A General Framework for Observation Driven
Time-Varying Parameter Models*

Drew Creal*, Siem Jan Koopmana,c, André Lucasb,c

(a) Department of Econometrics, VU University Amsterdam
(b) Department of Finance, VU University Amsterdam, and Duisenberg school of finance
(c) Tinbergen Institute, Amsterdam

Abstract

We propose a new class of observation driven time series models referred to as Generalized Autoregressive Score (GAS) models. The driving mechanism of the GAS model is the scaled score of the likelihood function. This approach provides a unified and consistent framework for introducing time-varying parameters in a wide class of non-linear models. The GAS model encompasses other well-known models such as the generalized autoregressive conditional heteroskedasticity, the autoregressive conditional duration, the autoregressive conditional intensity, and the single source of error models. In addition, the GAS specification provides a wide range of new observation driven models. Notable examples include non-linear regression models with time-varying parameters, multivariate point process models with time-varying parameters and pooling restrictions, new models for time-varying copula functions, models for time-varying higher order moments, and observation driven mixture models. We study the properties of GAS models and provide several non-trivial examples of their application.

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, Clive Granger, Andrew Harvey, Marius Ooms and Michel van der Wel for their comments on an earlier draft of the paper. We have also benefited from the comments of participants at the conference "High-Frequency Data Analysis in Financial Markets" in Hitotsubashi University, Tokyo, and at seminar presentations in University of Alicante, NAKE 2008 Utrecht research day, Tinbergen Institute and VU University. Correspondence: André Lucas, VU University Amsterdam, FEWEB, De Boelelaan 1105, 1081 HV Amsterdam, Netherlands, email: alucas@feweb.vu.nl.
1 Introduction

In many settings of empirical interest, time variation in a selection of parameters of a model is important for capturing the dynamic behavior of (multivariate) time series processes. Time series models with time-varying parameters have been categorized by Cox (1981) into two classes: observation driven models and parameter driven models. In this paper we develop a new, general framework for building observation driven time series models. In the observation driven approach, time variation of the parameters is introduced by making the parameters dependent on (functions of) lagged dependent values, exogenous variables, and past observations. Although the parameters are stochastic, they are perfectly predictable given past information. This approach simplifies likelihood evaluation and explains why these models have become popular in the applied econometrics and statistics literature. Typical examples of observation driven 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). Our approach encompasses many of the existing observation driven models as mentioned above. In addition, it allows the formulation of a wide range of new models.

The alternative to observation driven models are parameter driven models. In parameter driven models, the parameters are stochastic processes which are subject to their own source of error. Given past and concurrent observations, the parameters are not perfectly predictable. Typical examples include 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; see, e.g., Shephard and Pitt (1997). However, pa-
Parameter driven models offer a conceptually straightforward way of introducing time-varying parameters in a wide class of non-linear and non-Gaussian models.

The main contribution of this paper is the development of a common 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 are recovered. The scaled score is equally applicable to non-standard multivariate models and leads to the formulation of new observation driven models.

We will refer to our observation driven model with a 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. Other frameworks for observation driven models within the exponential family of distributions have been suggested in the literature, including 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). In contrast to these proposals, GAS models are able to exploit the complete density structure rather than only means and higher moments.

To illustrate the applicability of GAS models, we study a number of interesting, non-trivial settings for observation driven models. We consider linear and non-linear regression models with time-varying coefficients as a typical class of models that we can treat within the GAS framework. An example is the Nelson and Siegel (1987) model for analyzing the term structure of interest rates which emphasizes that GAS can also treat multivariate models. Multivariate non-Gaussian models for pooled marked point-processes with a GAS specification for latent factors driving the log-intensities is a new model specification for credit rating transitions. We also develop new models for time varying copula applications. Another challenging direction in the current literature is the modeling of higher-order moments of financial returns as time-varying processes. We show that GAS provides a generic framework for time-varying higher-
order moments. In particular, we consider linear regression models and GARCH models with Student $t$ distributions where the degrees of freedom parameter is allowed to be time-varying. Finally, we look at observation driven dynamic mixture models and compare these with a Hidden Markov Model (HMM).

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 contains a range of non-trivial examples of GAS models, where we develop several new observation driven models. Section 5 concludes and provides directions for future research.

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 alternative specifications of the model.

2.1 Basic model specification

Let $y_t$ denote the dependent variable of interest, $\psi_t$ the time-varying parameter vector, $x_t$ a vector of exogenous variables (covariates), all at time $t$, and $\theta$ a vector of static parameters. Define $Y^t_1 = \{y_1, \ldots, y_t\}$, $\Psi^t_1 = \{\psi_1, \ldots, \psi_t\}$, and $X^t_1 = \{x_1, \ldots, x_t\}$. The available information set at time $t$ consists of $Y^t_{t-1}$, $\Psi^t_1$ and $X^t_1$. We assume that $y_t$ is generated by the observation density

$$p(y_t | \psi_t, \Psi^t_{t-1}, Y^t_{t-1}, X^t_1; \theta),$$

(1)

for $t = 1, \ldots, n$. Furthermore, we assume that the mechanism for updating the time-varying parameter $\psi_t$ is given by the familiar autoregressive updating equation

$$\psi_t = \omega + \sum_{i=1}^{p} A_i s_{t-i} + \sum_{j=1}^{q} B_j \psi_{t-j},$$

(2)

where $\omega$ 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 appropriately scaled function, which depends on
the set of observations $Y_t^1$, the time-varying parameters in $\Psi_1^t$, and the static parameter vector $\theta$. Furthermore, all unknown coefficients in (2) are functions of $\theta$ as well, that is $\omega = \omega(\theta)$, $A_i = A_i(\theta)$, and $B_j = B_j(\theta)$. Our main contribution is the particular choice for the driving mechanism $s_t$ that is applicable uniformly over a wide class of densities and non-linear models.

Equation (1) differs from a typical parameter driven model specification. In a parameter driven model $\psi_t$ would evolve subject to its own source of error, say $\eta_t$. In particular, $\sum_{i=1}^p A_i s_{t-1}$ would be replaced by $\eta_t$ in (2). Estimation of $\psi_t$ is then based on the conditional (filtered) density $p(\psi_t | Y_t^1, X_t^1, Y_{t-1}^1; \theta)$. For linear Gaussian state space models, this density can be computed in closed form by the Kalman filter. In non-linear and non-Gaussian models, conditional densities are generally evaluated via simulation methods; see, e.g., Durbin and Koopman (2001) and Doucet, de Freitas, and Gordon (2001). The simulations are often most effective when they are based on second order expansions of the log observation density (1).

For observation driven models, we propose to use the same intuition to update the time-varying parameter from $\psi_{t-1}$ to $\psi_t$ via (2) with

$$s_t = S_t \cdot \nabla_t,$$

(3)

where

$$\nabla_t = \partial \ln p(y_t | \psi_t, \Psi_1^{t-1}, Y_1^{t-1}, X_1^t; \theta) / \partial \psi_t, \quad S_t = S(t, \Psi_1^{t-1}, Y_1^{t-1}, X_1^t; \theta),$$

(4)

with time dependent scaling matrix $S(\cdot)$. Given the reliance of the driving mechanism in (2) on the scaled score vector (3), we let the equations (1) – (3) constitute the generalized autoregressive score model with orders $p$ and $q$. We abbreviate the resulting model by GAS($p,q$).

There are several intuitive choices for the scaling matrix that we investigate here. Our first choice is to set $S_t$ equal to the (pseudo)-inverse information matrix based on the density (1), that is

$$S_t = \mathcal{I}_t^{-1}, \quad \mathcal{I}_t = \mathbb{E}_{t-1} (\nabla_t \nabla_t') = -\mathbb{E}_{t-1} \left[ \frac{\partial^2 \ln p(y_t | \psi_t, \Psi_1^{t-1}, Y_1^{t-1}, X_1^t; \theta)}{\partial \psi_t \partial \psi_t'} \right],$$

(5)

where $\mathbb{E}_{t-1}$ is expectation with respect to $p(y_t | \psi_t, \Psi_1^{t-1}, Y_1^{t-1}, X_1^t; \theta)$. The updating mechanism (2) for $\psi_t$ now reduces to something similar to a Newton or (Fisher’s) scoring updating step for every new observation $y_t$ that becomes available. Using this particular choice for scaling the
score vector, 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). When the scaling matrix is the identity matrix, that is $S_t = I$ in (3), the recursion captures models such as the autoregressive conditional multinomial (ACM) model of Russell and Engle (2005). In addition, it gives rise to a number of useful observation driven models that have not been investigated before. We first give some introductory examples of GAS models. In Section 4, we provide a more systematic review of new, non-trivial models within the GAS family.

2.2 Some examples

Example 1 (GARCH model): Consider the basic model $y_t = \sigma_t \varepsilon_t$ where the Gaussian disturbance $\varepsilon_t$ has zero mean and unit variance while $\sigma_t$ is a time-varying standard deviation. It is a basic exercise to show that the GAS(1,1) model with $S_t = I_{t|t-1}$ in (5) and for $\psi_t = \sigma_t^2$ reduces to

$$\psi_t = \omega + A_1 \left(y_{t-1}^2 - \psi_{t-1}\right) + B_1 \psi_{t-1},$$

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

$$\psi_t = \alpha_0 + \alpha_1 y_{t-1}^2 + \beta_1 \psi_{t-1}, \quad \psi_t = \sigma_t^2,$$

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 $\varepsilon_t$ follows a Student $t$ distribution with $\nu$ degrees of freedom scaled to have variance one, the GAS(1,1) specification for the conditional variance leads to the updating equation

$$\psi_t = \omega + A_1 \left(1 + 3\nu^{-1}\right) \left(1 - 2\nu^{-1}\right) \left(1 + \nu^{-1} y_{t-1}^2 / \left(1 - 2\nu^{-1}\right) \psi_{t-1} - \psi_{t-1}^2\right) + B_1 \psi_{t-1}.$$

The update (8) collapses to (6) in case of the Gaussian distribution, that is $\nu^{-1} = 0$. The recursion in (8), however, has an important difference with the standard t-GARCH(1,1) model of Bollerslev (1987) which has the Student $t$ density in (1) and 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 \(\nu\) 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 \(t\) errors therefore is substantially different from its familiar GARCH counterpart.

A similar specification for the Student \(t\) distribution is developed by Harvey and Chakravarty (2008) and its properties are discussed in detail. The GAS framework however also provides a range of alternative GARCH specifications for different heavy-tailed distributions. For example, unreported derivations for the class of Generalized Hyperbolic (GH) distributions for \(\varepsilon_t\) confirm the intuition for the Student \(t\) GAS model (8): under the GH assumption, large absolute realizations \(y_t\) have less impact on volatility increases. Similarly, for right-skewed GH innovations \(\varepsilon_t\), observations in the right-tail cause smaller volatility increases than equally sized observations in the other tail and vice-versa. These properties follow directly from inspecting the score of these models with respect to \(\sigma_t^2\).

**Example 2 (MEM, ACD and ACI models):** Consider the model \(y_t = \mu_t \varepsilon_t\) where \(\varepsilon_t\) has a gamma distribution with density \(p(\varepsilon_t; \alpha) = \Gamma(\alpha)^{-1} \varepsilon_t^{\alpha-1} \alpha^\alpha \exp(-\alpha \varepsilon_t)\). Using a change of variables, we have

\[
p(y_t | \mu_t; \alpha) = \Gamma(\alpha)^{-1} y_t^{\alpha-1} \alpha^\alpha \mu_t^{-\alpha} \exp(-\alpha \frac{y_t}{\mu_t}).
\]

Let \(\psi_t = \mu_t\), then the GAS(1,1) updating equation with \(S_t = T_{t-1}^{-1}\) in (5) is given by

\[
\psi_t = \omega + A_1 (y_{t-1} - \psi_{t-1}) + B_1 \psi_{t-1}. \tag{9}
\]

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 and consequently GAS. The ACD model of Engle and Russell (1998) follows straightforwardly with \(\alpha = 1\) and factor recursion (9). Suppose we parameterize the exponential density in terms of the intensity rather than the expected duration so that \(\lambda_t = 1/\mu_t\) and \(p(y_t | \lambda_t) = \lambda_t \exp(-\lambda_t y_t)\). By defining \(\tilde{\psi}_t = \log(\lambda_t)\), the GAS(1,1) updating equation becomes

\[
\tilde{\psi}_t = \omega + A_1 \left[1 - y_{t-1} \exp(\tilde{\psi}_{t-1})\right] + B_1 \tilde{\psi}_{t-1}, \tag{10}
\]
which is equivalent to the standard ACI(1, 1) model of Russell (2001).

**Example 3 (Regression model):** The linear regression model \( y_t = x'_t \beta_t + \varepsilon_t \) has a \( k \times 1 \) vector \( x_t \) of exogenous variables, a \( k \times 1 \) vector of time-varying regression coefficients \( \beta_t \) and normally distributed disturbances \( \varepsilon_t \sim N(0, \sigma^2) \). Let \( \psi_t = \beta_t \). The scaled score function based on \( S_t = I_{t-1}^{-1} \) in (5) is now given by

\[
s_t = (x'_t x_t)^{-1} x_t (y_t - x'_t \psi_t),
\]

where the inverse of \( I_{t-1} \) is now the Moore-Penrose pseudo inverse to account for the singularity of \( x_tx'_t \). The GAS(1, 1) specification for the time-varying regression coefficient becomes

\[
\psi_t = \omega + A_1(x'_{t-1} x_{t-1})^{-1} x_{t-1} (y_{t-1} - x'_{t-1} \psi_{t-1}) + B_1 \psi_{t-1}.
\]

This time-varying regression updating equation can be extended by including \( \sigma^2 \) as a time-varying factor and recomputing the scaled score function in (11) for the new time-varying parameter vector \( \psi_t = (\beta'_t, \sigma^2_t)' \).

**Example 4 (Dynamic exponential family models):** The exponential family of distributions can be represented by the density function

\[
\exp \left[ \eta(\theta)'T(y_t) - C(\theta) + h(y_t) \right],
\]

with column vector functions \( \eta(\cdot) \) and \( T(\cdot) \), scalar functions \( C(\cdot) \) and \( h(\cdot) \) and parameter vector \( \theta \). We consider replacing \( \theta \) by a time-varying parameter vector \( \theta_t = \Phi \psi_t \) where matrix \( \Phi \) is a \( k_1 \times k_2 \) full column-rank loading matrix with \( k_1 \geq k_2 \), such that \( \theta_t \) is time-varying according to a factor structure. We propose a GAS model specification for the factors \( \psi_t \). The GAS driving mechanism with information matrix scaling (5) is then given by

\[
s_t = (\Phi' I_{t-1} \Phi)^{-1} \Phi' \left[ i'_t T(y_t) - \dot{C}_t \right], \quad \text{with} \quad I_{t-1} = \frac{\partial^2 C(\theta_t)}{\partial \theta_t \partial \theta'_t},
\]

where \( \dot{C}_t = \partial C(\theta_t) / \partial \theta_t \) and \( i'_t = \partial \eta(\theta_t) / \partial \theta'_t \). The resulting dynamic framework applies to any member of the exponential family. Shephard (1995) and Benjamin et al. (2003) propose alternative observation driven models for the subclass of natural exponential family members.
given by (13) with \( \eta(\theta)^T(y_t) = \theta^Ty_t \). In this approach the link function \( E_{t-1}(y_t) = \dot{C}_t \) is made dependent on covariates and autoregressive/moving average terms. The advantage of the GAS model over these alternative specifications is its use of the full density structure to update the time-varying parameters. Explicit formulas for GAS driving mechanisms for a selection of members from the exponential family are provided in Creal, Koopman, and Lucas (2008).

### 2.3 Different GAS specifications

An important advantage of the GAS\((p, q)\) specification is that its applicability is not restricted to one specific model or choice of model parameterization. By contrast, the recursion scheme is applicable to a wide range of models that are characterized by a parametric likelihood specification. The GAS framework is particularly relevant for the applications in Section 4, where we generalize some well-known models with time-varying parameters outside their usual area of application. For example, if the time-varying parameter is common across different observations, the specification in (3) gives an automatic and model consistent way to weight the information provided by different observations.

A useful feature of the GAS updating equation (3) is that under the correct model specification, \( s_t \) is a martingale difference, that is \( E_{t-1}[s_t] = 0 \). This follows directly from the properties of the score vector. Due to scaling by the matrix \( S_t \), we also obtain \( E_{t-1}[s_tS_t'] = S_t \cdot \mathcal{I}_{t|t-1} \cdot S_t' \) which simplifies to \( \mathcal{I}_{t|t-1}^{-1} \) and \( \mathcal{I}_{t|t-1} \) for \( S_t = \mathcal{I}_{t|t-1}^{-1} \) and \( S_t = I \), respectively. The first two moments of the driving mechanism \( s_t \) are therefore easily linked to the theoretical properties of the postulated model density (1).

A key aspect of the GAS model is the scaling of the score. Selection of this matrix defines different GAS models. In this paper, we focus on the inverse of the information matrix due to its relationship with existing models in the literature, e.g. GARCH, ACD, and MEM models. This scaling matrix has the advantage that it emphasizes the local information that is available about the factor. The inverse of the information matrix may however not be full rank or may be numerically unstable for specific models making it potentially difficult to use as the scaling matrix. In this respect, we can refer to Example 3 in subsection 2.2 where we obtain a singular information matrix for the multiple regression model.

Another example is given by a time-varying autoregressive model of order one without
intercept, that is,

\[ y_t = \phi_t y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma^2), \]

where \( \phi_t \) is the time-varying coefficient with GAS specification (2) for \( \psi_t = \phi_t \). The information scaled score step in this case reduces to \( s_t = (y_t - \phi_t y_{t-1})/y_{t-1}^2 \). The GAS updating scheme becomes numerically unstable if \( y_{t-1} \) is close to zero. In this case, the information matrix \( \mathcal{I}_{t|t-1} = \psi_{t-1}^2 \) is close to zero and \( s_t \) can jump to extreme values. To circumvent such numerical instabilities, we introduce a form of information smoothing over the most recent stretch of observations, that is \( S_t = (\mathcal{I}_{t|t-1})^{-1} \) where

\[ \mathcal{I}_{t|t-1} = \alpha \mathcal{I}_{t-1|t-2} + (1 - \alpha) \mathcal{I}_{t|t-1}, \quad (14) \]

for some \( 0 \leq \alpha \leq 1 \). This is an exponentially weighted moving average (EWMA) smoothing scheme. Other weighting schemes for smoothing the information are also possible. The smoothing parameter \( \alpha \) determines the number of observations that \( S_t \) takes into account. For \( \alpha \to 0 \), we recover the standard GAS model with information scaling. For \( \alpha \to 1 \), the model tends to average the information over all past observations. The optimal smoothing parameter could be fixed a priori, or be determined from the data itself by treating \( \alpha \) as part of the static parameter vector \( \theta \) in the likelihood.

The basic dynamics of (2) may be further extended in obvious directions. For example, it may be interesting to include exogenous variables in (2), or to generalize the evolution of \( \psi_t \) to include other non-linearities such as regime-switching. In addition, it may be useful in some applications to consider long-memory versions of (2), for example

\[ \psi_t = \omega + \sum_{i=0}^{\infty} \frac{(i + d - 1)!}{i!(d - 1)!} s_{t-1-i}, \]

for fractional integration parameter \( d < 1/2 \), such that we obtain a fractionally integrated GAS or FIGAS model specification, similar to the ARFIMA and FIGARCH literature, see the seminal paper of Hosking (1981). We leave such extensions for future research.
3 Statistical properties

In this section we explore issues related to maximum likelihood estimation and parameter identification. We also discuss whether standard statistical and asymptotic results apply.

3.1 Estimation and inference

A convenient advantage of observation driven models is the relatively simple way of estimating parameters by maximum likelihood (ML). This advantage 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

\[
\max_{\theta} \sum_{t=1}^{n} \ell(\theta; y_t, \psi_t, \Psi_{t-1}^{l-1}, Y_{t-1}^{l-1}, X_t^l),
\]

where \( \ell(\theta; y_t, \psi_t, \Psi_{t-1}^{l-1}, Y_{t-1}^{l-1}, X_t^l) = \ln p(y_t|\psi_t, \Psi_{t-1}^{l-1}, Y_{t-1}^{l-1}, X_t^l; \theta) \) for an observed value \( y_t \). Similar to the GARCH model, the GAS model defines a filter for the time-varying parameters. This makes likelihood evaluation particularly simple. It only requires the implementation of the GAS updating function (2) and the evaluation of \( p(y_t|\psi_t, \Psi_{t-1}^{l-1}, Y_{t-1}^{l-1}, X_t^l; \theta^*) \) for a particular value \( \theta^* \) of \( \theta \).

It is possible to formulate recursions for computing the gradient of the likelihood with respect to the static parameters \( \theta \). Gradient recursions for the GARCH model have been developed by Fiorentini, Calzolari, and Panattoni (1996). For the GAS(1, 1) specification, we obtain

\[
\frac{d\ell_t}{\partial \theta'} = \frac{\partial \ln p_t}{\partial \theta'} + \frac{\partial \ln p_t}{\partial \psi_t} \frac{\partial \psi_t}{\partial \theta'},
\]

\[
\frac{\partial \psi_t}{\partial \theta'} = A_1 \frac{\partial s_{t-1}}{\partial \theta'} + B_1 \frac{\partial \psi_{t-1}}{\partial \theta'} + (s_{t-1} \otimes I) \frac{\partial \vec{A}_1}{\partial \theta'} + (\psi_{t-1} \otimes I) \frac{\partial \vec{B}_1}{\partial \theta'},
\]

\[
\frac{\partial s_{t-1}}{\partial \theta'} = S_t \frac{\partial \vec{\nabla}_{t-1}}{\partial \theta'} + (\vec{\nabla}_{t-1} \otimes I) \frac{\partial \vec{S}_{t-1}}{\partial \theta'},
\]

where \( \ell_t = \ell(\theta; y_t, \psi_t, \Psi_{t-1}^{l-1}, Y_{t-1}^{l-1}, X_t^l), p_t = p(y_t|\psi_t, \Psi_{t-1}^{l-1}, Y_{t-1}^{l-1}, X_t^l; \theta), \vec{A} = \text{vec}(A) \) denotes the vector with the stacked columns of the matrix \( A \), and \( \otimes \) is the Kronecker matrix product. Higher order GAS specifications can be dealt with similarly by formulating the GAS model updating equation in companion form. The log-likelihood derivatives can be computed simultaneously.
with the time-varying parameters $\psi_t$. 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 at the optimum to compute standard errors and $t$-values for the estimated parameters. In particular, if $\theta$ gathers all static parameters of the model, we conjecture that under standard regularity conditions, the maximum likelihood estimator $\hat{\theta}$ of $\theta$ is consistent and satisfies

$$T^{1/2}(\hat{\theta} - \theta) \xrightarrow{d} N(0, H^{-1}),$$

with $H = -\mathbb{E}(\partial^2 \ell / \partial \theta \partial \theta')$.

It is not clear, however, that standard statistical results apply directly. As an example, even though $\{s_t\}$ forms a martingale difference sequence, it is not directly evident that the GAS(1,1) model will be stable even if $|B_1| < 1$. Since the variance of $s_t$ changes over time in a stochastic way, precise conditions for stability need to be formulated. For example, the GAS specification does not prevent the variance of $s_t$ from becoming unbounded. If the model density is such that the inverse information matrix with respect to $\psi_t$ is uniformly bounded, standard stability results apply for $|B_1| < 1$ in the GAS(1,1) model. This holds for a number of examples we discuss in Section 4. It is clear that given the generality of the GAS specification, the conditions for standard asymptotic theory to apply need to be validated on a case by case basis. We leave this to future research and mostly concentrate on conceptual issues. To provide some indications of statistical convergence, simulation experiments for a selected set of examples are available in the working paper version, in Creal et al. (2008).

### 3.2 Parameterization and identification issues

The GAS specification allows a freedom of choice with respect to the parameterization of the model. In the GARCH example of Section 2.2, the time-varying parameter is $\psi_t = \sigma_t^2$. When it is preferred to enforce the positivity of $\sigma_t^2$, an obvious alternative would be to parameterize the model in terms of $\tilde{\psi}_t = \ln(\sigma_t^2)$. After some manipulations, the GAS(1,1) specification for this alternative model is

$$\tilde{\psi}_t = \omega + A_1 \left( \frac{y_{t-1}^2}{\sigma_{t-1}^2} - 1 \right) + B_1 \tilde{\psi}_{t-1}, \quad (19)$$
which is equivalent to the EGARCH model of Nelson (1991). The GAS dynamics automatically adapt to the choice of parameterization. In general, assume that one prefers a different parameterization \( \tilde{\psi}_t = h(\psi_t) \) for some invertible mapping \( h(\cdot) \). Let \( \dot{h}_t = \partial h(\psi_t)/\partial \psi_t \) and note that \( \dot{h}_t \) is deterministic given all information up to and including time \( t \). Let \( \dot{s}_t^\psi \) denote the GAS information scaled score step for parameterization \( \psi_t \). For well behaved densities, the information matrix equals both the expected outer product of scores and the expected second derivative of the log density. This allows the information scaled score step to be written as

\[
s_t^\psi = \left( E_{t-1}[(\dot{h}_{t-1})'\nabla_t \dot{h}_{t-1}^{-1}] \right)^{-1} (\dot{h}_{t-1})'\nabla_t = \dot{h}_{t-1}s_t^\psi.
\]

Reparameterizing the model thus reduces to rescaling the information scaled score step by the inverse gradient of the mapping \( h(\cdot) \) in each period. For example, this confirms the transition from (9) to (10) by defining \( \tilde{\psi}_t = -\log(\psi_t) \) in Example 2.

Another important issue concerns parameter identification. Consider model density

\[
p(y_t; \Phi \psi_t) = p(y_t; \Phi \psi_t, \Psi_{t-1}, Y_{t-1}, X_{t-1}, \theta),
\]

where the time-varying factor \( \psi_t \) follows a GAS(1,1) specification and \( \Phi \) is a loading matrix, see the discussion below (13) in Example 4. For example, \( \Phi \psi_t \) can be a vector of volatilities for a vector time series driven by a single common factor \( \psi_t \). In this case, it is not possible to define both \( \Phi \) and all GAS(1,1) parameters \( \omega, A_1 \) and \( B_1 \), simultaneously. Take the model in (21) and introduce an invertible matrix \( K \). Define \( \tilde{\psi}_t = K \psi_t, \tilde{s}_t = K s_t, \tilde{\Phi} = K \Phi^{-1}, \tilde{\omega} = K \omega, \tilde{A}_1 = KA_1K^{-1} \) and \( \tilde{B}_1 = KB_1K^{-1} \). The likelihoods for \( p(y_t; \Phi \psi_t) \) and \( p(y_t; \Phi \tilde{\psi}_t) \) are obviously identical. Pre-multiplying the GAS(1,1) transition equation for the original parameterization by \( K \), we obtain

\[
K \psi_t = K \omega + K A_1 K^{-1} K s_{t-1} + K B_1 K^{-1} K \psi_{t-1} \quad \Leftrightarrow \quad \tilde{\psi}_t = \tilde{\omega} + \tilde{A}_1 \tilde{s}_{t-1} + \tilde{B}_1 \tilde{\psi}_{t-1}.
\]

From (20) it follows directly that \( \tilde{s}_t \) is the GAS driver for the new parameterization \( \tilde{\psi}_t \). As \( K \) is an arbitrary invertible matrix, restrictions must be imposed on \( \Phi \) to ensure identification. For example, specific rows of \( \Phi \) can be set equal to corresponding rows of the identity matrix. Note, however, that the identification problem cannot be solved by only imposing restrictions
on the matrix $A_1$ in the GAS(1, 1) model. For example, the normalization condition $A_1 = I$ for $\psi_t$ also holds for $\dot{\psi}_t = K\psi_t$ as $\dot{A}_1 = KA_1K^{-1} = KK^{-1} = I$, but for example $\Phi \neq \bar{\Phi} = \Phi K^{-1}$ in general. The fact that the identification issue can be solved via restrictions on $\Phi$ but not on $A_1$ is a direct consequence of the equivariance of the score and information matrix as a driving mechanism for the recursions in the GAS model. We therefore need to take care in normalizing the parameter spaces of GAS models with a factor structure such as (21).

4 Applications and new models

In this section we present illustrations to highlight the variety of cases in which the GAS framework can be used. We provide several new models and non-trivial extensions of existing models. Additional non-trivial applications are reported in Creal et al. (2008).

4.1 Time-varying non-linear regression models

The term structure of interest rates plays a central role in both macroeconomics and finance as it describes the linkage between monetary policymakers’ impact on the short term interest rate and firms’ investment decisions at longer horizons. We develop an observation driven analogue to the popular term structure model of Nelson and Siegel (1987) given by the partial non-linear regression model

$$y_{t, \tau} = x_\tau(\lambda) \beta + \varepsilon_{t, \tau}, \quad t = 1, \ldots, n,$$

where $y_{t, \tau}$ is the interest rate at time $t$ for an investment that matures after $\tau$ months. The $1 \times 3$ covariate vector $x_\tau(\lambda)$ is defined as

$$x_\tau(\lambda) = [1, (\lambda \tau)^{-1}(1 - \exp(-\lambda \tau)), (\lambda \tau)^{-1}(1 - \exp(-\lambda \tau)) - \exp(-\lambda \tau)],$$

where the coefficients $\lambda$ and $\beta$ are unknown and with independent disturbance $\varepsilon_{t, \tau} \sim N(0, \sigma^2)$ for a given time $t$. For the particular choice of $x_\tau(\lambda)$, the three coefficients in $\beta$ can be interpreted as the level, slope and curvature of the term structure, respectively. The slope and curvature factors depend on the parameter $\lambda$ that is non-linear in $y_{t, \tau}$. At time $t$, interest rates for $m$ maturities can be observed such that $\tau = \tau_j$ for $j = 1, \ldots, m$. Based on these $m$ observations
(for a given $t$), parameters $\beta$, $\lambda$ and $\sigma^2$ can be estimated by non-linear least squares. This estimation procedure can be repeated at each time $t$, resulting in a time series of parameter estimates, see Diebold and Li (2006) for further details.

**Time-varying dynamic Nelson-Siegel model**

Diebold, Rudebusch, and Aruoba (2006) analyze the time series dimension simultaneously with the maturity dimension of interest rates by considering the Nelson-Siegel model as a multivariate state space model. For this purpose, they treat $\beta$ as a $3 \times 1$ dynamic latent vector $\beta_t$ that is modeled by a vector autoregressive (VAR) process. Furthermore, the interest rates for the $m$ maturities at time $t$ are put into the observation vector $y_t = (y_{t,\tau_1}, \ldots, y_{t,\tau_m})'$ and the model specification $y_t = X_t \beta_t + \varepsilon_t$ is considered with $X_t = [x_{\tau_1}(\lambda), \ldots, x_{\tau_m}(\lambda)']'$ and $\varepsilon_t \sim N(0, D)$ where $D$ is a $m \times m$ positive diagonal matrix. The parameters in the VAR model for $\beta_t$, as well as $D$ and $\lambda$ are estimated by ML using the Kalman filter in this parameter driven approach.

In our observation driven approach, we take $\psi_t = \beta_t$ with the GAS(1, 1) updating equation

$$\psi_t = \omega + A_1 s_{t-1} + B_1 \psi_{t-1}$$

and the driving mechanism given by

$$s_t = (X_t' D^{-1} X_t)^{-1} X_t' D^{-1} (y_t - X_t \psi_t),$$

assuming that $m > 3$ so that matrix $X_t' D^{-1} X_t$ is nonsingular. We can also extend the DNS model as in Koopman, Mallee, and van der Wel (2009) by allowing $\lambda$ to vary over time. Let $\tilde{\psi}_t = (\beta'_t, \lambda_t)'$ be a $4 \times 1$ vector of factors. We propose to model it as $\tilde{\psi}_t = \phi_0 + \Phi \psi_t$ where $\psi_t = \beta_t$ remains the $3 \times 1$ vector of factors. To identify the factor structure of the model, we set the upper $3 \times 3$ matrix of the $4 \times 3$ loading matrix $\Phi$ equal to the identity matrix and the upper three elements of the $4 \times 1$ vector $\phi_0$ equal to zero, see the discussion in Section 3.2. In this case, the driving mechanism for the GAS(1, 1) factor $\tilde{\psi}_t$ is derived as

$$\tilde{s}_t = (\tilde{X}_t' D^{-1} \tilde{X}_t)^{-1} \tilde{X}_t' D^{-1} (y_t - X_t \psi_t),$$

where $\tilde{X}_t = [X_t, (\partial X_t/\partial \lambda_t)\psi_t]$.
Illustration using the Fama-Bliss data-set

To illustrate the GAS specification for the dynamic Nelson-Siegel model, we analyze the Fama-Bliss data-set as in Diebold et al. (2006). It consists of monthly U.S. Treasury yields with maturities 3, 6, 9, 12, 15, 18, 21, 24, 30, 36, 48, 60, 72, 84, 96, 108, and 120 months over the period from January 1972 to December 2000. For comparison purposes, we have estimated both the parameter driven (DNS) and observation driven (GAS) models by ML, both with constant $\lambda$. The DNS estimates are close to those reported by Diebold et al. (2006) while the GAS estimates are different than those obtained for the parameter driven DNS model. For example, the estimate of $\lambda$ is 0.0778 in the DNS model while it is 0.0948 in the GAS model. The estimates of other coefficients are also different, which emphasizes that the interpretation of “comparable” coefficients in both models are different. Nevertheless, the time series of $\beta_t$ estimates appear similar for both classes of models when they are plotted against each other.

The estimation results for the model with time-varying $\lambda_t$ are presented in Figure 1. The first three factors are plotted with their empirical proxies from the data. These results are typical and close to those reported in Diebold et al. (2006). The fourth plot displays the estimates of the time-varying GAS parameter $\lambda_t$ when it is plotted with the nonlinear least squares estimates of $\lambda$ from the cross-section of yields in each period. The estimates of $\lambda_t$ from the GAS model vary considerably and roughly track the estimates of $\lambda$ from the cross-section. By allowing $\lambda$ to vary over time, the log-likelihood function substantially increases from $-3861$ to $-3611$ (an increase of 250 points for adding 4 parameters).

4.2 Pooled marked point-process models

Point process models

Models with time-varying intensities have received much attention in the finance and micro-econometric literature. 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. We develop a new and useful modeling framework that is based on the GAS specification.

Recently, a number of promising models with stochastically evolving intensities have been
proposed, see Bauwens and Hautsch (2006), Koopman et al. (2008), Duffie, Eckner, Horel, and Saita (2006), and Koopman, Lucas, and Schwaab (2008). The econometric handling of these parameter driven models is intricate while parameter estimation can be computationally demanding. In particular, likelihood evaluation for these models requires the computation of high-dimensional integrals using importance sampling techniques or Markov chain Monte Carlo algorithms. The use of such simulation-based techniques, however, may obstruct the widespread application of these models in practice. A computationally less-demanding alternative can be based on developing observation driven analogues of these models.

The first step would then be to consider multivariate generalizations of Russell (2001). However, this is not straightforward. Most of the models of Russell (2001) are developed in the
context of high frequency data and in particular for stock trades. The structure of data sets of trades is substantially different from the data sets that are used in credit risk. Whereas in high frequency data one typically observes many spells for a limited number of stocks, in modeling credit data one typically works with many different companies that only have very few spells each. This requires the pooling of data over different companies in the sample. Consequently, different events might carry information that is relevant for the dynamic parameter at any point in time. The GAS model provides a straightforward and consistent methodology to address this issue.

The GAS point process model

Let $y_k(t) = [y_{1k}(t), \ldots, y_{nk}(t)]'$ be a vector of marks of $n$ competing risk processes for firm $k = 1, \ldots, N$. We have $y_{jk}(t) = 1$ if event type $j$ materializes for firm $k$ at time $t$, and zero otherwise. By following the application in Koopman et al. (2008), we model the log intensities of these processes by

$$
\lambda_{jk}(t) = \eta_j + \gamma_j' \psi_t^*,
$$

(22)

where $\eta_j$ is the baseline intensity and $\gamma_j$ is the vector of factor loadings for $\psi_t$ with $t^*$ being the last event time before time $t$. The vector of dynamic factors $\psi_t$ is specified by the GAS(1,1) updating equation (2) with $\omega = 0$. Since $\psi_t$ is not observed directly, we need to impose a sign restriction on $\gamma_j$ to obtain economic interpretations for the factors. The intensities of all firms are driven by the same vector of time-varying systematic factors $\psi_t$. Model (22) nests the model of Russell (2001) when we set the dimension of $\psi_t$ equal to the number of firms $N$. In a credit risk context, we typically have $\text{dim}(\psi_t) << N$. Furthermore, we require parameter restrictions for model identification, see the discussion in Section 3. In the illustration below, it is sufficient to set one of the $\gamma_j$’s equal to unity.

The log-likelihood specification using (22) is

$$
\ell_t = \sum_{j,k} y_{jk}(t) \lambda_{jk}(t) - R_{jk}(t) \cdot (t - t^*) \cdot \exp[\lambda_{jk}(t^*)],
$$

(23)

where $R_{jk}(t)$ is a zero-one variable indicating whether company $k$ is potentially subject to risk.
$j$ at time $t$. Based on the first and second derivative of $\ell_t$, we obtain

$$s_t = \left[ \sum_j \gamma_j \gamma'_j \sum_k w_{jk}(t) \right]^{-1} \sum_j \left( \sum_k y_{jk}(t) - R_{jk}(t) \cdot (t - t^*) \cdot \exp[\lambda_{jk}(t)] \right) \gamma_j,$$  

(24)

where $w_{jk}(t) = R_{jk}(t) \cdot \exp[\lambda_{jk}(t)] / \sum_{j,k} R_{jk}(t) \cdot \exp[\lambda_{jk}(t)] = P[y_{jk}(t) = 1]$ is the probability of the next event being of type $j$ for company $k$. Combining these basic elements into a GAS specification, we have obtained a new observation driven model for credit rating transitions.

**Application to Standard and Poor’s rating data**

As an illustration, we adopt the model described above for the CreditPro 7.0 data set which contains the Standard and Poor’s (S&P) rating histories of all US corporates over the period 1981–2005. We distinguish two complementary credit rating classes: the investment grade (IG) and the sub-investment grade (SIG). Event 1 represents a rating transition from IG to SIG while events 2, 3 and 4 represent IG to default, SIG to IG and SIG to default, respectively. The GAS(1,1) model has a univariate (single) factor $\psi_t$ and the updating equation has the scaled score function (24). The parameters of this model are estimated under the restriction $\gamma_4 = 1$. The estimation results are presented in Table 1. The GAS parameter $B_1$ is estimated close to unity which implies a persistent dynamic process for $\psi_t$. Given the estimates of $\gamma_j$, the downgrades appear to be most sensitive to the common factor $\psi_t$. In particular, the baseline downgrade from investment grade to default is small with an estimate of -7.4 while it is strongly sensitive to the common factor $\psi_t$ with a loading estimate of 1.19. Interestingly, the estimated pattern (not shown) of the systematic intensity factor $\psi_t$ is close to the estimated pattern of the parameter driven model of Koopman et al. (2008). However, in a GAS framework we do not require their computationally intensive methods such as importance sampling methods for parameter and factor estimation.

It is straightforward in our GAS framework to generalize the model from a one-factor model to a three-factor model. In this case, $A_1$ and $B_1$ in the GAS updating equation become $3 \times 3$ matrices. To obtain identification, we set the loading vector $\gamma_j$ equal to the $j$th column of a $4 \times 4$ identity matrix for $j = 1, 2, 3$ while $\gamma_4 = \gamma_3$. This parsimonious specification implies that upgrades and downgrades between IG and SIG have different factors while transitions to default
Table 1: Estimation results for one-factor intensity model

<table>
<thead>
<tr>
<th></th>
<th>IG → SIG</th>
<th>IG → DEF</th>
<th>SIG → IG</th>
<th>SIG → DEF</th>
</tr>
</thead>
<tbody>
<tr>
<td>j</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>η</td>
<td>0.520</td>
<td>0.520</td>
<td>0.520</td>
<td>0.520</td>
</tr>
<tr>
<td></td>
<td>(0.076)</td>
<td>(0.076)</td>
<td>(0.076)</td>
<td>(0.076)</td>
</tr>
<tr>
<td>γ</td>
<td>-3.920</td>
<td>-7.360</td>
<td>-3.360</td>
<td>-3.330</td>
</tr>
<tr>
<td></td>
<td>(0.118)</td>
<td>(0.353)</td>
<td>(0.109)</td>
<td>(0.217)</td>
</tr>
<tr>
<td>A1</td>
<td>0.024</td>
<td>6.415</td>
<td>0.998</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.537)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Estimation results for the parameters in the one-factor GAS(1,1) intensity model (22) in a two-grade system, with \( \gamma_4 = 1 \), with the scaled scoring function (24) and based on the S&P ratings of all US corporates between 1981 and 2005. The estimates are reported with asymptotic standard errors in parentheses below the estimates. Parameter \( B_1 \) is subject to a logistic transformation during estimation and \( \gamma_3 \) is subject to a identifying restriction \( \gamma_3 < 0 \).

Table 1: Estimation results for one-factor intensity model

<table>
<thead>
<tr>
<th></th>
<th>IG → SIG</th>
<th>IG → DEF</th>
<th>SIG → IG</th>
<th>SIG → DEF</th>
</tr>
</thead>
<tbody>
<tr>
<td>j</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>η</td>
<td>-3.920</td>
<td>-7.360</td>
<td>-3.360</td>
<td>-3.330</td>
</tr>
<tr>
<td></td>
<td>(0.118)</td>
<td>(0.353)</td>
<td>(0.109)</td>
<td>(0.217)</td>
</tr>
<tr>
<td>γ</td>
<td>0.520</td>
<td>1.190</td>
<td>-0.470</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>(0.076)</td>
<td>(0.330)</td>
<td>(0.086)</td>
<td>——</td>
</tr>
</tbody>
</table>

To conclude, note that the model can be reparameterized as the time-varying multinomial ACM model of Russell and Engle (2005). For the case with information matrix scaling, we illustrate the extension by considering a setting of two competing risks characterized by the log intensities \( \psi_{1t} = \ln(\lambda_{1t}) \) and \( \psi_{2t} = \ln(\lambda_{2t}) \). The multinomial model by contrast uses the log intensity of the pooled process, \( \tilde{\psi}_{1t} = \ln(\lambda_{1t} + \lambda_{2t}) \), and the logit transform of the probability \( \pi_{1t} = \lambda_{1t}/(\lambda_{1t} + \lambda_{2t}) \) of observing state 1 (conditional on an event occurring), \( \tilde{\psi}_{2t} = \ln(\pi_{1t}/(1 - \pi_{1t})) = \ln(\lambda_{1t}/\lambda_{2t}) \). From (20), it follows that the GAS driver for the new parameterization \((\tilde{\psi}_{1t}, \tilde{\psi}_{2t})\) is obtained by pre-multiplying the GAS driver \( s_t \) in (24) by the matrix

\[
\begin{bmatrix}
\lambda_{1,t-1} / (\lambda_{1,t-1} + \lambda_{2,t-1}) & \lambda_{2,t-1} / (\lambda_{1,t-1} + \lambda_{2,t-1}) \\
1 & -1
\end{bmatrix}
\begin{bmatrix}
\pi_{1,t-1} & 1 - \pi_{1,t-1}
\end{bmatrix}.
\]
Figure 2: Marked point-process illustration: the estimated intensities in a two-grade system of the GAS(1,1) model with three credit risk factors, based on the scaled score function (24) and using the S&P rating histories of US corporates for the period 1981–2005.

4.3 Dynamic copula models

Copulas have become popular over the last decade in the literature on financial risk management. A copula is a multivariate distribution function over a hypercube with uniform marginals. The copula can be used to link marginal distributions into a multivariate distribution using Sklar’s theorem. In this section, we demonstrate that the GAS framework provides new model specifications for simple copulas such as the bivariate Gaussian copula. We then illustrate some of the numerical extensions of the GAS specification to mixture copulas that allow for asymmetric tail behavior.
Gaussian copulas

We first focus on 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). In the case of Patton (2006), the driving mechanism for the dynamic bivariate Gaussian copula is given by

\[
\psi_t = \omega + A_1 \cdot \sum_{i=1}^m \Phi^{-1}(u_{1,t-i})\Phi^{-1}(u_{2,t-i}) + B_1\psi_{t-1},
\]

where \(\Phi^{-1}\) is the inverse 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 \(\psi_t\). The (Gaussian) correlation parameter \(\rho_t\) is obtained via the transformation \(\rho_t = [1 - \exp(-\psi_t)] / [1 + \exp(-\psi_t)]\). Equation (25) 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. Our results therefore also apply to the Dynamic Conditional Correlation (DCC) framework of Engle (2002a).

Define \(x_t = \Phi^{-1}(u_{1t})^2 + \Phi^{-1}(u_{2t})^2\) and \(y_t = \Phi^{-1}(u_{1t})\Phi^{-1}(u_{2t})\). For \(m = 1\), Patton’s model (25) reduces to

\[
\psi_t = \omega + A_1 \cdot y_{t-1} + B_1 \cdot \psi_{t-1}.
\]

(26)

Deriving the score and information matrix of the bivariate normal for the transformed correlation parameter, the GAS(1,1) updating equation for \(\psi_t\) is obtained as

\[
\psi_t = \omega + A_1 \cdot \frac{2}{1 - \rho_{t-1}^2} [y_{t-1} - \rho_{t-1} - \rho_{t-1}(1 + \rho_{t-1}^2)^{-1}(x_{t-1} - 2)] + B_1 \psi_{t-1}.
\]

(27)

The similarities and differences between (26) and (27) are clear. Both models are driven by \(y_{t-1}\) as positively clustered transformed marginals should increase the correlation parameter. The additional scaling factor \(1/2(1 - \rho_{t-1}^2)\) in (27) is a consequence of modeling the transformed
correlation parameter $\psi_t$ rather than $\rho_t$ directly. The additional $\rho_{t-1}$ term in the numerator of the second term in (27) enforces $y_{t-1} - \rho_{t-1}$ to be a martingale difference. The most interesting difference between the two model specification is the final term involving $x_{t-1}$. The term $x_{t-1} - 2$ is a martingale difference. The value of $x_t$ is large when an extreme observation occurs in $u_{1t}$, $u_{2t}$ or particularly in both. The effect of such an event depends on the current estimate of the correlation parameter $\rho_t$. If the correlation is positive, the impact on the value of $x_t$ is negative. In this case, the $x_t$ term offsets part of the effect of $y_t$ if the latter has a positive value (for example, if $y_t$ corresponds with the current positive estimate of $\rho_t$). If $y_t$ has a negative value, however, the $x_t$ term reinforces the magnitude of the GAS step.

The effects are visualized in Figure 3 where the GAS (top graphs) and Patton (bottom graphs) drivers for different values of $(u_{1t}, u_{2t})$ and three different values of the correlation parameter, $\rho_{t-1} = -0.5, 0.2, 0.9$, are presented. Note that each pair of top and bottom graphs has the same scale on the vertical axis. If we consider the plot for $\rho_{t-1} = 0.9$, we see two clear differences. First, the GAS step results in a smaller increase in the correlation parameter along the $u_{1t} = u_{2t}$ axis. Particularly if $u_{1t}$ and $u_{2t}$ are both large or small, the step based on $y_t$ alone (Patton; lower panels) results in a more pronounced increase of the transformed correlation parameter. The same holds for the smaller positive correlation parameter of $\rho = 0.2$.

A more striking feature, however, is the increased sensitivity along the off-diagonal areas for positive $\rho$. If the current estimate of $\rho$ is positive and one observes a combination of $(u_{1t}, u_{2t})$ that signals negative rather than positive dependence, the GAS specification is more sensitive to this occurrence and is more inclined to rapidly adjust the current estimate of $\rho$ downwards compared to the Patton step. For negative values of $\rho$, the left panels show that the effects are reversed. The GAS specification becomes more sensitive to observations along the diagonal than the specification based on $y_t$ alone.

**Illustration for Gaussian copula**

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, 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 autoregressive model for the conditional mean and a GARCH model for the conditional
Figure 3: A bivariate Gaussian copula illustration: comparisons between the GAS and Patton drivers as a function of the uniforms $(u_{1t}, u_{2t})$. The top panels contain the graphs for the GAS step in (27) for $\rho_{t-1} = -0.5, 0.2, 0.9$ (left, middle, right). The lower graphs contain the (re-centered) steps $y_t - \rho_t$ of the Patton model, (26). The vertical axes have the same scale for each column of graphs.

variance (an AR-GARCH model). We construct the transformed series for $u_{1t}$ and $u_{2t}$ and use these as input for the Gaussian copula model. Apart from (26) and (27), we also estimate an ad-hoc implementation of the DCC framework of Engle (2002a). In particular, we model the correlation parameter directly using the updating equation

$$\rho_t = \omega + A_1 \cdot y_{t-1} + B_1 \cdot \rho_{t-1}. \tag{28}$$

To enforce the stationarity property of this process, we estimate the logit transform of $B_1$. The results are presented in Table 2 and Figure 4.

Table 2 shows that the GAS specification increases the log-likelihood value 25 to 125 points.
Table 2: Estimation results for different dynamic copula models

<table>
<thead>
<tr>
<th></th>
<th>$10^3 \omega$</th>
<th>$A_1$</th>
<th>$\ln(B_1 / 1 - B_1)$</th>
<th>$B_1$</th>
<th>log-lik</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>German Mark (Euro)–US $$, Japanese Yen–US $$</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GAS</td>
<td>6.11</td>
<td>0.058</td>
<td>5.30</td>
<td>0.995</td>
<td>1218.16</td>
</tr>
<tr>
<td></td>
<td>(2.48)</td>
<td>(0.009)</td>
<td>(0.37)</td>
<td>(0.990,0.998)</td>
<td></td>
</tr>
<tr>
<td>Patton</td>
<td>-1.60</td>
<td>0.036</td>
<td>4.27</td>
<td>0.986</td>
<td>1191.51</td>
</tr>
<tr>
<td></td>
<td>(0.85)</td>
<td>(0.003)</td>
<td>(0.10)</td>
<td>(0.983,0.989)</td>
<td></td>
</tr>
<tr>
<td>DCC</td>
<td>1.03</td>
<td>0.008</td>
<td>4.65</td>
<td>0.991</td>
<td>1184.13</td>
</tr>
<tr>
<td></td>
<td>(0.29)</td>
<td>(0.001)</td>
<td>(0.09)</td>
<td>(0.989,0.992)</td>
<td></td>
</tr>
<tr>
<td><strong>German Mark (Euro)–US $$, British Pound–US $$</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GAS</td>
<td>12.55</td>
<td>0.082</td>
<td>4.97</td>
<td>0.993</td>
<td>2218.82</td>
</tr>
<tr>
<td></td>
<td>(3.55)</td>
<td>(0.008)</td>
<td>(0.26)</td>
<td>(0.988,0.996)</td>
<td></td>
</tr>
<tr>
<td>Patton</td>
<td>-0.97</td>
<td>0.025</td>
<td>4.71</td>
<td>0.991</td>
<td>2090.42</td>
</tr>
<tr>
<td></td>
<td>(0.84)</td>
<td>(0.002)</td>
<td>(0.11)</td>
<td>(0.989,0.993)</td>
<td></td>
</tr>
<tr>
<td>DCC</td>
<td>2.64</td>
<td>0.004</td>
<td>4.84</td>
<td>0.992</td>
<td>2060.43</td>
</tr>
<tr>
<td></td>
<td>(0.39)</td>
<td>(0.000)</td>
<td>(0.11)</td>
<td>(0.990,0.994)</td>
<td></td>
</tr>
</tbody>
</table>

Parameter estimates for the GAS, Patton, and DCC drivers in (26)–(28). 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. Confidence interval in parentheses for $B_1$, otherwise standard errors in parentheses.

Figure 4: A copula illustration: comparisons of the correlation parameter estimates for the GAS, Patton, and DCC drivers in (26)–(28). 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.
for the same number of parameters. The figures show the empirical estimates of the time-varying correlation. Based on the estimates of the parameter $B_1$, the GAS specification leads to the most persistent correlation process, followed by the DCC and the Patton specifications. However, the increased sensitivity of the score mechanism to correlation shocks reveals an opposite pattern in the figures. Due to the sharper decline at the edges as visualized in Figure 3, the GAS specification reacts much more fiercely to exchange rate returns of opposite sign if the current correlation estimate is positive. This is most clearly seen for the Mark-Pound example, but also the Mark-Yen example shows similar features at the end of 1993 and 2003. The DCC dynamics, and to a lesser extent the Patton dynamics, are much smoother in this sense. The difference between the dynamics for the different specifications may be highly relevant for risk managers, where changes in correlations and in particular correlation breakdowns are a major concern.

**Clayton copula**

The GAS specification can also be considered for non-Gaussian copulas such as a mixture of Clayton-type copulas. Patton (2006) proposes a generally applicable driving mechanism for copula parameters

$$
\psi_t = \omega - m^{-1} A_1 \cdot \sum_{i=1}^{m} |u_{1,t-i} - u_{2,t-i}| + B_1 \cdot \psi_{t-1},
$$

(29)

where $\psi_t$ captures the dependence between the coordinates. The intuition for (29) is clear. If the most recent $u_{1t}$ and $u_{2t}$ are close together, this is a signal of strong dependence and, therefore, $\psi_t$ is increased. Similarly $\psi_t$ is decreased if $u_{1t}$ and $u_{2t}$ are far apart.

Though the driving mechanism in (29) is intuitively straightforward, two issues are less clear. First, (29) uses no information contained in the particular choice of the copula. As with the Gaussian copula, such information may be helpful in specifying the dynamics. Second, although (29) provides an easy updating scheme for the bivariate case, the extension to the multivariate case is less obvious. In particular, if one has an Archimedean copula characterized by a single dependence parameter, there are many different ways in which one could use the differences $|u_{it} - u_{jt}|$ for $i \neq j$ to update the dependence parameter. Equation (29) provides little guidance as to how these different and possibly conflicting signals should be weighed.
The Clayton copula for our example is a member of the Archimedian family. Its specification in dimension $d$ is given by

$$C(u_1, \ldots, u_d; \lambda) = \left( 1 - d + \sum_{i=1}^{d} u_{it}^{-\lambda} \right)^{-1/\lambda}. \quad (30)$$

The Clayton copula is characterized by the dependence parameter $\lambda$. Low values of $\lambda$ indicate high levels of dependence. This is also captured by the tail dependence coefficient, which measures the probability of joint extreme exceedances. For the Clayton, extreme joint crashes receive positive probability, while joint extreme upward shocks have zero probability.

We define $g_t(\lambda) = \sum_{i=1}^{d} u_{it}^{-\lambda}$ and consider a time-varying $\lambda$ by replacing it with a GAS parameter $\psi_t$. The Clayton copula has density

$$c(u_1, \ldots, u_d) = (1 - d + g_t(\psi_t))^{-1/\psi_t - d} \cdot \prod_{i=1}^{d} \left( 1 - i \cdot \psi_t u_{it}^{-\psi_t - 1} \right). \quad (31)$$

We obtain the score

$$\nabla_t = - \sum_{i=1}^{d} \left( \frac{i}{1 - i \cdot \psi_t} - \ln(u_{it}) \right) + \frac{1}{\psi_t^2} \ln(1 - d + g_t(\psi_t)) + \left( \frac{1}{\psi_t} + d \right) \sum_{i=1}^{d} u_{it}^{-\psi_t - 1} \ln(u_{it}) \frac{1}{1 - d + g_t(\psi_t)}. \quad (32)$$

The principal difficulty for some GAS-based dynamic copula models is deriving a closed-form expression for the information matrix. Even for simple copula models, this may quickly become unmanageable analytically. This certainly holds for mixtures of copulas that we consider next. To solve this analytical issue, we compute the information matrix numerically. In our current example, the information matrix can be written as

$$I_{t|t-1} = E_{t-1} \left[ (\nabla_t)^2 \right] \equiv h(\psi_t), \quad (33)$$

with the score vector $\nabla_t$ as defined in (32). Note that the function $h(\cdot)$ in (33) does not depend on time or on any parameter other than $\psi_t$. We can therefore construct a grid of values $\psi^{(0)} < \ldots < \psi^{(n)}$ and compute the function value $h(\psi^{(j)})$ at each of the grid points. 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. First, choose starting values of the parameter $\theta$ and set the starting value $\psi_0$. Using interpolation, compute $h(\psi_0)$ and use it to scale the score step $s_1 = \nabla_1 / h(\psi_0)$. Compute the new parameter value $\psi_1$ through the GAS recursion, and again use interpolation to obtain $h(\psi_1)$. This process is repeated for the complete sample $t = 1, \ldots, n$. Finally, the likelihood can be computed.

**Symmetrized Clayton copula**

The Clayton copula accounts for lower tail dependence but not for upper tail dependence. Therefore, it is useful to use a symmetrized version of the Clayton copula that allows for non-zero, but different upper and lower tail dependence. The symmetrized Clayton copula is a mixture of the Clayton and the survival Clayton copula. Consider a general mixture of $r$ copulas,

$$C(u_1t, \ldots, u_dt) = \sum_{i=1}^{r} p_i C_i(u_1t, \ldots, u_dt), \quad (34)$$

with copula functions $C_i$, corresponding density $c_i$ and positive scalar $p_i$ for $i = 1, \ldots, r$. Define $w_i = p_i c_i / \sum_{j=1}^{r} p_j c_j$ as the weight of copula $i$. It is straightforward to derive that

$$\frac{\partial \ln c}{\partial \psi} = \sum_{i=1}^{r} w_i \cdot \frac{\partial \ln c_i}{\partial \psi}, \quad (35)$$

and

$$\frac{\partial^2 \ln c}{\partial \psi \partial \psi'} = \sum_{i=1}^{r} w_i \cdot \left( \frac{\partial^2 \ln c_i}{\partial \psi \partial \psi'} + \frac{\partial \ln c_i}{\partial \psi} \frac{\partial \ln c_i}{\partial \psi'} \right) - \left( \sum_{i=1}^{r} w_i \cdot \frac{\partial \ln c_i}{\partial \psi} \right) \left( \sum_{i=1}^{r} w_i \cdot \frac{\partial \ln c_i}{\partial \psi} \right)', \quad (36)$$

where $\psi$ represents a potential GAS parameter of the symmetrized Clayton copula. We obtain

$$E_{t-1} \left[ \frac{\partial^2 \ln c}{\partial \psi \partial \psi'} \right] = -E_{t-1} \left[ \left( \sum_{i=1}^{r} w_i \cdot \frac{\partial \ln c_i}{\partial \psi} \right) \left( \sum_{i=1}^{r} w_i \cdot \frac{\partial \ln c_i}{\partial \psi} \right)' \right].$$

It follows that the scores of the individual copulas can be used directly to build a GAS driving mechanism of the mixture copula. We illustrate this for a mixture of $r = 2$ copulas. The
first one is the Clayton copula characterized by the parameter $\lambda_L$ that accounts for lower tail dependence. The second component of the mixture is the survival Clayton copula, characterized by the parameter $\lambda_U$ and accounting 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 either $\lambda_U$ or $\lambda_L$ (to either the upper or lower tail dependence, respectively). The contributions are measured in terms of the likelihood of each mixture component vis-a-vis the total likelihood. As a result, observations that cluster in the upper tail automatically contribute to the evolution of $\lambda_U$, and similarly in the lower tail for $\lambda_L$. By contrast, Patton’s methodology 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 these two, we construct a simulated example. We generate data from the symmetrized Clayton copula. The lower tail dependence coefficient follows a sinusoidal pattern. The pattern of the upper tail dependence is also specified by a sinusoidal function, but with a period that is half as long. This makes it difficult for a model with a uniform observation driving mechanism to capture both upper and lower tail dependence dynamics within a single model. We plot the results in Figure 5 for smoothing parameter values $m = 1$ and $m = 10$.

It is clear that the driving mechanism based on averages of $|u_{it} - u_{jt}|$ only captures some of the variation in the dependence coefficients. However, as the same mechanism underlies both types of dependence, it has difficulty in capturing the upper and lower tail dependence dynamics simultaneously. The GAS specification on the other hand is more successful in picking up both types of dynamics. The GAS(1, 1) estimate is noisier compared to the one obtained from the Patton model, but it follows the true dependence pattern more closely. As a result, a significant increase in the likelihood is achieved.

### 4.4 Time-varying higher order moments

Following the empirical successes in GARCH modeling, many authors have suggested further generalizations, in particular to the model with Student $t$ errors. Hansen (1994) proposes to allow the degrees of freedom parameter to vary over time. Harvey and Siddique (1999), Jondeau and Rockinger (2003) and Brooks, Burke, Heravi, and Persand (2005) consider models with time-varying skewness and kurtosis. We develop a t-GAS(1, 1) model for $y_t = \sigma_t \varepsilon_t$ where $\varepsilon_t \sim t(\nu_t)$. The error term is scaled to have unit variance such that $\sigma_t^2$ is the conditional variance while $\nu_t$ is the time-varying degrees of freedom parameter. Define the vector of factors as $\psi_t = (\sigma_t^2, - \ln \left\{ \frac{(b - a)}{(\nu_t - a)} \right\})'$ where the second element is the inverse of the logit transformation which is used to keep $\nu_t$ in the interval $[a, b]$ for $a < b$. In our empirical work, we select $a = 2.01$ and $b = 30$ to ensure that the conditional variance exists, that is $\nu_t > 2$. It is possible to select the conditional kurtosis as a factor instead of $\nu_t$ but for some time series the conditional kurtosis may not exist.

To obtain the GAS driving mechanism, we require the score vector $\nabla_t$ for $\sigma_t^2$ and $\nu_t$ which
\[ \nabla_{1t} = -\frac{1}{2\sigma_t^2} + \left( \frac{\nu_t + 1}{2} \right) \left( 1 + \frac{y_t^2}{(\nu_t - 2)\sigma_t^2} \right)^{-1} \frac{y_t^2}{(\nu_t - 2)\sigma_t^2}, \]
\[ \nabla_{2t} = \frac{1}{2} \left\{ \Gamma'(\frac{\nu_t + 1}{2}) - \Gamma'(\frac{\nu_t}{2}) \right\} - \frac{1}{2\nu_t} - \frac{1}{2} \ln \left( 1 + \frac{y_t^2}{(\nu_t - 2)\sigma_t^2} \right) + \frac{\nu_t + 1}{2} \left( 1 + \frac{y_t^2}{(\nu_t - 2)\sigma_t^2} \right)^{-1} \frac{y_t^2}{(\nu_t - 2)^2 \sigma_t^2}, \]
and the elements of the information matrix as given by
\[ I_{11,t|t-1} = -\frac{\nu_t}{2\sigma_t^4 (\nu_t + 3)}, \]
\[ I_{12,t|t-1} = I_{21,t|t-1} = -\frac{3}{2\sigma_t^4 (\nu_t + 1)(\nu_t + 3)(\nu_t - 2)}, \]
\[ I_{22,t|t-1} = \frac{1}{4} \left\{ \Gamma'' \left( \frac{\nu_t + 1}{2} \right) - \Gamma'' \left( \frac{\nu_t}{2} \right) \right\} + \frac{\nu_t + 4)(\nu_t - 3)}{2(\nu_t - 2)^2 (\nu_t + 1)(\nu_t + 3)}. \]

The functions \( \Gamma' \) and \( \Gamma'' \) are the digamma and trigamma functions and \( I_{ij,t|t-1} \) is the \((i,j)\) element of \( I_{t|t-1} \). Given the results above and the derivatives of the logit transformation, it is straightforward to construct a GAS(1,1) recursion using the reparameterization argument from (20). We label this model the tv-t-GAS(1,1) model.

We consider daily returns on the S&P 500 from February 1989 through April 2008 as an illustration. We compare the tv-t-GAS(1,1) model described above to a t-GAS(1,1) model with constant \( \nu \), that is equation (8), and a standard t-GARCH(1,1) model with constant \( \nu \) as in Bollerslev (1987). We estimate the t-GARCH(1,1) model using the recursion (6) instead of (7) to make the model’s parameters comparable. Parameter estimates from each of these models are reported in Table 3 and estimates of the conditional variance are plotted in panel (i) of Figure 6. The log-likelihood values for the t-GAS(1,1) and t-GARCH(1,1) models are close to each other. Estimates of the conditional variances in panel (i) of Figure 6 are similar except for those periods with outlying observations. The differences between the estimates from the two GAS models minus those from the GARCH model are presented in panel (ii) of Figure 6. In the first half of the sample before 1998, the level of volatility is lower and there are several outliers in the series. The estimated conditional variance from the t-GARCH(1,1) model is larger compared to those from both GAS models. The difference in estimated degrees
Figure 6: Time-varying degrees of freedom illustration: (i) estimated conditional variances from the t-GAS(1,1), t-GARCH(1,1), and tv-t-GAS(1,1) models; (ii) differences between the two GAS(1,1) models and the t-GARCH(1,1) model; (iii) estimated time-varying degrees of freedom from the tv-t-GAS(1,1) model; (iv) estimated time-varying degrees of freedom from the GARCH model of Brooks et al. (2005).

of freedom is due to the fact that the t-GAS model does not treat outliers like a standard t-GARCH model. From 1998-2003, volatility increases such that large returns are not necessarily treated as outliers. Estimates of the conditional variance from the GAS and GARCH models are still significantly different and economically meaningful during this period.

The estimated time-varying degrees of freedom from the tv-t-GAS(1,1) model is plotted in panel (iii) of Figure 6 and these estimates demonstrate significant variability. The log-likelihood value for our new time-varying GAS model increases appreciably relative to the t-GAS(1,1) model. Estimates of the conditional variance in panels (i) and (ii) are reasonably similar to the t-GAS(1,1) model with some differences in 1998-2004 when the time-varying degrees of freedom increases. We compare this model with the time-varying higher-order GARCH model of Brooks et al. (2005), which we label as the tv-t-GARCH(1,1) model. In their model, the
Table 3: Estimation results for different time-varying higher-order moment models

<table>
<thead>
<tr>
<th></th>
<th>tv-t-GAS</th>
<th>t-GARCH</th>
<th>t-GAS</th>
<th>tv-t-GARCH</th>
<th>t-GAS</th>
<th>t-GAS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Full sample</td>
<td>Sample I</td>
<td>Sample II</td>
<td>Sample I</td>
<td>Sample II</td>
<td>Sample II</td>
</tr>
<tr>
<td><strong>ω1</strong></td>
<td>0.007 (0.007)</td>
<td>0.004 (0.001)</td>
<td>0.004 (0.001)</td>
<td>0.004 (0.001)</td>
<td>0.002 (0.001)</td>
<td>0.007 (0.003)</td>
</tr>
<tr>
<td><strong>ω2</strong></td>
<td>-2.385 (0.738)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>A1(1,1)</strong></td>
<td>0.057 (0.007)</td>
<td>0.046 (0.007)</td>
<td>0.044 (0.006)</td>
<td>0.048 (0.007)</td>
<td>0.026 (0.006)</td>
<td>0.060 (0.009)</td>
</tr>
<tr>
<td><strong>A1(1,2)</strong></td>
<td>-0.131 (0.045)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>A1(2,1)</strong></td>
<td>-0.220 (0.040)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>A1(2,2)</strong></td>
<td>-1.375 (0.921)</td>
<td>-</td>
<td>-</td>
<td>0.004 (0.005)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>B1(1,1)</strong></td>
<td>0.994 (0.004)</td>
<td>0.996 (0.002)</td>
<td>0.996 (0.002)</td>
<td>0.997 (0.002)</td>
<td>0.997 (0.002)</td>
<td>0.996 (0.003)</td>
</tr>
<tr>
<td><strong>B1(1,2)</strong></td>
<td>0.000 (0.000)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>B1(2,1)</strong></td>
<td>0.982 (0.135)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>B1(2,2)</strong></td>
<td>0.026 (0.370)</td>
<td>-</td>
<td>-</td>
<td>0.966 (0.026)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>ν</strong></td>
<td>-</td>
<td>6.920 (0.606)</td>
<td>7.05 (0.680)</td>
<td>-</td>
<td>5.368 (0.610)</td>
<td>11.02 (2.091)</td>
</tr>
<tr>
<td><strong>log-lik</strong></td>
<td>-6138.44</td>
<td>-6155.88</td>
<td>-6159.15</td>
<td>-6154.39</td>
<td>-2359.69</td>
<td>-3780.95</td>
</tr>
</tbody>
</table>

Estimates from t-GARCH(1, 1), t-GAS(1, 1) and tv-t-GAS(1, 1) models applied to daily returns of the S&P500 from Feb. 1989 - April 2008 (full sample), Feb. 1989 - Dec 1998 (sample I) and Jan. 1999 - April 2008 (sample II). The tv-t-GARCH(1, 1) estimates are from Brooks et al. (2005). We report estimates for $ω = (ω_1, ω_2)'$, the $(i,j)$ elements of the matrices $A_1$ and $B_1$, for $i, j = 1, 2$, the degrees of freedom parameter $ν$ and the loglikelihood values.

The second factor is the conditional kurtosis and it evolves independently from $σ^2_t$ according to its own GARCH(1, 1) recursion. The implied estimates of $ν_t$ can be calculated straightforwardly.

We estimate their model using the factor recursion (6) to give their model’s parameters an equivalent interpretation.

It is a notable result that the estimates of $ν_t$ from our model shown in panel (iii) are significantly different than the implied estimates of $ν_t$ from the tv-t-GARCH(1, 1) model of Brooks et al. (2005) in panel (iv). In the literature on time-varying higher-order moments, the factors are typically forced to evolve independently by imposing zero restrictions on the
(1, 2) and (2, 1) elements of $B_1$. The estimated $B_1$ coefficient elements (2, 1) and (2, 2) reported in Table 3 for the GAS model imply that both $\sigma_t^2$ and $\nu_t$ are driven by the same factor since element (2, 2) of $B_1$ is estimated close to zero. Accordingly, the estimates of $\nu_t$ in panel (iii) of Figure 6 exhibit a similar pattern with the conditional variance in panel (i). Estimates of $\nu_t$ from the tv-t-GARCH(1,1) model are different while the estimate of element (2, 2) of $B_1$ is significant and persistent.

To complete the analysis, we split the sample in two halves: before and after 1998. The estimates of the parameters in the t-GAS(1,1) model for the two sub-samples are reported in the right-hand columns of Table 3. The degrees of freedom estimate and its standard error clearly increase in the second half of the sample. Estimates of $\nu$ for the t-GARCH(1,1) model (not reported here) are similar in the two sub-samples. This result may seem counterintuitive initially. However, large returns during this period are no longer extreme outliers because the conditional volatility $\sigma_t^2$ is higher. This provides support for estimates of $\nu_t$ from the tv-t-GAS(1,1) model and some evidence that modeling higher-order moments independently of the conditional variance may be inappropriate.

### 4.5 Dynamic mixtures of models

The GAS specification can provide a mixture framework for probabilities of several competing, possibly, time-varying models. Assume we have a mixture model with $J$ components where each component $j$ or sub-model $j$ has a likelihood contribution $\ell_{jt}$ at time $t$ and for $j = 1, \ldots, J$.

Define the vector of GAS factors as the time-varying mixture probabilities $\pi_{jt}$, which defines a new mixture model

$$\ell_t = \sum_{j=1}^{J} \pi_{jt} \ell_{jt}, \quad \sum_{j=1}^{J} \pi_{jt} = 1. \tag{37}$$

We parameterize the $\pi_{jt}$’s using the logit transformation to ensure that the probabilities remain in the zero-one interval. The GAS factors are

$$\pi_{jt} = \frac{e^{\psi_{jt}}}{1 + \sum_{k=1}^{J-1} e^{\psi_{kt}}} \iff \psi_{jt} = \ln(\pi_{jt}) - \ln \left(1 - \sum_{k=1}^{J-1} \pi_{kt}\right). \tag{38}$$

for $j = 1, \ldots, J - 1$ with the probability of the last component determined by the constraint $\pi_{Jt} = 1 - \sum_{k=1}^{J-1} \pi_{kt}$. Taking the derivative of the log-likelihood with respect to $\psi_{jt}$, we obtain
the elements of the score vector

\[ \nabla_{jt} = \frac{\partial \ell_t}{\partial \psi_{jt}} = \frac{\pi_{jt} \ell_{jt}}{\sum_{k=1}^{J} \pi_{kt} \ell_{kt}} - \pi_{jt}, \]

(39)

for \( j = 1, \ldots, J - 1 \). The interpretation of (39) is intuitive. The probability of model \( j \) is increased if the relative likelihood of model \( j \) is above its expectation \( \pi_{jt} \). Otherwise, it is decreased. The information matrix for this GAS model is not easy to compute analytically. In our empirical example below, we use a mixture of two normal densities \( \phi_j(y_t) \) for \( j = 1, 2 \) implying an information matrix of the form

\[ I_{t|t-1} = E_{t-1}[\nabla_{tt'} \nabla_{tt'}'] = \pi_{1,t}(1 - \pi_{1,t})E_{t-1} \left[ \left( \frac{\phi_1(y_t) - \phi_2(y_t)}{\pi_{1,t} \phi_1(y_t) + (1 - \pi_{1,t}) \phi_2(y_t)} \right)^2 \right], \]

where expectation \( E_{t-1} \) is taken with respect to the conditional mixture density. We use numerical integration to compute the information matrix, which is feasible when the mixture model (37) contains say \( J = 5 \) components or less.

To illustrate the methodology, we consider a time series of quarterly log U.S. real GDP growth rates from 1947Q2 to 2008Q2 obtained from the Federal Reserve Bank of St. Louis. The GAS model is a mixture of two normals with different means \( \mu_i \) for \( i = 1, 2 \) and a common variance \( \sigma^2 \). The probability that the data comes from the normal distribution with low mean indicates a recession. The GAS(1,1) updating equation is considered with information smoothing as in (14) and \( \alpha = 0.05 \). This GAS model provides an observation driven alternative to the hidden Markov model (HMM). We compare our model with a simplified version of the Hamilton (1989) model without autoregressive dynamics, that is

\[ y_t = \mu_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma^2), \quad \mu_t = \begin{cases} \mu_1 & \text{if } S_t = 0 \\ \mu_2 & \text{if } S_t = 1 \end{cases} \]

\[ p_{ij} = P(S_t = j | S_{t-1} = i), \quad i = 0, 1 \quad j = 0, 1. \]

In this model, the latent variable \( S_t \) is a regime-switching variable indicating whether the economy is in a recession or expansion. The comparisons focus on the one-step ahead predicted estimates from the hidden Markov model and the GAS(1, 1) model.
Estimates of the parameters of both models are reported in Table 4. The estimated values for each mean are reasonably close. The recession parameter $\mu_1$ for the HMM model is slightly smaller and negative. Panel (i) of Figure 7 presents the growth rate of log U.S. real GDP along with the estimated conditional mean $\pi_t \mu_1 + (1 - \pi_t) \mu_2$ from the GAS and HMM models. The GAS and HMM estimates nicely follow the changes in the mean of the series. The estimated probabilities of a recession from each model are plotted in panel (ii) of Figure 7. The estimated probabilities from the GAS model reflect the possibility of the model to rapidly adapt to new signals concerning the current behavior of the time series. As a result, we obtain a clear division of regimes (switches) over time as depicted in the graph. By contrast, the one-step ahead predicted probabilities produced by the hidden Markov model do not change as rapidly and are less clear. The GAS model offers a convenient method for forecasting economic downturns.
 Parameter estimates and log-likelihood values from the GAS(1,1) mixture and hidden Markov models applied to U.S. log real gdp growth rates from 1947(2) to 2008(2). Standard errors are in parentheses.

## 5 Conclusions

We have introduced Generalized Autoregressive Score (GAS) models. A 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 making a scaled (local density) score step, the time-varying parameter automatically tries to reduce its one-step ahead prediction error at the current observation with respect to current parameter values. Although it is based on a completely different paradigm, the GAS model provides a powerful and highly competitive alternative to other observation driven models as well as parameter driven models. We have illustrated this extensively by describing a number of non-trivial empirical and simulated examples. Some of these examples are interesting in their own right and provide interesting extensions or alternative specifications for parameter driven models with time-varying parameters, in particular for state space models with stochastically time-varying parameters, for multivariate marked point processes, and for time-varying copula models.

There are many interesting future research directions. The issues of identification, consistency, stationarity, and asymptotic distribution theory require more work than presented here. Due to its generality and applicability for a wide class of models, however, it appears difficult to come up with an uniform set of conditions for stationarity and consistency that is applicable
to all situations of interest. A more promising route may be to formulate conditions for particular sub-sets of models with a GAS specification. The investigation of finite-sample properties of GAS models in more detail is a second direction for further research. Although we have provided a number of interesting empirical and simulated examples, a more systematic study into the statistical properties of parameter estimates for GAS models may be appropriate.

A third direction for future research concerns the development of misspecification tests for GAS models. On the one hand, we require goodness-of-fit tests and model selection criteria for GAS models. Many of such tests and diagnostics are already developed for the class of GARCH models. On the other hand, the GAS model itself might provide a powerful and general basis for dynamic misspecification tests. Test diagnostics for the presence of possible ARCH effects are widely applied in empirical studies. Lagrange multiplier based tests for the possible presence of GAS effects are straightforward extensions of these. Such tests might provide a useful empirical tool for testing for possible time variation in parameters in the context of a general class of non-linear and non-Gaussian models.

A fourth direction of research is the application of the GAS specification to new models. In this paper, we have tried to review a number of interesting directions of new models with time-varying parameters. However, the GAS framework is not restricted to these, and other new and empirically relevant models with time-varying parameters would provide additional support for the usefulness of GAS as an empirical modeling tool.

Long-memory versions of the GAS model would be a fifth direction for a possible research project. However, the long-memory specification for GAS models is not trivial and therefore more theoretical and empirical research is needed. A related issue is that GAS models may be interpreted as discrete time approximation of their parameter driven counterparts. An interesting research project may be to bridge the gap between GAS models and parameter driven models in a similar continuous time limiting sense as obtained by Nelson (1990) who has bridged the gap between GARCH models and stochastic volatility model specifications.

A sixth direction of future research is to provide a systematic comparison of the advantages and disadvantages of parameter driven versus observation driven models in a wider setting than GARCH and ACI. Given numerical advances for non-linear and non-Gaussian state space models, and given the general applicability of the current GAS specification, such comparisons have become feasible.
Developing Bayesian inference procedures for the GAS framework is an interesting seventh possibility. For models with a small number of parameters, we anticipate that it will be relatively straightforward to simulate from the posterior densities using Markov chain Monte Carlo as the likelihood of the model is readily available. Intelligent proposal distributions may need to be designed for models with a large number of parameters. One potential advantage of a Bayesian approach is that parameter uncertainty is easily taken into account when estimating the factors.

Finally, there are various computational details that need to be studied in further detail. Three issues of particular interest are: finding starting values, finding the required degree of information smoothing for the GAS updating step in particular models, and finding better numerical approximations to the scaling matrix if it cannot be computed analytically. With respect to the first issue, our findings so far are mixed. In relatively straightforward models, the problem of finding appropriate starting values does not exist. In particular, if the information matrix is clearly non-singular for all sample observations, the maximum likelihood maximization algorithm converges quickly and robustly. Introducing information smoothing as well as finding reasonable starting values become more relevant when an observation contains limited or no information on the parameter of interest. This is particularly relevant if there are regions with a degenerate information matrix. In our experience, some degree of information smoothing is indispensable in such cases. In addition, automatic smoothing by estimating the smoothing parameter directly from the data has increased the likelihood value in several cases. In our current implementations, however, the information smoothing is rather rigid. One could consider more involved specifications, where the degree of smoothing also depends on the current position in the sample and the parameter space. The third issue concerns further progress that is needed for models where the information matrix cannot be computed analytically. In the illustration of time-varying copulas in Section 4.3, we provided some suggestions based on numerical interpolation techniques using kernel smoothing in low-dimensional parameter spaces. Further extensions are needed to develop computationally feasible estimation methods for GAS models with large parameter spaces and possibly more complicated specifications.
References


