JOURNAL OF APPLIED ECONOMETRICS
J. Appl. Econ. **28**: 777–795 (2013)
Published online 20 January 2012 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/jae.1279

GENERALIZED AUTOREGRESSIVE SCORE MODELS WITH APPLICATIONS

DREW CREAL[a]*, SIEM JAN KOOPMAN[b,d] AND ANDRÉ LUCAS[c,d]

[a] *University of Chicago Booth School of Business, Chicago, IL, USA*
[b] *Department of Econometrics, VU University Amsterdam, Netherlands*
[c] *Department of Finance, VU University Amsterdam, and Duisenberg School of Finance, Amsterdam, Netherlands*
[d] *Tinbergen Institute, Amsterdam, Netherlands*

SUMMARY

We propose a class of observation-driven time series models referred to as generalized autoregressive score (GAS) models. The mechanism to update the parameters over time is the scaled score of the likelihood function. This new approach provides a unified and consistent framework for introducing time-varying parameters in a wide class of nonlinear models. The GAS model encompasses other well-known models such as the generalized autoregressive conditional heteroskedasticity, autoregressive conditional duration, autoregressive conditional intensity, and Poisson count models with time-varying mean. In addition, our approach can lead to new formulations of observation-driven models. We illustrate our framework by introducing new model specifications for time-varying copula functions and for multivariate point processes with time-varying parameters. We study the models in detail and provide simulation and empirical evidence. Copyright © 2012 John Wiley & Sons, Ltd.

Received 21 June 2011; Revised 24 August 2011

Supporting information may be found in the online version of this article.

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) models 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 *et al*. (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

* Correspondence to: Drew Creal, University of Chicago Booth School of Business, Chicago, IL, USA.
E-mail: Drew.Creal@chicagobooth.edu

of parameter-driven models 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 *et al.* (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 (MCMC) (see, for example, Shephard and Pitt (1997).

The main contribution of this paper is the development of a framework for time-varying parameters which is based on the score function of the predictive model density at time t. We will argue that the score function is an effective choice for introducing a driving mechanism for time-varying parameters. In particular, by scaling the score function appropriately, standard observation-driven models such as the GARCH, ACD, and ACI models can be recovered. Application of this framework to other nonlinear, non-Gaussian, possibly multivariate, models will lead to the formulation of new observation driven models.

We refer to our observation-driven model based on the score function as the generalized autoregressive score (GAS) model. The GAS model has the advantages of other observation-driven models. Likelihood evaluation is straightforward. Extensions to asymmetric, long memory, and other more complicated dynamics can be considered without introducing further complexities. Since the GAS model is based on the score, it exploits the complete density structure rather than means and higher moments only. It differentiates the GAS model from other observation-driven models in the literature, such as the generalized autoregressive moving average models of Shephard (1995) and Benjamin *et al.* (2003) and the vector multiplicative error models of Cipollini *et al.* (2006).

In our first illustration, we develop new models for time-varying copulas. The copula function provides an important tool for the econometrics of financial risk measurement. Patton (2006) introduced the notion of time-varying copulas and provided the main properties of dynamic copula functions. Other models for time-varying copulas include Giacomini *et al.* (2007), who developed locally constant copula models, and the stochastic copula model of Hafner and Manner (2011). Another interesting copula-based model is developed by Lee and Long (2009) where the multivariate GARCH model is extended with copula functions to capture any remaining dependence in the volatility of the time series. An extended review of the recent developments of copula functions in time series models is given by Patton (2009).

In our second illustration, we create a new class of multivariate point-process models for credit risk. Models for counterparty default and rating transition risk are an important element in the current regulatory system for financial institutions. Many of the new models are based on marked point processes with time-varying intensities for different levels of risk. Parameter estimation relies on computationally demanding methods (see for example, Duffie *et al.* (2009). One of the main challenges when modeling credit events is the sparse number of transitions for each individual company. We show how a multi-state model for pooled marked point processes follows naturally within our framework. We analyze an extensive data set of Moody's rating histories of more than 8000 US corporates over a time span of almost 30 years. We compare the results of the GAS model with those of its parameter-driven counterpart. The parameters in the benchmark model need to be estimated using a MCMC method, which is computationally more demanding compared to our maximum likelihood procedure. Despite the substantial differences in computing time, the GAS model produces almost identical estimates of time-varying defaults and rating transition probabilities when compared with those of the parameter-driven model.

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 describes several new copula models with time-varying parameters. Section 4 presents the model for marked point processes with time-varying parameters. Section 5 concludes.

DOI: 10.1002/jae

2. MODEL SPECIFICATION AND PROPERTIES

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 maximum likelihood estimation and model specification.

2.1. Basic Model Specification

Let $N \times 1$ vector y_t denote the dependent variable of interest, f_t the time-varying parameter vector, x_t a vector of exogenous variables (covariates), all at time t, and θ a vector of static parameters. Define $Y^t = \{y_1, \ldots, y_t\}$, $F^t = \{f_0, f_1, \ldots, f_t\}$, and $X^t = \{x_1, \ldots, x_t\}$. The available information set at time t consists of $\{f_t, F_t\}$, where

$$\mathcal{F}_t = \left\{ Y^{t-1}, F^{t-1}, X^t \right\}, \quad \text{for } t = 1, \ldots, n$$

We assume that y_t is generated by the observation density

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

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

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

where ω is a vector of constants, coefficient matrices A_i and B_j have appropriate dimensions for $i = 1, \ldots, p$ and $j = 1, \ldots, q$, while s_t is an appropriate function of past data, $s_t = s_t(y_t, f_t, F_t; \theta)$. 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 $i = 1, \ldots, p$ and $j = 1, \ldots, q$. The main contribution of this paper is the particular choice for the driving mechanism s_t that is applicable over a wide class of observation densities and nonlinear 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 f_t to the next period $t+1$ using (2) with

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

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 (1–4) define the generalized autoregressive score model with orders p and q. We may abbreviate the resulting model as GAS(p, q).

The use of the score for updating f_t is intuitive. It defines a steepest ascent direction for improving the model's local fit in terms of the likelihood or density 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 of the 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 warrants separate inspection.

DOI: 10.1002/jae

In many situations, it is natural to consider a form of scaling that depends on the variance of the score. For example, we can define the scaling matrix as

$$S_t = \mathcal{I}_{t|t-1}^{-1}, \quad \mathcal{I}_{t|t-1} = \mathrm{E}_{t-1}\left[\nabla_t \nabla_t'\right] \tag{4}$$

where E_{t-1} denotes an expectation with respect to $p(y_t|f_t, F_t; \theta)$. For this choice of S_t, the GAS model encompasses the well-known observation-driven GARCH model of Engle (1982) and Bollerslev (1986), the ACD model of Engle and Russell (1998), and the ACI model of Russell (2001), as well as most of the Poisson count models considered by Davis *et al*. (2003). Another possibility that we consider in this paper is the GAS model with scaling matrix

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

where S_t is defined as the square root matrix of the (pseudo-)inverse information matrix for (1) with respect to f_t. An advantage of this specific choice for S_t is that the statistical properties of the corresponding GAS model become more tractable. This follows from the fact that for $S_t = J_{t|t-1}$ the GAS step s_t has constant unit variance.

Another convenient choice is $S_t = \mathrm{I}$. The GAS model then captures models such as the autoregressive conditional multinomial (ACM) model of Russell and Engle (2005) or the GARMA models of Benjamin et al. (2003). In the context of a fully generic observation density $p(y_t|f_t, F_t; \theta)$, 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 nonlinear 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 the fractionally integrated GAS model specification in the same vein as the well-known ARFIMA and FIGARCH models (see the contributions of Hosking, 1981, and Baillie *et al*., 1996, respectively).

2.2. Special Cases 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: 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 $f_t = \sigma_t^2$ reduces to

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

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

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

DOI: 10.1002/jae

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 v degrees of freedom and unit variance, the GAS(1, 1) specification for the conditional variance leads to the updating equation

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

In case $v^{-1} = 0$, the Student's t distribution reduces to the Gaussian distribution and update (8) collapses to (6) as required. The recursion in (8), however, has an important difference from 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 on the right-hand side of (8) causes a more moderate increase in the variance for a large realization of $|y_t|$ as long as v 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 to (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. Recently, Creal *et al*. (2011) have extended this model to the fully multivariate case with further generalizations and compared it to the popular DCC model of Engle (2002a).

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 $f_t = log\left(\sigma_t^2\right)$, the GAS step takes the form

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

where $1_A(x)$ is the indicator function for the set A; i.e. $1_A(x) = 1$ 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

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

where $\tilde{\vartheta}_1 = \left(\vartheta^2 - 1\right)/\vartheta$ and $\tilde{\vartheta}_2 = \left(\vartheta^2 + 1\right)/\vartheta$. Specification (10) is equivalent to the driving mechanism of the EGARCH model of Nelson (1991), who used the generalized error distribution (GED) instead of the asymmetric Laplace described here.

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

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

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

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

DOI: 10.1002/jae

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 $\alpha = 1$. Hence ACD and ACI models are special cases of the MEM class. The ACD model of Engle and Russell (1998) follows directly from (11) for $\alpha = 1$ and factor recursion (12). In the case where we specify the exponential density in terms of its intensity rather than its expected duration, we obtain $p(y_t|\lambda_t) = \lambda_t \exp(-\lambda_t y_t)$ with intensity $\lambda_t = 1/\mu_t$. Let $\tilde{f}_t = log(\lambda_t)$, and the GAS $(1, 1)$ updating equation becomes

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

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

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

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 $S_t = \mathcal{I}_{t|t-1}^{-1}$ is given by

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

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). Additional examples of GAS models can be found in the supplementary appendix available on the journal's website.

2.3. Maximum Likelihood Estimation

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 $y_1, \ldots, y_n$ and by adopting the standard prediction error decomposition, we can express the maximization problem as

$$\hat{\theta} = \arg\max_{\theta} \sum_{t=1}^{n} \ell_t \tag{15}$$

where $\ell_t = \ln\ p(y_t|f_t, F_t; \theta)$ for a realization of y_t. Evaluating the log-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

DOI: 10.1002/jae

Fiorentini *et al*. (1996). In the case of the GAS(1, 1) specification, the gradient is computed via the chain rule, i.e.

$$\frac{\partial \ell_t}{\partial \theta'} = \frac{\partial \ln p_t}{\partial \theta'} + \frac{\partial \ln p_t}{\partial f_t'} \cdot \frac{\partial f_t}{\partial \theta'} \tag{16}$$

with $p_t = p(y_t | f_t, F_t; \theta)$ and

$$\frac{\partial f_t}{\partial \theta'} = \frac{\partial \omega}{\partial \theta'} + A_1 \frac{\partial s_{t-1}}{\partial \theta'} + B_1 \frac{\partial f_{t-1}}{\partial \theta'} + \left(s_{t-1}' \otimes \mathrm{I}\right) \frac{\partial \overrightarrow{A_1}}{\partial \theta'} + \left(f_{t-1}' \otimes \mathrm{I}\right) \frac{\partial \overrightarrow{B_1}}{\partial \theta'} \tag{17}$$

$$\frac{\partial s_{t-1}}{\partial \theta'} = S_{t-1} \frac{\partial \nabla_{t-1}}{\partial \theta'} + \left(\nabla_{t-1}' \otimes \mathrm{I}\right) \frac{\partial \overrightarrow{S_{t-1}}}{\partial \theta'} \tag{18}$$

where $\overrightarrow{A} = \mathrm{vec}(A)$ 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 $\partial \overrightarrow{S_{t-1}} / \partial \theta'$ should also consider the effect of θ through f_t as in (16).

The log-likelihood derivatives can be computed simultaneously with the time-varying parameters f_t. The analytic derivatives, particularly for (18), may be cumbersome to compute in specific cases. We then turn to likelihood maximization based on numerical derivatives.

We propose to compute standard errors and t-values for the estimated parameters based on the inverse Hessian of the log-likelihood evaluated at the optimum. In particular, if θ gathers all static parameters of the model, we conjecture that under suitable regularity conditions such as those of White (1994) and Wooldridge (1994), the maximum likelihood estimator $\hat{\theta}$ of θ is consistent and satisfies

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

where $H = lim_{n \to \infty} \mathrm{E}[(\partial \ell / \partial \theta)(\partial \ell / \partial \theta')]/n$ and $\ell = \sum_{t=1}^{n} \ell_t$. A formal proof of these results for the general class of GAS models is beyond the scope of the present paper. The results have been established for specific subclasses of GAS models. For example, Davis *et al*. (2005) prove consistency and asymptotic normality of the ML estimator for first-order Poisson count models. Straumann and Mikosch (2006) provide a set of conditions for consistency and asymptotic normality for the Gaussian GARCH model and for more general GARCH specifications. The main challenges for proving the result for the general class of GAS models lie in verifying the stochastic equicontinuity of the likelihood function and in establishing a contracting property for the nonlinear stochastic recurrence equation (2). A contracting property is needed to prove the stationarity and ergodicity of the data-generating process.

A nice feature of the model is that under the assumption of a correct model specification the series s_t forms a martingale difference series, $\mathrm{E}_{t-1}[s_t] = 0$. In particular, if we set the scaling matrix $S_t = J_{t|t-1}$, s_t is a martingale difference with unit variance. If we then express the updating equation for GAS(1, 1) in its infinite-order moving-average form, we obtain

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

Therefore, it is necessary for the covariance stationarity of f_t that the roots of B_1 lie inside the unit circle. Such necessary conditions are helpful for establishing the limiting distribution results mentioned above. For other choices of S_t, the derivation of such properties is less evident.

DOI: 10.1002/jae

2.4. Parameterizations

The GAS specification adapts naturally to different parameterizations of the observation density (1). In the GARCH example of Section 2.2, for example, the time-varying parameter is $f_t = \sigma_t^2$. If 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 the information set $\mathcal{F}_t$. 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

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

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

$$\tilde{\nabla}_t = \frac{\partial \ln p(y_t|f_t, \mathcal{F}_t; \theta)}{\partial \tilde{f}_t} = \left(\dot{h}_t'\right)^{-1}\nabla_t \tag{20}$$

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

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

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

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

where the last equality follows from (19). The choice of parameterization therefore only has a minor effect on the form of the updating step s_t if we adopt $J_{t|t-1}$ 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 $\mathcal{I}_{t|t-1}^{-1}$, it is easy to derive that the updating step $\tilde{s}_t$ for $\tilde{f}_t$ equals $\tilde{s}_t = \dot{h}_t s_t$.

3. DYNAMIC COPULA MODELS

In this section, we introduce several new dynamic copula models. Patton (2006) introduced the notion of time-varying copulas (for reviews see also Dias and Embrechts, 2004; van den Goorbergh *et al*., 2005; Lee and Long, 2009; Patton, 2009).

3.1. The Dynamic Gaussian Copula Model

Copulas have recently become popular in financial risk management. A copula is a multivariate distribution function over a hypercube with uniform marginals. It can be used to link marginal

DOI: 10.1002/jae

distributions into a multivariate distribution using Sklar's theorem in Sklar (1959). 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. The (Gaussian) correlation parameter ρ_t is modeled by the transformed parameter $\rho_t = [1 - exp(-f_t)]/[1 + exp(-f_t)]$. In Patton (2006), the driving mechanism for the dynamic bivariate Gaussian copula is given by

$$f_{t+1} = \omega + A_1 \cdot \sum_{i=1}^{m} \Phi^{-1}\left(u_{1,t-i+1}\right)\Phi^{-1}\left(u_{2,t-i+1}\right) + B_1 f_t \tag{23}$$

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 (23) 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 (23) reduces to

$$f_{t+1} = \omega + A_1 \cdot y_t + B_1 \cdot f_t \tag{24}$$

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

$$f_{t+1} = \omega + A_1 \frac{2}{\left(1 - \rho_t^2\right)}\left[y_t - \rho_t - \rho_t \frac{(x_t - 2)}{\left(1 + \rho_t^2\right)}\right] + B_1 f_t \tag{25}$$

where $x_t = \Phi^{-1}(u_{1t})^2 + \Phi^{-1}(u_{2t})^2$. The similarities and differences between (24) and (25) 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 additional scaling factor $2/\left(1 - \rho_t^2\right)$ in (25) is a consequence of modeling the transformed correlation parameter f_t rather than ρ_t directly. 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 possible scenarios: we might observe $\Phi^{-1}(u_{1t}) = 1$ and $\Phi^{-1}(u_{2t}) = 1$ or, alternatively, $\Phi^{-1}(u_{1t}) = 0.25$ and $\Phi^{-1}(u_{2t}) = 4$. In both cases, the cross-product term $y_t = 1$ is the same and the recursion in (24) will cause f_{t+1} to be the same regardless of which of the two scenarios 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} will depend on the current value of the correlation ρ_t. If the correlation is positive, the impact on the value 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 $(y_t - \rho_t)$ has a negative value, however, the $(x_t - 2)$ term reinforces the magnitude of the GAS step for negative ρ_t.

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 the 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 inputs for the Gaussian copula model.

DOI: 10.1002/jae

Table I 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

Table I. Estimation results for different dynamic copula models

	$10^3\omega$	A_1	$\ln(B_1/1-B_1)$	B_1	Log-likelihood
German mark (euro)–US $, Japanese yen–US $					
GAS	6.11	0.058	5.30	0.995	1218.16
	(2.48)	(0.009)	(0.37)	(0.990, 0.998)	
Patton	–1.60	0.036	4.27	0.986	1191.51
	(0.85)	(0.003)	(0.10)	(0.983, 0.989)	
German mark (euro)–US $, British pound–US $					
GAS	12.55	0.082	4.97	0.993	2218.82
	(3.55)	(0.008)	(0.23)	(0.988, 0.996)	
Patton	–0.97	0.025	4.71	0.991	2090.42
	(0.84)	(0.002)	(0.11)	(0.989, 0.993)	

Note: Parameter estimates for the GAS and Patton models in (24)–(25). 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) andfor 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.

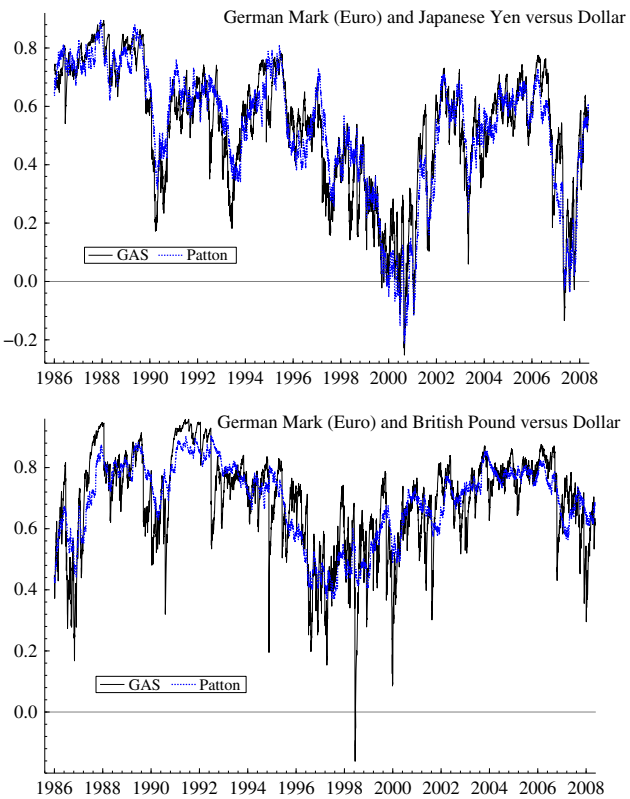

Figure 1. A copula illustration: comparisons of the correlation parameter estimates for the GAS and Patton models in (24)–(25). The data are the marginal AR-GARCH transforms of log exchange rates for the German mark–US dollar and Japanese yen–US dollar (top panel) and for the German mark–US dollar and British pound–US dollar (bottom panel). The sample period is January 1986–August 2008

DOI: 10.1002/jae

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 in Figure 1 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 a major concern.

3.2. Static Mixtures of Dynamic Copulas

Patton (2006) amends the driving mechanism (23) of the dynamic bivariate Gaussian copula towards a generally applicable driving mechanism for copula parameters. The general updating equation of Patton for a bivariate model is given by

$$f_{t+1} = \omega - m^{-1}A_1 \cdot \sum_{i=1}^{m} \left| u_{1,t-i+1} - u_{2,t-i+1} \right| + B_1 \cdot f_t \tag{26}$$

where u_{1t}, u_{2t} and m are defined below (23). 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 (26) is intuitive and simple, two issues are less clear. First, the updating mechanism is not influenced by the particular choice of the copula. As was shown for the Gaussian copula in the previous subsection, particular features of the copula can be useful for the specification of f_t. Second, although (26) 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 methods exist in which the differences $|u_{it} - u_{jt}|$ for $i \neq j$ can be combined to update the dependence parameter. Equation (26) provides little guidance as to how different and possibly conflicting signals should be weighed.

To illustrate how the GAS framework can cope with these issues, consider a static mixture of r dynamic copulas as given by

$$C(u_{1t}, \ldots, u_{Nt}; \lambda_t) = \sum_{i=1}^{r} w_i^* C_i\left(u_{1t}, \ldots, u_{N,t}; \lambda_t^{(i)}\right) \tag{27}$$

where $w_i^* \geq 0$ and $w_1^* + \ldots w_r^* = 1$, $C_i(\cdot)$ is the ith copula function with parameter $\lambda_t^{(i)}$, and $\lambda_t = \left(\lambda_t^{(1)}, \ldots, \lambda_t^{(r)}\right)$. Define $w_{i,t} = w_i^* C_{i,t}^* / \sum_{j=1}^{r} w_j^* C_{j,t}^*$ as the weight of copula i at time t, where $C_{i,t}^*$ is the density function corresponding to copula $C_i\left(u_{1t}, \ldots, u_{N,t}; \lambda_t^{(i)}\right)$, for $i = 1, \ldots, r$. Let f_t represents all time-varying coefficients in $\lambda_t^{(i)}$ for $i = 1, \ldots, r$. The score function is then given by

$$\frac{\partial \ln C_t^*}{\partial f_t} = \sum_{i=1}^{r} w_{i,t} \cdot \frac{\partial \ln C_{i,t}^*}{\partial f_t}$$

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

DOI: 10.1002/jae

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

To obtain the driving mechanism s_t, we note that

$$I_{t|t-1} = -\mathrm{E}_{t-1} \left(\frac{\partial^2 \ln C_t^*}{\partial f_t \partial f_t^{'}} \right) = \mathrm{E}_{t-1} \left[\left(\sum_{i=1}^{r} w_{i,t} \cdot \frac{\partial \ln C_{i,t}^*}{\partial f_t} \right) \left(\sum_{i=1}^{r} w_{i,t} \cdot \frac{\partial \ln C_{i,t}^*}{\partial f_t} \right)^{'} \right] \tag{28}$$

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 $r = 2$. The Appendix discusses the details of the simple Clayton copula ($r = 1$) with a GAS time-varying dependence. The Clayton copula, however, only accounts for lower tail and not for upper tail dependence. To allow for both types of dependence at the same time, we mix the standard Clayton with the survival Clayton as in Patton (2006) to obtain a symmetrized Clayton copula. The first element of the mixture is the standard Clayton copula characterized by the parameter $\lambda_{t,L}$ that accounts for lower tail dependence. The second component of the mixture is the survival Clayton copula and is characterized by the parameter $\lambda_{t,U}$ 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 $\lambda_{t,U}$ or the lower tail dependence $\lambda_{t,L}$. The contributions are measured in terms of the likelihood of each mixture component *vis-à-vis* the total likelihood. As a result, observations that cluster in the upper tail automatically contribute to the evolution of $\lambda_{t,U}$, and similarly in the lower tail for $\lambda_{t,L}$. 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. (Figure 1)

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 $\lambda_{t,L}$ follows a time-varying sinusoidal pattern. The time-varying pattern of $\lambda_{t,U}$ 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 2, where the scaling matrix in (28) is computed numerically. In the case of the updating mechanism of Patton in (26), the smoothing parameter m has taken 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 f_t is noisier compared to the one obtained from the Patton model, but GAS captures the true dependence pattern more closely.

4. DYNAMIC POOLED MARKED POINT PROCESS MODELS

In this section we present a new model for credit risk and rating transitions. We apply the GAS framework based on the square root information matrix, i.e. $S_t = J_{t|t-1}$ as defined in (5), to create a new multivariate point process model with time-varying intensities. Empirical results are presented for a dataset based on Moody's rating histories of all US corporates.

DOI: 10.1002/jae

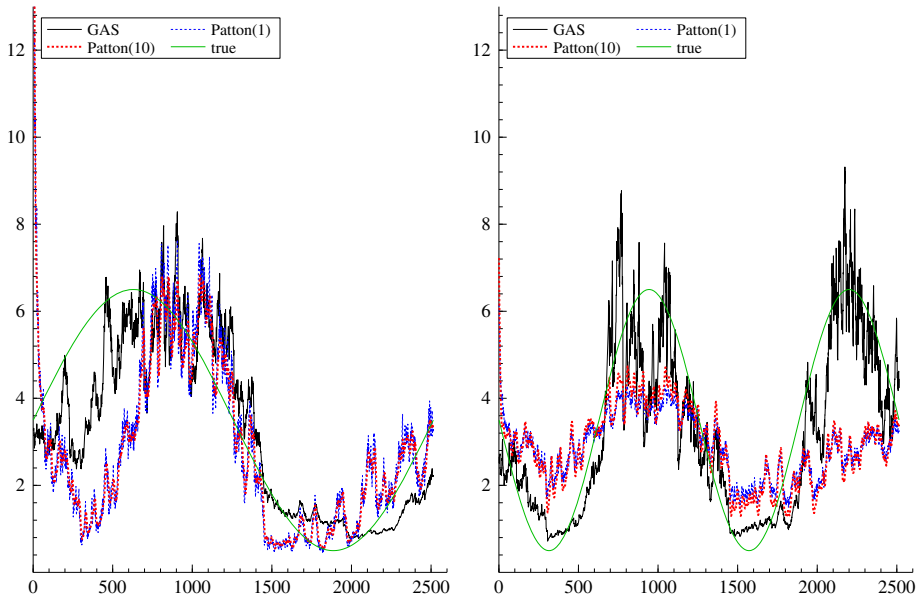

Figure 2. Symmetrized Clayton copula illustration: comparisons between the lower tail ($\lambda_{t,L}$, left panel) and upper tail ($\lambda_{t,U}$, right panel) dependence coefficients from the GAS framework and the Patton model based on a simulated dataset

4.1. Point Process Models

Statistical models with time-varying intensities have received much attention in finance and econometrics. The principal areas of application in economics include intra-day 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.

A number of different models with stochastically evolving intensities have been proposed (see, for example, Bauwens and Hautsch, 2006; Koopman *et al.*, 2008, 2011; Duffie *et al.*, 2009). The econometric treatment of parameter driven models is intricate, while parameter estimation can be computationally demanding regardless of whether frequentist or Bayesian methods are used. In particular, likelihood evaluation for these models requires the computation of high-dimensional integrals using importance sampling techniques or MCMC algorithms. Here we propose an alternative, observation-driven model for time-varying intensities. The formulation of the model follows naturally within the GAS framework.

4.2. The Dynamic Marked Point Process Model

Let $y_{k,t} = \left(y_{1k,t}, \ldots, y_{Jk,t}\right)'$ be a vector of marks of J competing risk processes for firms $k = 1, \ldots, N$. 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 $\lambda_{k,t} = \left(\lambda_{1k,t}, \ldots, \lambda_{Jk,t}\right)'$ be a $J \times 1$ vector of log-intensities. We model the log intensities by

DOI: 10.1002/jae

$$\lambda_{k,t} = d + Zf_t + X_{k,t}\beta \tag{29}$$

where d is a $J \times 1$ vector of baseline intensities, Z is a $J \times r$ matrix of factor loadings, and β is a $p \times 1$ vector of regression parameters for the exogenous covariates $X_{k,t}$. 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 therefore require parameter restrictions for model identification, which we discuss further below.

Model (29) 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 feasible and we typically have $r << N$. This is due to the fact that credit rating events are much sparser than trades in a liquid stock. The former are 1 up to 12 over a span of 30 years, while the latter may be already substantial if only counted within a 5-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 (29) is given by

$$\ell_t = \sum_{j=1}^{J}\sum_{k=1}^{N} y_{jk,t}\lambda_{jk,t} - R_{jk,t}\cdot(t - t^*)\cdot\exp\left(\lambda_{jk,t^*}\right) \tag{30}$$

where $R_{k,t} = \left(R_{1k,t}, \ldots, R_{Jk,t}\right)'$ and $R_{jk,t}$ 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 jth diagonal element $p_{j,t} = \sum_k R_{jk,t}\cdot\exp\left[\lambda_{jk,t}\right] / \sum_{j,k} R_{jk,t}\cdot\exp\left[\lambda_{jk,t}\right] = \mathrm{P}\left[\sum_k y_{jk,t} = 1 \middle| \sum_{j,k} y_{jk,t} = 1\right]$, i.e. the probability that the next event is of type j given that an event happens for firm k. Based on the first and second derivatives of ℓ_t and setting $S_t = J_{t|t-1}$, we obtain the score and scaling matrix

$$\nabla_t = Z'\left(\sum_{k=1}^{N} y_{k,t} - R_{k,t}\cdot(t - t^*)\cdot exp\left(\lambda_{k,t^*}\right)\right) \tag{31}$$

$$S_t = \left(Z'PZ\right)^{-\frac{1}{2}} \tag{32}$$

By combining these basic elements into a GAS specification, we have obtained a new observation-driven model for credit rating transitions. In comparison with its parameter-driven counterparts, parameter estimation for the current model is much easier.

4.3. Application to Moody's Credit Rating Data

For our illustration, we adopt the marked point-process model (29), (30) and (2) with $s_t = S_t \nabla_t$ given by (31) and (32) for a dataset which contains Moody's rating histories of all US corporates over the period January 1981 to March 2010. We set $\omega = 0$ to identify the intercept d in (29). The initial credit ratings for each firm are known at the beginning of the sample and we observe the transitions from one rating category to another over time. Moody's ratings include 21 different categories, some of which are sparsely populated. For the sake of this illustration, therefore, we pool the ratings into a much smaller set of complementary credit classes: investment grade (IG, containing Moody's rating categories Aaa to Baa3), and sub-investment grade (SIG, containing the remaining Moody's rating categories). Default is treated as an absorbing category. This makes for $J = 4$ possible events, which

DOI: 10.1002/jae

are the rating transitions from $j = 1, \ldots, 4$ as IG $\rightarrow$ SIG, IG $\rightarrow$ Default, SIG $\rightarrow$ IG, and SIG $\rightarrow$ Default, respectively.

It is often concluded in credit risk studies that default probabilities are countercyclical. We therefore allow the log-intensities (29) to depend upon the annual growth rate (standardized) of US industrial production as an exogenous variable. This time series has been downloaded from the Federal Reserve Bank of St Louis. Finally, during several days (one in April 1982 and two in October 2007), Moody's redefined several ratings categories. This caused incidental re-rating and resulted in several outliers in the sample. We eliminate the days with outliers from the sample.

Parameter estimates for β, d, Z, A_1 and B_1 are reported in Table II for GAS models with $r = 1, 2, 3$ factors. In order to identify the parameters in Z, we have to impose a number of restrictions. For the different models, we set

$$r = 1 : \begin{bmatrix} * \\ * \\ * \\ 1 \end{bmatrix}; \quad r = 2 : \begin{bmatrix} * & * \\ * & * \\ 1 & 0 \\ 0 & 1 \end{bmatrix}; \quad r = 3 : \begin{bmatrix} 1 & 0 & 0 \\ * & 0 & * \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

where * indicates a coefficient that is estimated. These restrictions imply that for $r = 1$ the factor is common to all transition types and is identified as the sub-investment grade (SIG) to default factor. For $r = 2$, we have a factor for upgrades and another factor for downgrades. For the model with $r = 3$, we have an IG downgrade factor, an IG and SIG default factor, and an SIG upgrade factor.

Table II. Estimation results for the GAS marked point process models

	1 factor		2 factors		3 factors	
θ	j,k		j,k		j,k	
β_j	1	− 0.165 (0.045)	1	− 0.123 (0.069)	1	− 0.137 (0.068)
	2	− 0.510 (0.239)	2	− 0.500 (0.241)	2	− 0.509 (0.241)
	3	0.273 (0.054)	3	0.221 (0.062)	3	0.254 (0.073)
	4	− 0.278 (0.067)	4	− 0.310 (0.069)	4	− 0.314 (0.075)
d_j	1	− 3.708 (0.170)	1	− 3.702 (0.130)	1	− 3.728 (0.141)
	2	− 8.094 (0.389)	2	− 8.084 (0.368)	2	− 8.176 (0.532)
	3	− 3.680 (0.049)	3	− 3.689 (0.075)	3	− 3.693 (0.070)
	4	− 3.289 (0.343)	4	− 3.262 (0.317)	4	− 3.728 (0.141)
$Z_{j,k}$	1, 1	0.492 (0.091)	1, 1	− 1.794 (0.402)	1, 1	1
	2, 1	0.538 (0.618)	2, 1	− 0.903 (1.457)	2, 1	0.576 (0.826)
	3, 1	− 0.075 (0.088)	1, 2	0.331 (0.160)	2, 3	0.347 (0.665)
	4, 1	1	2, 2	0.496 (0.627)	3, 2	1
			3, 1	1	4, 3	1
			4, 2	1		
$A_{1,jj}$	1	0.023 (0.003)	1	0.021 (0.005)	1	0.017 (0.003)
			2	0.021 (0.003)	2	0.009 (0.004)
					3	0.021 (0.003)
$B_{1,jj}$	1	0.996 (0.003)	1	0.990 (0.005)	1	0.993 (0.004)
			2	0.996 (0.003)	2	0.983 (0.011)
					3	0.997 (0.005)
Log-likelihood		− 12806.5		− 12752.1		− 12758.1
AIC		25639.0		25536.2		25547.8
BIC		25711.8		25625.8		25637.4
No. of parameters		13		16		16

Note: Estimation results for the parameters in the marked point process model (29) with $r = \mathbf{1, 2, 3}$, (30) and (2) with $\omega = 0$. The data are Moody's rating histories of all US corporates between January 1981and March 2010. The estimates are reported with asymptotic standard errors in parentheses next to the estimates. Notation: β_j is the jth element of β, d_j is the jth element of d, $Z_{j,k}$ is the (j,k)th element of Z, $A_{1,jj}$ is the jth diagonal element of A_1 and $B_{1,jj}$ is the jth diagonal element of B_1.

DOI: 10.1002/jae

We impose the additional restriction that the upgrade factor from SIG to IG does not impact the IG to default transitions. All of the estimated baseline intensities in d as well as the coefficients in β are significant. The baseline downgrade from IG to default is the smallest, with an estimate of -8.094 for $r=1$. The estimated βs have signs and magnitudes that are consistent with the notion that default intensities are countercyclical. The log-likelihood, AIC, BIC for each of the three models are reported at the bottom of Table II. 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 $r=1$ *versus* the models with $r=2$ and $r=3$ is indicative that more than one factor is needed to fit the data. The GAS parameter B_1 is estimated close to unity for all factors and all specifications, which implies a persistent dynamic process for f_t.

For the one-factor model $9r=1$), we perform a full benchmark analysis for the new GAS model in relation to its parameter-driven counterpart. We therefore take the model of Koopman *et al*. (2008) as our benchmark, hereafter referred to as KLM08. The marked point process KLM08 model has the same observation density (30) as the GAS model. However, the time-varying parameter f_t follows an Ornstein–Uhlenbeck process driven by an independent stochastic process. We refer to Koopman *et al*. (2008) for further details. Parameter estimation for the KLM08 model is more involved than for the GAS model due to the presence of a dynamic, non-predictable stochastic component. In this paper we estimate the KLM08 model using MCMC techniques and compute the one-step-ahead predicted estimates of f_t by means of a particle filter; see Creal (2012) for a detailed survey.

Figure 3 compares the estimates of f_t obtained from the two model specifications. For each of the four possible rating transitions, we plot the intensity of the transition (in basis points on a log scale). These intensities, after dividing them by the number of days in a year, can approximately be interpreted as the daily transition probabilities for each rating transition type. We learn from Figure 3 that the estimates of the time-varying probabilities of the GAS model are almost identical to those of the

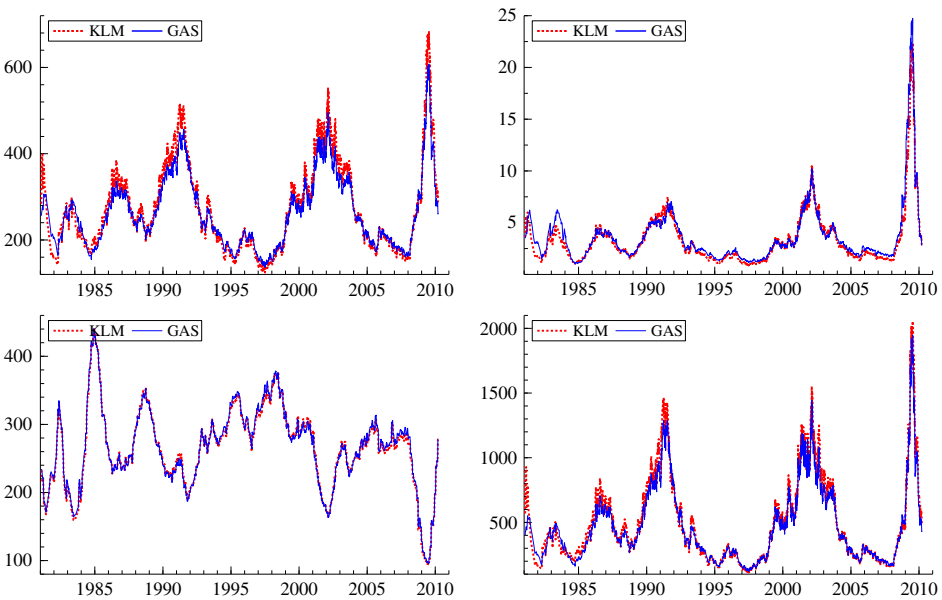

Figure 3. The estimated intensities (in basis points) for each transition type for the one-factor marked point process model. Moody's rating histories are for all US corporates between January 1981 and March 2010

DOI: 10.1002/jae

parameter-driven KLM08 model. However, in our current GAS framework, the results can be obtained without the need for computationally intensive simulation methods required for parameter-driven models. It underlines an attractive feature of our GAS approach.

5. CONCLUSIONS

We have introduced the generalized autoregressive score model as a uniformly applicable observation-driven model specification to capture time variation in parameters. A clear 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. We have illustrated our framework by empirical studies based on dynamic copula models and multivariate marked point process models.

ACKNOWLEDGEMENTS

Lucas acknowledges the financial support of the Dutch Science Foundation (NWO). 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, NAKE 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.

REFERENCES

Baillie RT, Bollerslev T, Mikkelsen HO. 1996. Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **74**(1): 3–30.

Bauwens L, Hautsch N. 2006. Stochastic conditional intensity processes. *Journal of Financial Econometrics* **4**(3): 450–493.

Benjamin MA, Rigby RA, Stanispoulos M. 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**(3): 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, Engle RF, Gallo GM. 2006. Vector multiplicative error models: representation and inference. NBER Working Paper.

Cox DR. 1981. Statistical analysis of time series: some recent developments. *Scandinavian Journal of tatistics* **8**(2): 93–115.

Creal DD. 2012. A survey of sequential Monte Carlo methods for economics and finance. *Econometric Reviews* **31**(3): 245–296.

Creal DD, Koopman SJ, Lucas A. 2011. A dynamic multivariate heavy-tailed model for time-varying volatilities and correlations. *Journal of Business and Economic Statistics* **29**(4): 552–563.

Davis RA, Dunsmuir WTM, Streett S. 2003. Observation driven models for Poisson counts. *Biometrika* **90**(4): 777–790.

DOI: 10.1002/jae

Davis RA, Dunsmuir WTM, Streett S. 2005. Maximum likelihood estimation for an observation driven model for Poisson counts. *Methodology and Computing in Applied Probability* **7**: 149–159.
Dias A, Embrechts P. 2004. Dynamic copula models for multivariate high-frequency data in finance. ETH Zurich.
Duffie D, Eckner A, Horel G, Saita L. 2009. Frailty correlated default. *Journal of Finance* **64**(5): 2089–2123.
Engle RF. 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **50**(4): 987–1007.
Engle RF. 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 RF. 2002b. New frontiers for ARCH models. *Journal of Applied Econometrics* **17**: 425–446.
Engle RF, Bollerslev T. 1986. Modelling the persistence of conditional variances. *Econometric Reviews* **5**(1): 1–50.
Engle RF, Gallo GM 2006. A multiple indicators model for volatility using intra-daily data. *Journal of Econometrics* **131**(1–2): 3–27.
Engle RF, Manganelli S. 2004. CAViaR: conditional autoregressive value at risk by regression quantiles. *Journal of Business and Economic Statistics* **22**(4): 367–382.
Engle RF, Russell JR. 1998. Autoregressive conditional duration: a new model for irregularly spaced transaction data. *Econometrica* **66**(5): 1127–1162.
Fiorentini G, Calzolari G, Panattoni L. 1996. Analytic derivatives and the computation of GARCH estimates. *Journal of Applied Econometrics* **11**: 399–417.
Giacomini E, Härdle W, Spokoiny V. 2007. Inhomogeneous dependence modeling with time-varying copulae. *Journal of Business and Economic Statistics* **27**(2): 224–234.
Hafner CM, Manner H. 2011. Dynamic stochastic copula models: estimation, inference and applications. *Journal of Applied Econometrics* doi: 10.1002/jae.1197.
Hamilton J. 1989. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* **57**(2): 357–384.
Harvey AC. 1989. *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press: Cambridge, UK.
Harvey AC, Chakravarty T. 2008. Beta–t–(E)GARCH. Working Papers in Economics, Faculty of Economics, Cambridge University, UK.
Hosking JRM. 1981. Fractional differencing. *Biometrika* **68**(1): 165–176.
Koopman SJ, Lucas A, Monteiro A. 2008. The multi-state latent factor intensity model for credit rating transitions. *Journal of Econometrics* **142**(1): 399–424.
Koopman SJ, Lucas A, Schwaab B. 2011. Modeling frailty-correlated defaults using many macroeconomic covariates. *Journal of Econometrics* **162**(2): 312–325.
Lee T-H, Long X. 2009. Copula-based multivariate GARCH model with uncorrelated dependent errors. *Journal of Econometrics* **150**(2): 207–218.
Nelson DB. 1991. Conditional heteroskedasticity in asset returns: a new approach. *Econometrica* **59**: 347–370.
Patton AJ. 2006. Modelling asymmetric exchange rate dependence. *International Economic Review* **47**(2): 527–556.
Patton AJ. 2009. Copula-based models for financial time series. In *Handbook of Financial Time Series*, Anderson TG, Davis RA, Kreiss JP, Mikosch T (eds.). Springer: New York; 767–786.
Russell JR. 2001. Econometric modeling of multivariate irregularly-spaced high-frequency data. University of Chicago, Graduate School of Business.
Russell JR, Engle RF. 2005. A discrete-state continuous-time model of financial transactions prices and times: the autoregressive conditional multinomial-autoregressive conditional duration model. *Journal of Business and Economic Statistics* **23**(2): 166–180.
Shephard N. 1995. Generalized linear autoregressions. Nuffield College, University of Oxford.
Shephard N. 2005. *Stochastic Volatility: Selected Readings*. Oxford University Press: Oxford.
Shephard N, Pitt MK. 1997. Likelihood analysis of non-Gaussian measurement time series. *Biometrika* **84**(3): 653–667.
Sklar A. 1959. Fonctions de répartition à n dimensions et leurs marges. Publications de l'Institut de Statistique de L'Université de Paris 8; 229–231.
Straumann D, Mikosch T. 2006. Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: a stochastic recurrence equations approach. *The Annals of Statistics* **34**(5): 2449–2495.
van den Goorbergh RWJ, Genest C, Werker BJM. 2005. Bivariate option pricing using dynamic copula models. *Insurance Mathematics and Economics* **37**(1): 101–114.
White H. 1994. *Estimation, Inference and Specification Analysis*. Cambridge University Press: Cambridge, UK.
Wooldridge JM. 1994. Estimation and inference for dependent processes. In *Handbook of Econometrics*, Engle RF, McFadden DL (eds.). North-Holland: Amsterdam; 2639–2738.

DOI: 10.1002/jae

APPENDIX

GAS SPECIFICATION FOR CLAYTON COPULA

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

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

where N is the dimension of the observation vector $y_t=(y_{1t},\ldots,y_{N,t})'$, $u_{i,t}$ 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 log-density

$$\ln C^{*}\left(u_{1t},\ldots,u_{N,t}\right)=-\left(\frac{1}{f_t}+N\right)\ln \mathrm{c}_t(f_t)+\sum_{i=1}^{N}\ln(1+(i-1)\cdot f_t)-(f_t+1)\sum_{i=1}^{N}\ln u_{it} \tag{34}$$

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

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

The principal difficulty here is to derive a closed-form expression for the information matrix. This difficulty also arises for even simpler copula models. An alternative method is to compute the information matrix numerically. In our current example, the information matrix is given by

$$I_{t|t-1}=\mathrm{E}_{t-1}\left[(\nabla_t)^2\right]\equiv h(f_t) \tag{36}$$

with ∇_t given by (35) and where $h(\cdot)$ does not depend on time or on any parameter other than f_t. For numerical evaluation of (36), we construct a grid of values $f_t^{(0)}<\ldots<f_t^{(n)}$ for some positive integer n and compute the function value $h(f_t^{(j)})$ at each of the grid points $j=0,\ldots,n$. 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 $\theta=\theta^{*}$ and initial value f_1, compute $h(f_1)$ and use it to scale the score step $s_1=\nabla_1/h(f_1)$ for a scalar f_1. Obtain the new parameter value f_2 from the GAS update equation (2) and compute $h(f_2)$ via interpolation. This process is repeated for each t. Finally, the log-likelihood function can be computed for $\theta=\theta^{*}$.

DOI: 10.1002/jae