(y_t|\mu_t; \alpha) = \Gamma(\alpha)^{-1} y_t^{\alpha-1} \alpha^\alpha \mu_t^{-\alpha} \exp\left(-\alpha \frac{y_t}{\mu_t}\right). \quad (11)$$

In case we set $f_t = \mu_t$, the GAS(1, 1) updating equation with $S_t = \mathcal{I}_{t|t-1}^{-1}$ becomes

$$f_{t+1} = \omega + A_1 (y_t - f_t) + B_1 f_t. \quad (12)$$

This specification is equivalent to the multiplicative error model (MEM) proposed by Engle (2002b) and extended in Engle and Gallo (2006). The exponential distribution is a special case of the gamma distribution when $\alpha = 1$. This makes the ACD and ACI models a special case of MEM. The ACD model of Engle and Russell (1998) follows from (11) straightforwardly with $\alpha = 1$ and factor recursion (12). In case we specify the exponential density in terms of the intensity rather than the expected duration, we obtain the model density $p(y_t | \lambda_t) = \lambda_t \exp(-\lambda_t y_t)$ with intensity $\lambda_t = 1/\mu_t$. Let $\tilde{f}_t = \log(\lambda_t)$ and it follows that the GAS(1, 1) updating equation becomes

$$\tilde{f}_{t+1} = \omega + A_1 \left[1 - y_t \exp(\tilde{f}_t) \right] + B_1 \tilde{f}_t, \quad (13)$$

which is equivalent to the standard ACI(1, 1) model of Russell (2001).

Example 3 (Dynamic exponential family models): The class of natural exponential family models for a vector of observations y_t can be represented by the density function

$$p(y_t | f_t, \mathcal{F}_t; \theta) = \exp [\gamma' y_t - c(\gamma) + h(y_t)], \quad (14)$$

with scalar functions $c(\cdot)$ and $h(\cdot)$ and $m \times 1$ parameter vector γ . We consider replacing γ by a time-varying parameter vector γ_t that is specified as

$$\gamma_t = d + Z f_t,$$

with $m \times 1$ constant vector d and $m \times r$ factor loading matrix Z . The unknown coefficients in d and Z are placed in parameter vector θ . Further, we impose a GAS specification on the time-varying factor f_t . The GAS driving mechanism with $S_t = \mathcal{J}_{t|t-1}$ is given by

$$s_t = [Z' \ddot{c}(\gamma_t) Z]^{-1/2} Z' [y_t - \dot{c}(\gamma_t)],$$

where $\dot{c}(\gamma_t) = \partial c(\gamma_t) / \partial \gamma_t$ and $\ddot{c}(\gamma_t) = \partial^2 c(\gamma_t) / \partial \gamma_t \partial \gamma_t'$. This model directly encompasses some well-known models from the literature if we change the scaling choice. For example, for a Poisson density in (14) and $S_t = \mathcal{I}_{t|t-1}^{-1}$ we recover the observation-driven model for Poisson counts of Davis et al. (2003).

3 Statistical properties

3.1 Estimation, inference and stationarity

A convenient property of observation driven models is the relatively simple way of estimating parameters by maximum likelihood (ML). This feature applies to the GAS model as well. For an observed time series y_1, \dots, y_n and by adopting the standard prediction error decomposition, we can express the maximization problem as

$$\hat{\theta} = \arg \max_{\theta} \sum_{t=1}^n \ell_t, \quad (15)$$

where $\ell_t = \ell(\theta; y_t, f_t, \mathcal{F}_t) = \ln p(y_t | f_t, \mathcal{F}_t; \theta)$ for a realization of y_t . Evaluating the likelihood function of the GAS model is particularly simple. It only requires the implementation of the GAS updating equation (2) and the evaluation of ℓ_t for a particular value θ^* of θ .

It is possible to formulate recursions for computing the gradient of the likelihood with respect to the static parameter vector θ . Gradient recursions for the GARCH model have been developed by Fiorentini, Calzolari, and Panattoni (1996). In case of the GAS(1, 1) specification, the gradient is computed via the chain rule, that is

$$\frac{\partial \ell_t}{\partial \theta'} = \frac{\partial \ln p_t}{\partial \theta'} + \frac{\partial \ln p_t}{\partial f_t'} \cdot \frac{\partial f_t}{\partial \theta'}, \quad (16)$$

with $p_t = p(y_t | f_t, \mathcal{F}_t; \theta)$ and

$$\frac{\partial f_t}{\partial \theta'} = \frac{\partial \omega}{\partial \theta'} + A_1 \frac{\partial s_{t-1}}{\partial \theta'} + B_1 \frac{\partial f_{t-1}}{\partial \theta'} + (s'_{t-1} \otimes \mathbf{I}) \frac{\partial \vec{A}_1}{\partial \theta'} + (f'_{t-1} \otimes \mathbf{I}) \frac{\partial \vec{B}_1}{\partial \theta'}, \quad (17)$$

$$\frac{\partial s_{t-1}}{\partial \theta'} = S_{t-1} \frac{\partial \nabla_{t-1}}{\partial \theta'} + (\nabla'_{t-1} \otimes \mathbf{I}) \frac{\partial \vec{S}_{t-1}}{\partial \theta'}, \quad (18)$$

where $\vec{A} = \text{vec}(A)$ denotes the vector with the stacked columns of the matrix A , and \otimes is the Kronecker matrix product. The derivations for $\partial \nabla_{t-1} / \partial \theta'$ and $\partial \vec{S}_{t-1} / \partial \theta'$ should also consider the effect of θ through f_t as in (16). Higher order GAS specifications can be dealt with similarly by formulating the GAS model updating equation in a companion form. The log-likelihood derivatives can be computed simultaneously with the time-varying parameters f_t . However, computing the analytic derivatives, particularly for (18), may be cumbersome. In practice, we therefore often turn to likelihood maximization based on numerical derivatives.

The easiest way to conduct inference for GAS models is to apply a standard limiting result and use the inverse information matrix of the likelihood at the optimum to compute standard errors and t -values for the estimated parameters. In particular, if θ gathers all static parameters of the model, we conjecture that under suitable regularity conditions, the maximum likelihood estimator $\hat{\theta}$ of θ is consistent and satisfies

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} \text{N}(0, H^{-1}),$$

where $H = \text{E}[(\partial \ell / \partial \theta)(\partial \ell / \partial \theta)']$ and $\ell = \sum_{t=1}^n \ell_t$.

A key condition for applying a central limit theorem is a stationary and ergodic process for f_t . Though different alternatives for S_t are possible, it is here that our choice of $S_t = \mathcal{J}_{t|t-1}$ proves convenient. We write the GAS(1,1) updating equation in its infinite order moving average form as

$$f_t = (\text{I} - B_1)^{-1} \omega + A_1 \sum_{i=0}^{\infty} B_1^i s_{t-i}.$$

For $S_t = \mathcal{J}_{t|t-1}$, it follows immediately that $\text{E}_{t-1}[s_t] = 0$, and $\text{E}_{t-1}[s_t s_t'] = \text{I}$. The series s_t thus forms a martingale difference series with unit variance. Therefore, it is sufficient for covariance stationarity of f_t that the roots of B_1 are inside the unit circle. For other scaling choices, such conditions are less evident in general.

A second advantage of setting $S_t = \mathcal{J}_{t|t-1}$ follows from Nelson and Foster (1994). They derive optimality properties for the choice of $S_t = \mathcal{J}_{t|t-1}$ from an optimal filtering perspective. They thus show that $S_t = \mathcal{J}_{t|t-1}$ can be statistically optimal in at least one well-defined setting. Our perspective in the current paper, however, is entirely different. We focus on GAS models as a new class of observation driven models for complex statistical problems with dynamic

coefficients. Moreover, the results in Nelson and Foster (1994) cannot immediately be applied to all models such as those presented in Section 4. The results may, however, open up a different route to study the statistical behavior of GAS models through their continuous record asymptotics.

3.2 Specification and identification

An important feature of the GAS(p, q) specification is that its applications are not restricted to one specific model or choice of model parameterization. More specifically, the GAS framework is applicable to a large set of models that is characterized by a parametric likelihood function. This is particularly relevant for the new applications in Section 4, where we generalize models with static parameters outside their usual area of application. For example, if the time-varying parameter is common across different observations, equations (1)–(3) provides a natural and automatic way to weight observed information in a manner that is consistent with the model of interest.

Congruent with the observation density (1), the GAS specification allows for different parameterizations. In the GARCH example of Section 2.2, for example, the time-varying parameter is $f_t = \sigma_t^2$. When it is preferred to enforce the positivity of σ_t^2 , an obvious alternative is to parameterize the model in terms of $\tilde{f}_t = \log(\sigma_t^2)$. The GAS dynamics automatically adapt to the choice of parameterization. In general, assume that one prefers a different parameterization $\tilde{f}_t = h(f_t)$ for some continuous and invertible mapping $h(\cdot)$. Let $\dot{h}_t = \partial h(f_t)/\partial f_t'$ which is deterministic given all information up to and including time t , that is, given \mathcal{F}_t . For well behaved densities, the information matrix equals both the expected outer product of scores and the expected second derivative of the log density. Therefore,

$$\tilde{\mathcal{J}}'_{t|t-1} \tilde{\mathcal{J}}_{t|t-1} = \left(\mathbb{E}_{t-1}[(\dot{h}_t^{-1})' \nabla_t \nabla_t' \dot{h}_t^{-1}] \right)^{-1} = \dot{h}_t \mathcal{I}_{t|t-1}^{-1} \dot{h}_t' = \dot{h}_t \mathcal{J}'_{t|t-1} \mathcal{J}_{t|t-1} \dot{h}_t', \quad (19)$$

where tildes denote that derivatives are taken with respect to \tilde{f}_t rather than f_t . Similarly, we have

$$\tilde{\nabla}_t = \partial \ln p(y_t; f_t, \mathcal{F}_t; \theta) / \partial \tilde{f}_t = (\dot{h}_t')^{-1} \nabla_t. \quad (20)$$

The GAS updating step for \tilde{f}_t with square root information scaling is then given by

$$\tilde{s}_t = \tilde{\mathcal{J}}_{t|t-1} \tilde{\nabla}_t = \tilde{\mathcal{J}}_{t|t-1} (\dot{h}'_t)^{-1} \mathcal{J}_{t|t-1}^{-1} s_t, \quad (21)$$

since $s_t = \mathcal{J}_{t|t-1} \nabla_t$. For the univariate case, it is easy to see that $\tilde{\mathcal{J}}_{t|t-1} (\dot{h}'_t)^{-1} \mathcal{J}_{t|t-1}^{-1} = 1$. For the multivariate case with a vector f_t , it follows that the updating step under the reparameterization is an orthogonal linear transformation of the original step since

$$\left(\tilde{\mathcal{J}}_{t|t-1} (\dot{h}'_t)^{-1} \mathcal{J}_{t|t-1}^{-1} \right) \left(\tilde{\mathcal{J}}_{t|t-1} (\dot{h}'_t)^{-1} \mathcal{J}_{t|t-1}^{-1} \right)' = \tilde{\mathcal{J}}_{t|t-1} (\dot{h}'_t)^{-1} \mathcal{I}_{t|t-1} (\dot{h}_t)^{-1} \tilde{\mathcal{J}}'_{t|t-1} = \mathbf{I}, \quad (22)$$

where the last equality follows from (19). The choice of parameterization thus only has a minor effect on the form of the updating step s_t if we adopt $\mathcal{J}_{t|t-1}$ as our scaling matrix. In particular, the new \tilde{s}_t is also a unit variance martingale difference series. Other forms of scaling have different implications. For example, if we scale the score by the inverse information matrix $\mathcal{I}_{t|t-1}^{-1}$, the updating step \tilde{s}_t for \tilde{f}_t instead equals $\dot{h}_t s_t$, or \dot{h}_t times the updating step for f_t .

Another important issue concerns parameter identification, particularly in case the time-varying parameter f_t obeys a factor structure. Consider a model density of the form

$$p(y_t; Z f_t) = p(y_t; Z f_t, \mathcal{F}_t; \theta), \quad (23)$$

where the time-varying parameter f_t follows a GAS(1, 1) specification and Z is a loading matrix; also see the discussion below (14) in Example 3. For example, $Z f_t$ can be a vector of volatilities for a vector time series driven by a single common parameter f_t . In this case, it is not possible to identify both Z and all GAS(1, 1) parameters ω , A_1 and B_1 , simultaneously. To illustrate the identification concern, consider model (23) and introduce an invertible matrix K . Define $\bar{f}_t = K f_t$, $\bar{s}_t = K s_t$, $\bar{Z} = Z K^{-1}$, $\bar{\omega} = K \omega$, $\bar{A}_1 = K A_1$ and $\bar{B}_1 = K B_1 K^{-1}$. The likelihood values associated with $p(y_t; Z f_t)$ and $p(y_t; \bar{Z} \bar{f}_t)$ are obviously identical. Pre-multiplying the GAS(1, 1) updating equation for f_t by K and using our result (21), we obtain

$$K f_{t+1} = K \omega + K A_1 s_t + K B_1 K^{-1} K f_t \quad \Leftrightarrow \quad \bar{f}_{t+1} = \bar{\omega} + \bar{A}_1 \bar{s}_t + \bar{B}_1 \bar{f}_t.$$

Since K is an arbitrary invertible matrix, restrictions must be imposed on matrix Z to ensure

identification. For example, specific rows of Z can be set equal to rows of the identity matrix.

4 Applications and new models

In this section we present two examples to illustrate our GAS modeling framework. In the two illustrations the GAS updating equation is based on square root information matrix scaling, that is $S_t = \mathcal{J}_{t|t-1}$ as defined in (5).

4.1 Pooled marked point-process models

Point process models

Models with time-varying intensities have received attention in finance and econometrics. The principal areas of application in economics include intraday trade data (market microstructure), defaults of firms, credit rating transitions and (un)employment spells over time. To illustrate the GAS model in this setting, we consider an application from the credit risk literature in which pooled marked point-processes play an important role.

Promising models with stochastically evolving intensities have been proposed by Bauwens and Hautsch (2006), Koopman et al. (2008), Duffie, Eckner, Horel, and Saita (2006), and Koopman, Lucas, and Schwaab (2008). The econometric treatment of parameter driven models is intricate while parameter estimation can be computationally demanding regardless whether classical or Bayesian methods are used. In particular, likelihood evaluation for these models requires the computation of high-dimensional integrals using importance sampling techniques or Markov chain Monte Carlo algorithms. Here we propose an alternative, observation driven model for time-varying intensities. The model follows naturally by adopting the GAS framework.

The GAS marked point process model

Let $y_{k,t} = (y_{1k,t}, \dots, y_{Jk,t})'$ be a vector of marks of J competing risk processes for firms $k = 1, \dots, N$. We have $y_{jk,t} = 1$ if event type j out of J materializes for firm k at time t , and zero otherwise, and we assume that the pooled point process is orderly, such that with probability 1 precisely one event occurs at each event time. Let t^* denote the last event time before time t

and let $\lambda_{k,t} = (\lambda_{1k,t}, \dots, \lambda_{Jk,t})'$ be a $J \times 1$ vector of log-intensities. We model the log intensities by

$$\lambda_{k,t} = d + Z f_t, \quad (24)$$

where d is a $J \times 1$ vector of baseline intensities and Z is a $J \times r$ matrix of factor loadings. The $r \times 1$ vector of dynamic factors f_t is specified by the GAS(1,1) updating equation (2) with $\omega = 0$. Since f_t is not observed directly, we need to impose a sign restriction on Z to obtain economic interpretations for the time-varying parameters. We assume the model has a factor structure: the intensities of all firms are driven by the same vector of time-varying systematic parameters f_t . We thus require parameter restrictions for model identification such as those discussed in Section 3.2.

Model (24) nests the model of Russell (2001) when we set the dimension of f_t equal to the number of firms, $r = N$. In a credit risk context, this is not realistic and we typically have $r \ll N$. This is due to the fact that credit rating events are much sparser than trades in a liquid stock. The former are one up to 12 over a span of thirty years, while the latter may be already substantial if only counted within a five-minute time span. This difference in the structure of the data makes the empirical modeling process for credits substantially different from that for trade intensities.

The log-likelihood specification using (24) is given by

$$\ell_t = \sum_{j=1}^J \sum_{k=1}^N y_{jk,t} \lambda_{jk,t} - R_{jk,t} \cdot (t - t^*) \cdot \exp(\lambda_{jk,t^*}), \quad (25)$$

where $R_{jk,t}$ is a zero-one variable indicating whether company k is potentially subject to risk j at time t . Define P as a $J \times J$ diagonal matrix with j th diagonal element $p_{j,t} = \sum_k R_{jk,t} \cdot \exp[\lambda_{jk,t}] / \sum_{j,k} R_{jk,t} \cdot \exp[\lambda_{jk,t}] = \text{P}[\sum_k y_{jk,t} = 1]$, i.e., the probability that the next event is of type j . Also define the $J \times 1$ vector ∇_t with j th element $\nabla_{j,t} = \sum_k y_{jk,t} - R_{jk,t} \cdot (t - t^*) \cdot \exp(\lambda_{jk,t^*})$. Based on the first and second derivative of ℓ_t and setting $S_t = \mathcal{J}_{t|t-1}$, we obtain the scaled score function

$$s_t = (Z' P Z)^{-\frac{1}{2}} Z' \nabla_t. \quad (26)$$

By combining these basic elements into a GAS specification, we have obtained a new observation driven model for credit rating transitions. Compared to its parameter driven counterparts, the

current model is much easier to estimate.

Application to Moody's credit rating data

For our illustration, we adopt the marked point-process model (24), (25) and (2) with $\omega = 0$ and s_t given by (26) for a data set which contains Moody's rating histories of all US corporates over the period January 1980 to June 2008. Firms' initial credit ratings are noted at the beginning of the sample and we observe the transitions they make over time from one rating category to another. Conditional upon observing a transition, the number of transitions of each type are pooled into two complementary credit rating classes: the investment grade (IG) and the sub-investment grade (SIG). There are $J = 4$ possible events (one upgrade \uparrow and three downgrades \downarrow) which are the rating transitions

$$j = 1 : \text{IG to SIG } \downarrow; \quad j = 2 : \text{IG to default } \downarrow; \quad j = 3 : \text{SIG to IG } \uparrow; \quad j = 4 : \text{SIG to default } \downarrow .$$

First, we consider a GAS(1,1) model with a common time-varying parameter f_t , that is (24) with $r = 1$ and identifying restriction $Z_{4,1} = 1$ where $Z_{i,j}$ is the (i,j) element of Z . The estimates for d , Z , A_1 and B_1 are presented in Table 1. All estimated baseline coefficients in d are significant. The baseline downgrade from IG to default is the smallest with an estimate of -9.961 . Its associated loading with the common factor $Z_{2,1}$ is the largest with an estimate of 2.143 but the estimate is not highly significant which is partly due to the relatively low number of such transitions in our dataset. From the estimates for $Z_{1,1}$ and $Z_{2,1}$ (and with $Z_{4,1} = 1$), it appears that downgrades are most sensitive to the common factor f_t . The GAS parameter B_1 is estimated close to unity which implies a persistent dynamic process for f_t . Interestingly, the estimated pattern (not shown) of the systematic intensity factor f_t is close to the estimated pattern of the parameter driven model of Koopman et al. (2008). However, in our current GAS framework we do not require their computationally intensive methods such as importance sampling methods for parameter and factor estimation.

Next, following Koopman et al. (2008) we consider a more general model with three common factors ($r = 3$). For this model, equation (24) applies with $r = 3$ and the matrices A_1 and B_1 of the GAS(1,1) updating equation are $r \times r$ diagonal matrices. For identification purposes,

we take the loading matrix Z as

$$r = 1 : \begin{bmatrix} * \\ * \\ * \\ 1 \end{bmatrix}, \quad r = 3 : \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & * \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

where $*$ indicates a coefficient that is estimated. In case $r = 3$, we have a factor for upgrades from SIG to IG, a factor for downgrades from IG to SIG and a common factor for defaults. The estimate 2.133 for $Z_{2,3}$ can be compared with the estimate for $Z_{2,1}$ of the 1 factor model; they are statistically equal to each other and they are not strongly significant. Although we do not provide a theory for comparing pooled marked point-process factor models with different numbers of factors, the large difference between the log-likelihood values for $r = 1$ and $r = 3$ (with additional parameter and factor) is indicative that more than one factor is needed to fit the data. For the model with $r = 3$, Figure 1 presents the estimated intensities of each transition type. We obtain similar estimated patterns (not shown) for the three factors in f_t compared to the more involved parameter driven model of Koopman et al. (2008) with three factors. In particular, we corroborate the finding that the dynamics of upgrades are substantially different from those of downgrades and defaults. We do so without leaving the convenience of the observation driven modeling framework.

4.2 The dynamic Gaussian copula model

Copulas have recently become popular in financial risk management. The copula is a multivariate distribution function over a hypercube with uniform marginals. It can be used to link marginal distributions into a multivariate distribution using Sklar's theorem. In this section, we demonstrate that the GAS framework can provide a new model specification for the bivariate Gaussian copula.

We consider a simple Gaussian copula where the GAS model suggests an alternative dynamic structure compared to earlier suggestions in the literature. Patton (2006) introduced the notion of time-varying copulas, see also Dias and Embrechts (2004) and van den Goorbergh, Genest, and Werker (2005). The (Gaussian) correlation parameter ρ_t is modeled by the transformed

parameter $\rho_t = [1 - \exp(-f_t)] / [1 + \exp(-f_t)]$. In the case of Patton (2006), the driving mechanism for the dynamic bivariate Gaussian copula is given by

$$f_{t+1} = \omega + A_1 \cdot \sum_{i=1}^m \Phi^{-1}(u_{1,t-i+1})\Phi^{-1}(u_{2,t-i+1}) + B_1 f_t, \quad (27)$$

where $\Phi^{-1}(\cdot)$ is the inverse of the normal distribution function, u_{1t} and u_{2t} are the probability integral transforms using the univariate marginals, and m is a positive integer determining the smoothness of f_t . Equation (27) is intuitively appealing and builds on our understanding of covariances: if the transformed marginals have the same sign, the correlation should increase. The reverse holds if the transformed marginals are of opposite sign.

By using the density of the Gaussian copula, we can derive the GAS specification for the time-varying correlation parameter. The score with respect to the correlation parameter is the same for the Gaussian copula and for the bivariate normal distribution. For $m = 1$, Patton's model (27) reduces to

$$f_{t+1} = \omega + A_1 \cdot y_t + B_1 \cdot f_t, \quad (28)$$

where $y_t = \Phi^{-1}(u_{1t})\Phi^{-1}(u_{2t})$. The GAS(1, 1) updating equation for f_t is obtained as

$$f_{t+1} = \omega + A_1 \frac{\sqrt{1 + \rho_t^2}}{(1 - \rho_t^2)} \left[y_t - \rho_t - \rho_t \frac{(x_t - 2)}{(1 + \rho_t^2)} \right] + B_1 f_t, \quad (29)$$

where $x_t = \Phi^{-1}(u_{1t})^2 + \Phi^{-1}(u_{2t})^2$. The similarities and differences between (28) and (29) are as follows. Both models are driven by y_t so that positively clustered transformed marginals lead to an increase of the correlation parameter. The most interesting difference between the two model specifications is that the GAS model includes the term x_t , where $x_t - 2$ is a martingale difference. To understand the impact of this term, consider two different scenarios where we observe $\Phi^{-1}(u_{1t}) = 1$ and $\Phi^{-1}(u_{2t}) = 1$ or, alternatively, $\Phi^{-1}(u_{1t}) = 1/4$ and $\Phi^{-1}(u_{2t}) = 4$. In both cases, the value of $y_t = 1$ is the same and the recursion in (28) causes f_{t+1} to be the same regardless of which scenario we observe. Conversely, the sum of squares term x_t in the GAS model provides information to distinguish between these two cases. The behavior of f_{t+1} depends on the current value of the correlation ρ_t . If the correlation is positive, the impact of $(x_t - 2)$ is negative. In this case, the $(x_t - 2)$ term offsets part of the effect of $(y_t - \rho_t)$ if the

latter has a positive value. If $(y_t - \rho_t)$ is negative, however, the x_t term reinforces the magnitude of the GAS step. In this way for positive ρ_t the observation $(\Phi^{-1}(u_{1t}), \Phi^{-1}(u_{2t})) = (1, 1)$ drives the correlation ρ_t further up, whereas the observation $(1/4, 4)$ may drive the correlation down if ρ_t is sufficiently large.

For illustrative purposes, we extend the example from Patton (2006) to investigate the dependence of the daily exchange rates of the German Mark (later Euro), against the US dollar, with the Japanese Yen and with the British Pound, also both against the US dollar. The sample period is January 1986 through August 2008. The log returns of the exchange rate series are analyzed by an AR-GARCH model: an autoregressive process for the conditional mean and a GARCH process for the conditional variance. We construct the transformed series for u_{1t} and u_{2t} and use these as input for the Gaussian copula model.

Table 2 reports that the log-likelihood value increases 25 to 125 points when considering GAS instead of Patton for the same number of parameters. The estimates of the parameter B_1 imply that the GAS specification leads to a more persistently time-varying correlation process. However, the increased sensitivity of the score mechanism to correlation shocks in the GAS specification allows f_t to react more fiercely to exchange rate returns of opposite sign if the current correlation estimate is positive. This can be observed clearly for the Mark-Pound example, but also the Mark-Yen example shows similar features at the end of 1993 and 2003. The difference between the dynamics for the different specifications may be highly relevant to risk managers where changes in correlations and in particular correlation breakdowns are of major concern.

4.3 Static mixtures of dynamic copulas

The GAS framework can also be applied to mixtures of possibly non-Gaussian densities. To illustrate this level of flexibility next, we consider a mixture of dynamic copulas. Patton (2006) amends (27) towards a generally applicable driving mechanism for copula parameters. The general updating equation of Patton for a bivariate model is given by

$$f_{t+1} = \omega - m^{-1}A_1 \cdot \sum_{i=1}^m |u_{1,t-i+1} - u_{2,t-i+1}| + B_1 \cdot f_t, \quad (30)$$

where u_{1t} , u_{2t} and m are defined below (27). The time-varying f_t captures the dependence between the coordinates. Assume that $\omega > 0$, $A_1 > 0$ and $1 > B_1 > 0$. When concurrent and recent values of u_{1t} and u_{2t} are close together, f_{t+1} is likely to increase by a value less than ω . The increase represents stronger dependence. Similarly, when concurrent and recent values of u_{1t} and u_{2t} are far apart, it is likely that f_{t+1} decreases.

Although the driving mechanism in (30) is intuitive and simple, two issues are less clear. First, the updating mechanism is not influenced by the particular choice of the copula. As shown for the Gaussian copula, particular features of the copula can be useful for the dynamic specification of f_t . Second, although (30) provides an updating scheme for the bivariate case, the extension to the multivariate case is less obvious. In particular, in case of a copula characterized by a single dependence parameter, different ways exist in which the differences $|u_{it} - u_{jt}|$ for $i \neq j$ can be used to update the dependence parameter. Equation (30) provides little guidance as to how these different and possibly conflicting signals should be weighed.

To illustrate how the GAS framework can cope with these issues we consider the Clayton copula as an alternative to the Gaussian copula discussed in Section 4.2. The details of the Clayton copula with a GAS time-varying dependence are given in the Appendix. The Clayton copula (33) accounts for lower tail dependence but not for upper tail dependence. Therefore, it can be beneficial to use a symmetrized version of the Clayton copula that allows for different upper and lower tail non-zero dependencies. The symmetrized Clayton copula is a mixture of copulas. Therefore, we first derive the general GAS specification for a static mixture of J copulas.

Consider

$$C(u_{1t}, \dots, u_{Nt}; \lambda) = \sum_{j=1}^J w_j^* C_j(u_{1t}, \dots, u_{Nt}; \lambda^{(j)}), \quad (31)$$

where w_j^* is a positive scalar, $C_j(\cdot)$ is the j th copula function (e.g., the Clayton copula as defined in (33) in the Appendix) with λ replaced by $\lambda^{(j)}$ and where u_{it} is defined below (27), for $j = 1, \dots, J$ and $i = 1, \dots, N$. Define $w_{i,t} = w_i^* C_{i,t}^* / \sum_{j=1}^J w_j^* C_{j,t}^*$ as the weight of copula i at time t where $C_{j,t}^*$ is the density function corresponding to copula $C_j(u_{1t}, \dots, u_{N,t}; \lambda^{(j)})$, for $j = 1, \dots, J$. Let f_t represent all time-varying coefficients in $\lambda^{(j)}$ for $j = 1, \dots, J$. The score

function is then given by

$$\frac{\partial \ln C_t^*}{\partial f_t} = \sum_{j=1}^J w_{j,t} \frac{\partial \ln C_{j,t}^*}{\partial f_t},$$

where C_t^* is the density function corresponding to $C(u_{1t}, \dots, u_{N,t}; f_t)$. The Hessian function is

$$\frac{\partial^2 \ln C_t^*}{\partial f_t \partial f_t'} = \sum_{j=1}^J w_{j,t} \left(\frac{\partial^2 \ln C_{j,t}^*}{\partial f_t \partial f_t'} + \frac{\partial \ln C_{j,t}^*}{\partial f_t} \frac{\partial \ln C_{j,t}^*}{\partial f_t'} \right) - \left(\sum_{j=1}^J w_{j,t} \frac{\partial \ln C_{j,t}^*}{\partial f_t} \right) \left(\sum_{j=1}^J w_{j,t} \frac{\partial \ln C_{j,t}^*}{\partial f_t'} \right)'.$$

To obtain the driving mechanism s_t , we notice that

$$\mathcal{I}_{t|t-1} = E_{t-1} \left(\frac{\partial^2 \ln C_t^*}{\partial f_t \partial f_t'} \right) = -E_{t-1} \left[\left(\sum_{j=1}^J w_{j,t} \frac{\partial \ln C_{j,t}^*}{\partial f_t} \right) \left(\sum_{j=1}^J w_{j,t} \frac{\partial \ln C_{j,t}^*}{\partial f_t'} \right)' \right]. \quad (32)$$

It follows that the scores of the individual copulas can be used directly to build a GAS driving mechanism for the mixture copula model.

We illustrate the mixture model for $J = 2$ copulas. The first one is the Clayton copula characterized by the parameter λ_L that accounts for lower tail dependence. The second component of the mixture is known as the survival Clayton copula and is characterized by the parameter λ_U that accounts for upper tail clustering. The GAS mechanism for the mixture of copulas has an intuitive interpretation. A given observation may have a contribution to the evolution of upper tail dependence λ_U or the lower tail dependence λ_L . The contributions are measured in terms of the likelihood of each mixture component vis-a-vis the total likelihood, as summarized by the weights $w_{i,t}$. As a result, observations that cluster in the upper tail automatically contribute to the evolution of λ_U , and similarly in the lower tail for λ_L . By contrast, the framework of Patton (2006) for the symmetrized copula cannot make automatic use of such features, as its driving mechanism is given by averages of $|u_{it} - u_{jt}|$ for both upper and lower tail dependence.

To illustrate the differences between the time-varying dependencies implied by Patton and GAS, we carry out a simulation experiment. We generate data from the symmetrized Clayton copula. The lower tail dependence coefficient λ_L follows a time-varying sinusoidal pattern. The time-varying pattern of λ_U is also sinusoidal but with a period that is half as long. It is difficult for a model with a uniform driving mechanism to capture both upper and lower tail dependence dynamics within a single model. We present the results in Figure 3 for the GAS model with

$S_t = \mathcal{J}_{t|t-1}$, where $\mathcal{I}_{t|t-1}$ is computed numerically. In case of the updating mechanism of Patton in (30), the smoothing parameter m takes the values $m = 1$ and $m = 10$.

We find that the Patton driving mechanism based on averages of $|u_{it} - u_{jt}|$ only captures some of the variation in the dependence coefficients. It has difficulty in capturing the upper and lower tail dependence dynamics simultaneously, since the same mechanism applies to both types of dependence. The GAS specification is more successful in tracking both types of dynamics. The GAS(1, 1) estimate of the copula parameters is noisier for high values of λ_U or λ_L compared to the one obtained from the Patton model, but GAS captures the true dependence pattern more closely. For low values of the copula parameter, the GAS estimates are also more smooth.

5 Conclusions

We have introduced the generalized autoregressive score (GAS) model. The GAS model is a uniformly applicable observation driven model specification to capture the time variation in parameters. We have shown how GAS models encompass other well-known models, such as generalized autoregressive conditional heteroskedasticity models and autoregressive conditional duration and intensity models as well as multiplicative error models and single source of error models. The advantage of the GAS model is that it exploits the full likelihood information. By taking a scaled (local density) score step as a driving mechanism, the time-varying parameter automatically reduces its one-step ahead prediction error at the current observation with respect to current parameter values. Since it is based on a different paradigm, the GAS model provides an alternative to other observation driven models as well as parameter driven models. We have illustrated our framework by empirical and simulated examples based on models for multivariate marked point processes and models with time-varying copula functions. These illustrations show that our framework can provide interesting alternative specifications for parameter driven models with time-varying parameters.

Appendix: GAS specification for Clayton copula

Consider the Clayton copula as a member of the Archimedean family and given by

$$C(u_{1t}, \dots, u_{N,t}; \lambda) = c_t(\lambda)^{-1/\lambda}, \quad c_t(\lambda) = 1 - N + \sum_{i=1}^N u_{it}^{-\lambda}, \quad (33)$$

where N is the dimension of the observation vector $y_t = (y_{1t}, \dots, y_{N,t})'$, $u_{i,t}$ is the probability integral transform based on the univariate marginal density function of $y_{i,t}$ and parameter λ determines the dependence in y_t . In particular, the tail dependence coefficient λ measures the probability of joint extreme exceedances. Low values of λ indicate high levels of dependence. A particular feature of the Clayton copula is that extreme joint crashes receive positive probability while joint extreme upward shocks obtain zero probability.

The Clayton copula has logdensity

$$\ln C^*(u_{1t}, \dots, u_{N,t}) = - \left(\frac{1}{f_t} + N \right) \ln c_t(f_t) + \sum_{i=1}^N \ln(1 - i \cdot f_t) - (f_t + 1) \sum_{i=1}^N \ln u_{it}, \quad (34)$$

with time-varying dependence parameter $f_t = \lambda_t$. For the GAS updating equation we require

$$\nabla_t = \frac{1}{f_t^2} \ln c_t(f_t) - \sum_{i=1}^N \left[\frac{1}{1 - i \cdot f_t} - \ln(u_{it}) \right] + \left(\frac{1}{f_t} + N \right) c_t(f_t)^{-1} \sum_{i=1}^N u_{it}^{-f_t} \ln(u_{it}). \quad (35)$$

We require a closed-form expression for the information matrix. An alternative method is to compute the information matrix numerically. Here, the information matrix is given by

$$\mathcal{I}_{t|t-1} = \mathbb{E}_{t-1} [(\nabla_t)^2] \equiv h(f_t), \quad (36)$$

with ∇_t given by (35) and where $h(\cdot)$ does not depend on time or on any parameter other than f_t . For the numerical evaluation of (36), we construct a grid of values $f_t^{(0)} < \dots < f_t^{(n)}$ for some positive integer n and compute the function value $h(f_t^{(j)})$ at each of the grid points $j = 0, \dots, n$. Values at intermediate points can be obtained by cubic spline interpolation or non-parametric kernel smoothing to ensure continuity of first and second derivatives of the likelihood function. The numerical procedure is then as follows. For a given value of the parameter vector $\theta = \theta^*$, set $t = 1$ and have a starting value f_1 . Compute $h(f_t)$ via interpolation and use it to scale the score step, that is $s_1 = \nabla_1 / h(f_1)$ for a scalar f_t . Obtain the new parameter value f_{t+1} from the GAS update equation (2) and compute $h(f_{t+1})$

via interpolation. When the information matrix is numerically evaluated, the matrix decomposition $\mathcal{J}'_{t|t-1}\mathcal{J}_{t|t-1} = \mathcal{I}_{t|t-1}^{-1}$ is computed to set the scaling matrix to $S_t = \mathcal{J}_{t|t-1}$. This process is repeated for each t . Finally, the log-likelihood function can be computed for $\theta = \theta^*$.

References

- Baillie, R. T., T. Bollerslev, and H. O. Mikkelsen (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 74, 3–30.
- Bauwens, L. and N. Hautsch (2006). Stochastic conditional intensity processes. *Journal of Financial Econometrics* 4(3), 450.
- Benjamin, M. A., R. A. Rigby, and M. Stanispoulos (2003). Generalized autoregressive moving average models. *Journal of the American Statistical Association* 98(461), 214–223.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31, 307–327.
- Bollerslev, T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. *The Review of Economics and Statistics* 69(3), 542–547.
- Cipollini, F., R. F. Engle, and G. M. Gallo (2006). Vector multiplicative error models: representation and inference. *NBER Working Paper*.
- Cox, D. R. (1981). Statistical analysis of time series: some recent developments. *Scandinavian Journal of Statistics* 8, 93–115.
- Creal, D. D., S. J. Koopman, and A. Lucas (2010). A dynamic multivariate heavy-tailed model for time-varying volatilities and correlations. Tinbergen Institute Discussion paper, 032/2.
- Davis, R. A., W. T. M. Dunsmuir, and S. Streett (2003). Observation driven models for Poisson counts. *Biometrika* 90(4), 777–790.
- Dias, A. and P. Embrechts (2004). Dynamic copula models for multivariate high-frequency data in finance. *Manuscript, ETH Zurich*.
- Duffie, D., A. Eckner, G. Horel, and L. Saita (2006). Frailty correlated default. *manuscript Stanford University*.

- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50(4), 987–1007.
- Engle, R. F. (2002a). Dynamic conditional correlation: a simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business and Economic Statistics* 20(3), 339–350.
- Engle, R. F. (2002b). New frontiers for ARCH models. *Journal of Applied Econometrics* 17, 425–446.
- Engle, R. F. and T. Bollerslev (1986). Modelling the persistence of conditional variances. *Econometric Reviews* 5(1), 1–50.
- Engle, R. F. and G. M. Gallo (2006). A multiple indicators model for volatility using intradaily data. *Journal of Econometrics* 131, 3–27.
- Engle, R. F. and S. Manganelli (2004). CAViaR: conditional autoregressive value at risk by regression quantiles. *Journal of Business & Economic Statistics* 22(4), 367–382.
- Engle, R. F. and J. R. Russell (1998). Autoregressive conditional duration: a new model for irregularly spaced transaction data. *Econometrica* 66(5), 1127–1162.
- Fiorentini, G., G. Calzolari, and L. Panattoni (1996). Analytic derivatives and the computation of GARCH estimates. *Journal of Applied Econometrics* 11, 399–417.
- Hamilton, J. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* 57(2), 357–384.
- Harvey, A. C. (1989). *Forecasting, structural time series models and the Kalman filter*. Cambridge, UK: Cambridge University Press.
- Harvey, A. C. and T. Chakravarty (2008). Beta-t-(E)GARCH. Unpublished manuscript, Faculty of Economics, Cambridge University.
- Hosking, J. R. M. (1981). Fractional differencing. *Biometrika* 68(1), 165–176.
- Koopman, S. J., A. Lucas, and A. Monteiro (2008). The multi-state latent factor intensity model for credit rating transitions. *Journal of Econometrics* 142, 399–424.
- Koopman, S. J., A. Lucas, and B. Schwaab (2008). Forecasting cross-sections of frailty-correlated default. *Tinbergen Institute Discussion paper 08-029/4*.

- McNeil, A., R. Frey, and P. Embrechts (2005). *Quantitative Risk Management*. Princeton, New Jersey: Princeton University Press.
- Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: a new approach. *Econometrica* 59, 347–370.
- Nelson, D. B. and D. P. Foster (1994). Asymptotic filtering theory for univariate ARCH models. *Econometrica* 62(1), 1–41.
- Patton, A. J. (2006). Modelling asymmetric exchange rate dependence. *International Economic Review* 47(2), 527–556.
- Russell, J. R. (2001). Econometric modeling of multivariate irregularly-spaced high-frequency data. *Unpublished manuscript, University of Chicago, Graduate School of Business*.
- Russell, J. R. and R. F. Engle (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.
- Shephard, N. (1995). Generalized linear autoregressions. *Unpublished manuscript, Nuffield College, University of Oxford*.
- Shephard, N. (2005). *Stochastic Volatility: Selected Readings*. Oxford: Oxford University Press.
- Shephard, N. and M. K. Pitt (1997). Likelihood analysis of non-Gaussian measurement time series. *Biometrika* 84, 653–67.
- van den Goorbergh, R. W. J., C. Genest, and B. J. M. Werker (2005). Bivariate option pricing using dynamic copula models. *Insurance Mathematics and Economics* 37(1), 101–114.

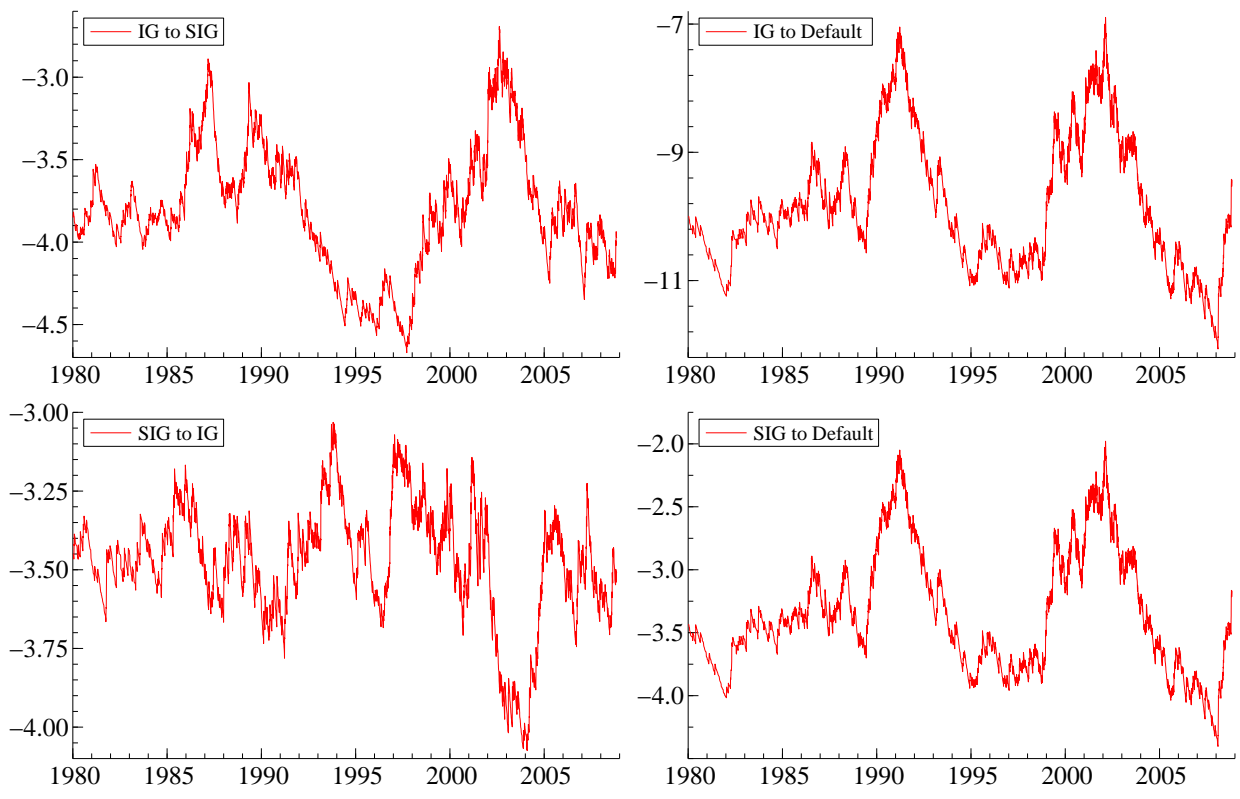


Figure 1: *Marked point-process illustration: the estimated intensities for each transition type. The results are obtained from the GAS(1,1) model with three time-varying factors in (24) and based on the data-set of Moody's ratings for all US corporates between January 1980 and June 2008.*

Table 1: Estimation results for the factor marked point process models with $r = 1, 3$

θ	j, k	1 factor		j, k	3 factors	
d_j	1	-3.866	(0.129)	1	-3.851	(0.147)
	2	-9.961	(0.921)	2	-9.995	(0.916)
	3	-3.463	(0.061)	3	-3.451	(0.078)
	4	-3.422	(0.180)	4	-3.432	(0.176)
$Z_{j,k}$	1, 1	0.712	(0.082)	2, 3	2.133	(1.118)
	2, 1	2.143	(1.201)			
	3, 1	-0.281	(0.077)			
$A_{1,jj}$	1	0.033	(0.004)	1	0.029	(0.006)
				2	0.023	(0.007)
				3	0.036	(0.005)
$B_{1,jj}$	1	0.998	(0.001)	1	0.998	(0.001)
				2	0.995	(0.002)
				3	0.998	(0.001)
log-lik		-12027.7			-12006.4	
nr.par		9			11	

Estimation results for the parameters in the marked point-process model (24) with $r = 1, 3$, (25) and (2) with $\omega = 0$ and s_t given by (26) using Moody's ratings of all US corporates between January 1980 and June 2008. The estimates are reported with asymptotic standard errors in parantheses next to the estimates. Notation: d_j is the j th element of d , $Z_{j,k}$ is the (j, k) element of Z , $A_{1,jj}$ is the j th diagonal element of A_1 and $B_{1,jj}$ is the j th diagonal element of B_1 .

Table 2: Estimation results for different dynamic copula models

	$10^3 \cdot \omega$		A_1		$\ln(B_1 / 1 - B_1)$		B_1		log-lik
German Mark (Euro)–US \$, Japanese Yen–US \$									
GAS	6.11	(2.48)	0.058	(0.009)	5.30	(0.37)	0.995	(0.990, 0.998)	1218.16
Patton	-1.60	(0.85)	0.036	(0.003)	4.27	(0.10)	0.986	(0.983, 0.989)	1191.51
German Mark (Euro)–US \$, British Pound–US \$									
GAS	12.55	(3.55)	0.082	(0.008)	4.97	(0.26)	0.993	(0.988, 0.996)	2218.82
Patton	-0.97	(0.84)	0.025	(0.002)	4.71	(0.11)	0.991	(0.989, 0.993)	2090.42

Parameter estimates for the GAS and Patton models in (28)–(29). The data are the marginal AR-GARCH transforms of log exchange rates for the German Mark-US dollar and Japanese Yen-US dollar (left panel) and for the German Mark-US dollar and British Pound-US dollar (right panel), January 1986–August 2008. The asymmetric confidence interval is in parentheses for B_1 , otherwise the standard error is in parentheses.

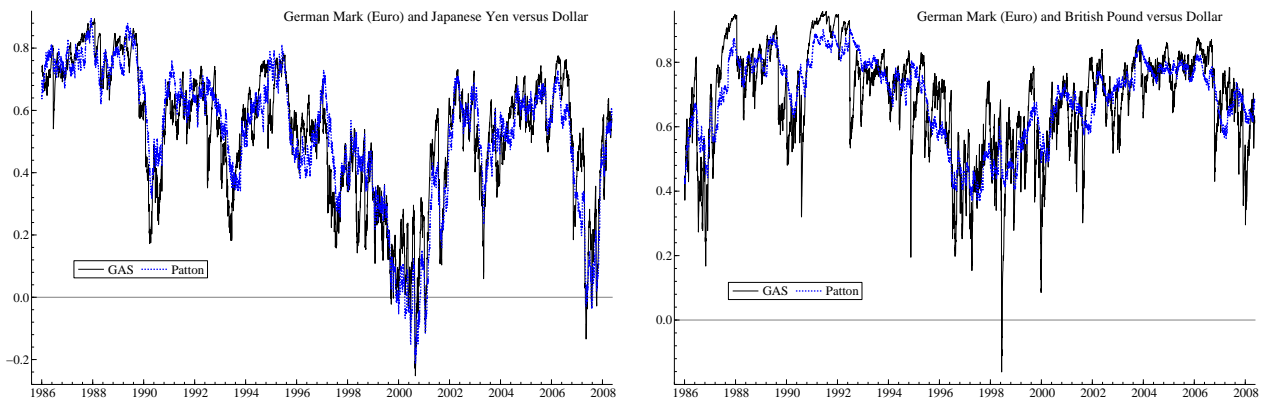


Figure 2: A copula illustration: comparisons of the correlation parameter estimates for the GAS and Patton models in (28)–(29). The data are the marginal AR-GARCH transforms of log exchange rates for the German Mark-US dollar and Japanese Yen-US dollar (left panel) and for the German Mark-US dollar and British Pound-US dollar (right panel). The sample period is January 1986–August 2008.

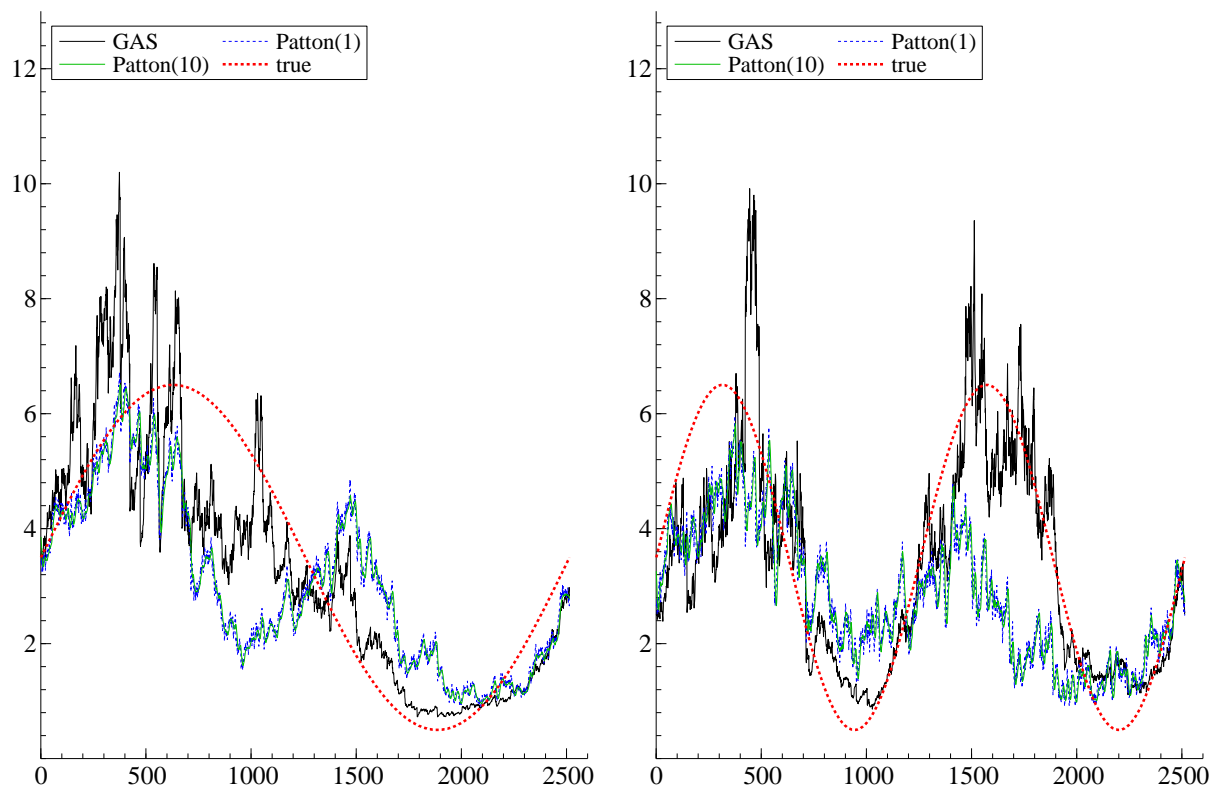


Figure 3: *Symmetrized Clayton copula illustration: comparisons between the estimates of the copula parameters λ_L and λ_U from the GAS framework and the Patton model based on a simulated data set. The copula is the symmetrized Clayton specification from the Appendix.*