A General Framework for Observation Driven Time-Varying Parameter Models

Koopman, S.J.; Creal, D.D.; Lucas, A.

2008

TI 2008-108/4
Tinbergen Institute Discussion Paper

A General Framework for Observation Driven Time-Varying Parameter Models

Drew Creal[1]
Siem Jan Koopman[1,2]
André Lucas[1,2,3]

[1] *VU University Amsterdam;*
[2] *Tinbergen Institute;*
[3] *Duisenberg School of Finance.*

Tinbergen Institute

The Tinbergen Institute is the institute for economic research of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam, and Vrije Universiteit Amsterdam.

Tinbergen Institute Amsterdam
Roetersstraat 31
1018 WB Amsterdam
The Netherlands
Tel.: +31(0)20 551 3500
Fax: +31(0)20 551 3555

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900
Fax: +31(0)10 408 9031

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

Drew Creal[a], Siem Jan Koopman[a,c], André Lucas[b,c]

[(a)] *Department of Econometrics, VU University Amsterdam*

[(b)] *Department of Finance, VU University Amsterdam, Duisenberg school of finance*

[(c)] *Tinbergen Institute, Amsterdam*

November 5, 2008

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. Examples include non-linear regression models with time-varying parameters, observation driven analogues of unobserved components time series models, multivariate point process models with time-varying parameters and pooling restrictions, new models for time-varying copula functions, and models for time-varying higher order moments. 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.

*Corresponding author: André Lucas, VU University Amsterdam, FEWEB, De Boelelaan 1105, 1081 HV Amsterdam, Netherlands, phone: +31 20 598 6039, fax: +31 20 598 6020, 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, 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 that can be used for the modeling of credit rating transitions. A new class of time-varying copulas models based on the GAS framework can also be formulated. 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 tool to develop models that have time-varying higher-order moments. A particular case is to consider linear regression models and GARCH models with Student t distributions where the degrees of freedom parameter is allowed to be time-varying. Observing trade by trade transaction prices on a discrete grid leads to some interesting research directions as well. For example, price changes can be viewed as realizations of a multinomial distribution which need to be subject to time-varying processes. We will discuss a GAS treatment for the modeling of discrete price changes. Finally, a methodology for dynamic mixture distributions based on the GAS framework can also be developed.

The remainder of the paper is organised 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. In Section 5 we provide simulation evidence for the statistical properties of the estimators. Finally, Section 6 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, 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 $Y_1^t = \{y_1, \ldots, y_t\}$, $F_1^t = \{f_1, \ldots, f_t\}$, and $X_1^t = \{x_1, \ldots, x_t\}$. The available information set at time t consists of Y_1^{t-1}, $F_1^{t-1} = \{f_{t-1}, F_1^{t-2}\}$, and X_1^t. We assume that y_t is generated by the observation density

$$p(y_t|f_{t-1}, Y_1^{t-1}, X_1^t, F_1^{t-2}; \theta), \tag{1}$$

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

$$f_t = \omega + \sum_{i=0}^{p-1} A_i s_{t-i} + \sum_{j=1}^{q} B_j f_{t-j}, \tag{2}$$

where ω is a vector of constants, coefficient matrices A_i and B_j have appropriate dimensions for $i = 0, \ldots, p-1$ and $j = 1, \ldots, q$, while s_t is an appropriately scaled function, which depends on past observations Y_1^{t-1}, the time-varying parameters in F_1^{t-1}, and the static parameter vector θ. Furthermore, all unknown coefficients in (2) are functions of θ 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, the parameter f_t would evolve subject to its own source of error, say η_t. In particular, s_t would be replaced by η_t in (2). Estimation of f_t is then based on the conditional (filtered) density $p(f_t|Y_1^t, X_1^t, F_1^{t-1}; \theta)$. 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 f_{t-1} to f_t via (2) with

$$s_t = S(t, Y_1^{t-1}, F_1^{t-1}) \cdot \nabla_t = S_{t-1} \cdot \nabla_t, \tag{3}$$

where

$$\nabla_t = \partial \ln p(y_t | f_{t-1}, Y_1^{t-1}, X_1^t, F_1^{t-2}; \theta) \; / \; \partial f_{t-1}, \qquad S_{t-1} = S(t, Y_1^{t-1}, X_1^t, F_1^{t-1}; \theta), \tag{4}$$

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-1} equal to the (pseudo)-inverse information matrix based on the density (1), that is

$$S_{t-1} = \mathcal{I}_{t-1}^{-1} = \mathrm{E}_{t-1} \left[\nabla_t \nabla_t' \right]^{-1} = -\mathrm{E}_{t-1} \left[\frac{\partial^2 \ln p(y_t | f_{t-1}, Y_1^{t-1}, X_1^t, F_1^{t-2}; \theta)}{\partial f_{t-1} \partial f_{t-1}'} \right]^{-1}. \tag{5}$$

The updating mechanism (2) for f_t now reduces to something close to a Gauss-Newton updating step for every new observation y_t that becomes available through time. 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 $S_{t-1} = I$ 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. It will be shown that the GAS framework offers many interesting directions for future research.

2.2 Some examples

Example 1 (GARCH model): Consider the basic model $y_t = \sigma_{t-1} \varepsilon_t$ 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 $S_{t-1} = \mathcal{I}_{t-1}^{-1}$ for $f_t = \sigma_t^2$ reduces to

$$f_t = \omega + A_0 \left(y_t^2 - f_{t-1} \right) + B_1 f_{t-1}, \tag{6}$$

which is equivalent to the standard GARCH(1,1) specification of Bollerslev (1986). However, if we assume that ε_t follows a Student t distribution scaled to have variance one and with ν degrees of freedom, that is $\varepsilon_t \sim t_\nu$, the GAS(1,1) specification for the conditional variance leads to the updating equation

$$f_t = \omega + A_0 \left(1 + 3\nu^{-1} \right) \cdot \left(\frac{(1+\nu^{-1})}{(1-2\nu^{-1})(1+\nu^{-1} y_t^2 \ / (1-2\nu^{-1}) \ f_{t-1})} y_t^2 - f_{t-1} \right) + B_1 f_{t-1}. \tag{7}$$

The update (7) collapses to (6) in case of the Gaussian distribution, that is $\nu^{-1} = 0$. The recursion in (7), 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 (7) causes a more moderate increase in the variance for a large realization of $|y_t|$ as long as ν is finite. The intuition is clear: if the errors are modeled by a fat-tailed distribution, a large realization in 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. We return to this example in more detail in Section 4.

Example 2 (MEM, ACD, and ACI models): Consider the model $y_t = \mu_{t-1}\varepsilon_t$ where ε_t has a gamma distribution with density $p(\varepsilon_t|\alpha) = \Gamma(\alpha)^{-1}\varepsilon_t^{\alpha-1}\alpha^\alpha \exp(-\alpha\varepsilon_t)$. Using a change of variables, we have $p(y_t|\alpha, \mu_{t-1}) = \Gamma(\alpha)^{-1} y_t^{\alpha-1}\alpha^\alpha \mu_{t-1}^{-\alpha} \exp(-\alpha \frac{y_t}{\mu_{t-1}})$. Let $f_t = \mu_t$, then the GAS(1, 1) model with $S_{t-1} = \mathcal{I}_{t-1}^{-1}$ is given by

$$f_t = \omega + A_0 \left(y_t - f_{t-1} \right) + B_1 f_{t-1}. \tag{8}$$

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 (8). Suppose we parameterize the exponential density in terms of the intensity rather than the expected duration so that $\lambda_t = 1/\mu_t$ and $p(y_t|\lambda_{t-1}) = \lambda_{t-1} \exp(-\lambda_{t-1} y_t)$. Let $\tilde{f}_t = \log(\lambda_t)$. The GAS(1,1) model now has the updating equation

$$\tilde{f}_t = \omega + A_0 \left(1 - y_t \exp(\tilde{f}_{t-1}) \right) + B_1 \tilde{f}_{t-1}, \tag{9}$$

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

Example 3 (Regression model): The linear regression model $y_t = x_t'\beta_{t-1} + \varepsilon_t$ has a $k \times 1$ vector x_t of exogenous variables, a $k \times 1$ vector of time-varying regression coefficients β_{t-1} and normally distributed disturbances $\varepsilon_t \sim \mathrm{N}(0, \sigma^2)$. Let $f_t = \beta_t$. It follows that the scaled score function based on $S_{t-1} = \mathcal{I}_{t-1}^{-1}$ is given by

$$s_t = (x_t'x_t)^{-1}x_t(y_t - x_t'f_{t-1}), \tag{10}$$

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

$$f_t = \omega + A_0(x_t'x_t)^{-1}x_t(y_t - x_t'f_{t-1}) + B_1 f_{t-1}. \tag{11}$$

In case $x_t \equiv 1$, the updating equation (11) for the time-varying intercept reduces to the exponentially weighted moving average (EWMA) recursion by setting $\omega = 0$ and $B_1 = 1$, that is

$$f_t = f_{t-1} + A_0(y_t - f_{t-1}). \tag{12}$$

In this case, we obtain the observation driven analogue of the local level (parameter driven) model,

$$y_t = \mu_{t-1} + \varepsilon_t, \qquad \mu_t = \mu_{t-1} + \eta_t,$$

where the unobserved level component μ_t is modeled by a random walk process and the disturbances ε_t and η_t are mutually and serially independent, and normally distributed, see Durbin and Koopman (2001, Chapter 2). A direct link between the parameter and observation driven models is established when we set $\eta_t = \alpha(y_t - \mu_{t-1}) = \alpha\varepsilon_t$ while in (12) we set $\alpha \equiv A_0$ and consider f_{t-1} as the (filtered) estimate of μ_{t-1}. The local level model example illustrates that GAS models are closely related to the single source of error (SSOE) framework as advocated by Ord, Koehler, and Snyder (1997). However, the GAS framework allows for straightforward extensions for this class of models. For example, the EWMA scheme in (12) can be extended by including σ^2 as a time-varying factor and recomputing the scaled score function in (10) for the new time-varying parameter vector $f_{t-1} = (\beta_{t-1}' \ , \ \sigma_{t-1}^2)'$.

The GAS updating function (11) reveals that if $x_t'x_t$ is close to zero, the GAS driving mechanism can become unstable. As a remedy for such instabilities, we provide an information smoothed variant of the GAS driving mechanism which we discuss in the next subsection. Alternatively, we may want to consider the identity matrix to scale the score with $S_{t-1} = I$ and $s_t = x_t(y_t - x_t'f_{t-1})$.

Example 4 (Dynamic exponential family models): Consider the exponential family of distributions represented by

$$\exp(\eta(\theta)'T(y_t) - C(\theta) + h(y_t)), \tag{13}$$

with scalar function C and vector function η. Let $\theta = \Phi f_{t-1}$, such that the parameters in θ are time-varying according to a factor structure. It is well-known that

$$\mathrm{E}_{t-1}[\dot{\eta}' T(y_t)] = \dot{C}, \tag{14}$$

and

$$\mathrm{E}_{t-1}[\dot{\eta}' T(y_t) T(y_t)' \dot{\eta}] = \frac{\partial^2 C}{\partial\theta\partial\theta'} + \frac{\partial C}{\partial\theta}\frac{\partial C}{\partial\theta'}.$$

with $\dot{C} = \partial C/\partial\theta$, $\dot{\eta} = \partial\eta/\partial\theta'$, see Lehmann and Casella (1998). The GAS driving mechanism with information matrix scaling is given by

$$s_t = \left(\Phi' \mathcal{I}_{t-1} \Phi\right)^{-1} \Phi'(\dot{\eta}' T(y_t) - \dot{C}),$$

and

$$\mathcal{I}_{t-1} = \frac{\partial^2 C}{\partial\theta\partial\theta'}.$$

This is a general expression for any member of the exponential family. Shephard (1995) and Benjamin et al. (2003) proposed observation-driven models for the subclass of natural exponential family members when $\eta(\theta)' T(y_t) = \theta' y_t$ in (13). Expression (14) then reduces to $\mathrm{E}_{t-1}[y_t] = \partial C/\partial\eta = g(f_{t-1}, Y_1^{t-1}, X_1^t, F_1^{t-2})$ where $g(\cdot)$ is known as the link function. They then model the link function using explanatory variables and autoregressive/moving average terms. The advantage of the GAS model over these alternative specifications is that it exploits the full density structure to update the time-varying parameters.

The main obstacle for using GAS models may be the computation of the information matrix given a specific parameterization. To facilitate this task, we present the elements of the gradient vector and the information matrix for a variety of exponential family models in Table 1. In addition to the GARCH and MEM classes of models, the GAS framework also encompasses the time-varying binomial models of Cox (1958) and Rydberg and Shephard (2003), the ACM model of Russell and Engle (2005), and some of the Poisson models in Davis et al. (2003). The latter three models can be obtained by scaling the relevant score vector from Table 1 with either an identity scaling matrix, $S_{t-1} = I$, or the matrix square root of $S_{t-1} = \mathcal{I}_{t-1}^{-1}$.

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. In 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

Table 1: Details for the GAS updates for a selection of exponential family distributions

Distribution	f_t	∇_t	$\mathcal{I}_t$
Normal (1)	μ_t	$0.5(y_t-\mu_t)/\sigma_t^2$	$\mathcal{I}_{t,11}=0.5\sigma_t^{-2}$
$\frac{\exp(-0.5(y-\mu)^2)}{(2\pi\sigma^2)^{1/2}}$	σ_t^2	$-0.5\sigma_t^{-2}+0.5\sigma_t^{-4}(y_t-\mu_t)^2$	$\mathcal{I}_{t,22}=0.5\sigma_t^{-4}$
			$\mathcal{I}_{t,12}=0$
Normal (2)	μ_t	$0.5(y_t-\mu_t)/\sigma_t^2$	$\mathcal{I}_{t,11}=0.5\sigma_t^{-2}$
$\frac{\exp(-0.5(y-\mu)^2)}{(2\pi\sigma^2)^{1/2}}$	$\ln(\sigma_t^2)$	$-0.5+0.5\sigma_t^{-2}(y_t-\mu_t)^2$	$\mathcal{I}_{t,22}=0.5$
			$\mathcal{I}_{t,12}=0$
Exponential	$\ln(\lambda_t)$	$1-\lambda_t y_t$	$\mathcal{I}_t=1$
$\lambda\exp(-\lambda y)$			
Gamma	$\ln(\alpha_t)$	$\alpha_t\left(\ln(y_t)-\ln(\beta_t)-\Psi(\alpha_t,1)\right)$	$\mathcal{I}_{t,11}=\alpha_t^2\Psi(\alpha_t,2)$
$\frac{y^{\alpha-1}\exp(-y/\beta)}{\beta^\alpha\Gamma(\alpha)}$	$\ln(\beta_t)$	$y_t/\beta_t-\alpha_t$	$\mathcal{I}_{t,22}=\alpha_t$
			$\mathcal{I}_{t,12}=\alpha_t$
Dirichlet	$\ln(\alpha_{it})$	$\alpha_{it}\left(\Psi\left(\sum\alpha_{jt},1\right)-\Psi(\alpha_{it},1)\right)$	$\mathcal{I}_{t,ii}=\alpha_{it}\left[1+\Psi(\alpha_{it},1)+\right.$
		$+\alpha_{it}\ln(y_{it})$	$\alpha_{it}\Psi(\alpha_{it},2)-$
			$\left.\Psi\left(\sum\alpha_{jt},1\right)-\alpha_{it}\Psi\left(\sum\alpha_{jt},2\right)\right]$
			$\mathcal{I}_{t,ij}=\alpha_{it}\alpha_{jt}\Psi\left(\sum\alpha_{jt},2\right)$
Poisson	$\ln(\mu_t)$	$y_t-\mu_t$	$\mathcal{I}=\mu_t$
$\frac{e^{-\mu}\mu^y}{y!}$			
Negative	$\ln(r_t)$	$r_t(\ln(p_t)+\Psi(y_t+r_t,1)-\Psi(r_t,1))$	$\mathcal{I}_{t,11}=r_t^2\left(\Psi(r_t,2)-\right.$
Binomial			$\left.\mathrm{E}[\Psi(r_t+y_t,2)]\right)$
$\binom{y+r-1}{k}p^r(1-p)^y$	$\ln(p_t/(1-p_t))$	$r_t(1-p_t)-y_tp_t$	$\mathcal{I}_{t,22}=r_t(1-p_t)$
			$\mathcal{I}_{t,12}=-p_t$
Multinomial	$\ln\left(\frac{p_{it}}{1-\sum_{j=1}^{J-1}p_{jt}}\right)$	$y_{it}-np_{it}$	$\mathcal{I}_{t,ii}=np_{it}(1-p_{it})$
$\frac{n!\prod_{j=1}^J p_j^{y_j}}{y_1!\cdots y_J!}\cdot$	$j=1,\ldots,J-1$		$\mathcal{I}_{t,ij}=-np_{it}p_{jt}$
$y_J=n-\sum_{j<J}y_j$			
$p_J=1-\sum_{j<J}p_j$			

The GAS model specification is given by the equations (1) *and* (2). *We have defined* ∇_t *in* (4) *and* $\mathcal{I}_t$ *in* (5). *The* (i,j) *element of* $\mathcal{I}_t$ *is denoted by* $\mathcal{I}_{t,ij}$. *We further note that* $\Psi(x,k)=\partial^k\ln\Gamma(x)/\partial x^k$.

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 $\mathrm{E}_{t-1}[s_t]=0$. This follows directly from the properties of the score vector. Due to scaling by the matrix S_{t-1}, we also obtain $\mathrm{E}_{t-1}[s_ts_t']=S_{t-1}\cdot\mathcal{I}_{t-1}\cdot S_{t-1}'$. This simplifies to $\mathcal{I}_{t-1}^{-1}$ and $\mathcal{I}_{t-1}$ for $S_{t-1}=\mathcal{I}_{t-1}^{-1}$ and $S_{t-1}=I$, 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).

An important ingredient of the GAS model is the scaling of the score based on the inverse of the information matrix. A simpler alternative for scaling is the unit matrix, $S_{t-1} \equiv I$. In this case, the updating mechanism only uses the unscaled gradient making it close to a steepest-descent optimization step of the likelihood at time t. Our experience, however, is that this type of updating mechanism is often less stable. We therefore propose to scale the score vector by the inverse of the information matrix whenever possible, $S_{t-1} = \mathcal{I}_{t-1}^{-1}$. A potential difficulty with this scaling is the computation of the inverse when the information matrix is not full rank or numerically unstable for specific models. 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,

$$y_t = \phi_{t-1} y_{t-1} + \varepsilon_t,$$

where ε_t is standard normally distributed. The information scaled score step in this case reduces to $s_t = (y_t - \phi_{t-1} y_{t-1})/y_{t-1}^2$. The GAS updating scheme becomes numerically unstable if y_{t-1} is close to zero. In this case, the information matrix $\mathcal{I}_{t-1} = y_{t-1}^2$ is close to zero and s_t can jump to extreme values. In such cases, we introduce a form of information smoothing over the most recent stretch of observations, that is $S_{t-1} = (\mathcal{I}_{t-1}^c)^{-1}$ where

$$\mathcal{I}_{t-1}^c = \alpha \mathcal{I}_{t-2}^c + (1-\alpha)\mathcal{I}_{t-1}. \tag{15}$$

for some $0 \leq \alpha \leq 1$. This is an EWMA smoothing scheme. Other weighting schemes for smoothing the information are also possible. The smoothing parameter α determines the number of observations that S_{t-1} 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 α as part of the static parameter vector θ 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 f_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

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

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 in our GAS framework.

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

$$\max_{\theta} \sum_{t=1}^{n} \ell(\theta; y_t, f_t, Y_1^{t-1}, X_1^t, F_1^{t-1}), \tag{16}$$

where $\ell(\theta; y_t, f_t, Y_1^{t-1}, X_1^t, F_1^{t-1}) = \ln p(y_t|f_t, Y_1^{t-1}, X_1^t, F_1^{t-1}; \theta)$ 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 $p(y_t|f_t, Y_1^{t-1}, X_1^t, F_1^{t-1}; \theta^*)$ 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 parameters θ. Gradient recursions for the GARCH model have been developed by Fiorentini, Calzolari, and Panattoni (1996). For the GAS(1,1) specification, we obtain

$$\frac{d\ell_t}{d\theta'} = \frac{\partial \ln p_t}{\partial \theta'} + \frac{\partial \ln p_t}{\partial f_{t-1}'} \frac{\partial f_{t-1}}{\partial \theta'}, \tag{17}$$

$$\frac{\partial f_t}{\partial \theta'} = \frac{\partial \omega}{\partial \theta'} + A_0 \frac{\partial s_t}{\partial \theta'} + B_1 \frac{\partial f_{t-1}}{\partial \theta'} + (s_t' \otimes I) \frac{\partial \vec{A}_0}{\partial \theta'} + \left(f_{t-1}' \otimes I\right) \frac{\partial \vec{B}_1}{\partial \theta'}, \tag{18}$$

$$\frac{\partial s_t}{\partial \theta'} = S_{t-1} \frac{\partial \nabla_t}{\partial \theta'} + (\nabla_t' \otimes I) \frac{\partial \vec{S}_{t-1}}{\partial \theta'}, \tag{19}$$

where $\ell_t = \ell(\theta; y_t, f_t, Y_1^{t-1}, X_1^t, F_1^{t-1})$, $p_t = p(y_t|f_t, Y_1^{t-1}, X_1^t, F_1^{t-1}; \theta)$, $\vec{A} = \text{vec}(A)$ 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 f_t. However, computing the analytic derivatives, in particularly for (19), 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 θ gathers all static parameters of the model, we conjecture that under standard regularity conditions, the maximum likelihood estimator $\hat{\theta}$ of θ is consistent and satisfies

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

with $H = -\mathrm{E}[\partial^2 \ell / \partial\theta\partial\theta']$.

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 $|B| < 1$. 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 f_{t-1} is uniformly bounded, standard stability results apply for $|B| < 1$. 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, we complement our empirical examples of Section 4 by Monte Carlo simulation experiments for a selected set of examples in Section 5.

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 Subsection 2.2, the time-varying parameter is $f_t = \sigma_t^2$. When it is preferred to enforce the positivity of σ_t^2, an obvious alternative would be to parameterize the model in terms of $\tilde{f}_t = \ln(\sigma_t^2)$. After some manipulations, the GAS(1,1) specification for this alternative model is

$$\tilde{f}_t = \omega + A_1 \cdot \left(\frac{y_t^2}{\sigma_{t-1}^2} - 1 \right) + B_1 \tilde{f}_{t-1}. \tag{20}$$

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 invertible mapping $h(\cdot)$. Let $\dot{h}_t = \partial h(f_t)/\partial f_t'$ and note that $\dot{h}_t$ is deterministic given all information up to and including time t. Let s_t^f denote the GAS information scaled score step for parameterization 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. This allows the information scaled score step to be written as

$$s_t^{\tilde{f}} = \left(\mathrm{E}_{t-1}[(\dot{h}_{t-1}')^{-1} \nabla_t \nabla_t' \dot{h}_{t-1}^{-1}] \right)^{-1} (\dot{h}_{t-1}')^{-1} \nabla_t = \dot{h}_{t-1} s_t^f. \tag{21}$$

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 (8) to (9) by defining $\tilde{f}_t = -\log(f_t)$.

Another important issue concerns parameter identification. Consider a model density of the following form,

$$p(y_t; \Phi f_{t-1}), \tag{22}$$

where f_t follows a GAS(1,1) specification and Φ is a matrix of constants. For example, Φf_t can be a vector of volatilities of a vector time series driven by a single common factor f_t. In this case, it is not possible to define both Φ and all GAS parameters ω, A, and B, simultaneously. Take the model in (22) and introduce an invertible matrix K. Define $\tilde{f}_t = Kf_t$, $\tilde{s}_t = Ks_t$, $\tilde{\Phi} = \Phi K^{-1}$, $\tilde{\omega} = K\omega$, $\tilde{A} = KAK^{-1}$, $\tilde{B} = KBK^{-1}$. The likelihoods for $p(y_t; \Phi f_{t-1})$ and $p(y_t; \tilde{\Phi}\tilde{f}_{t-1})$ are obviously identical. Pre-multiplying the GAS(1,1) transition equation for the original parameterization by K, we obtain

$$\begin{aligned} Kf_t &= K\omega + KAK^{-1}Ks_t + KBK^{-1}Kf_{t-1} \qquad \Leftrightarrow & (23) \\ \tilde{f}_t &= \tilde{\omega} + \tilde{A}\tilde{s}_t + \tilde{B}\tilde{f}_{t-1}. & (24) \end{aligned}$$

From (21) it follows directly that $\tilde{s}_t$ is the GAS driver for the new parameterization $\tilde{f}_t$. As K is an arbitrary invertable matrix, restrictions must be imposed on Φ to ensure identification. For example, specific rows of Φ 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 in the GAS equation. For example, assume we impose the normalization condition $A = I$. Then this normalization constraint also holds in the reparameterized model $\tilde{f}_t = Kf_t$ as $\tilde{A} = KAK^{-1} = KK^{-1} = I$ where K is an arbitrary invertable matrix. The other parameters in the model, however, will not be the same as for f_t (e.g., Φ versus $\tilde{\Phi} = \Phi K^{-1}$), but the likelihood value remains unchanged. The fact that the identification issue can be solved via restrictions on Φ but not on A in this illustration is a direct consequence of the equivariance of the score and information matrix as a basis for the recursions in the GAS model. We therefore need to take some care in normalizing the parameter spaces of models with a factor structure such as those in (22).

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.

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

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

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

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

where the coefficients λ and β are unknown and with independent disturbance $\varepsilon_{t,\tau} \sim \mathrm{N}(0, \sigma^2)$ for a given time t. For the particular choice of $x_\tau(\lambda)$, the three coefficients in β can be interpreted as the level, slope, and curvature of the term structure, respectively. The slope and curvature factors depend on the parameter λ 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 β, λ and σ^2 can be estimated by non-linear least squares methods. 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.

4.1.1 The dynamic Nelson-Siegel model

Diebold, Rudebusch, and Aruoba (2006) proposed analyzing 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 β as a 3×1 vector of unobservables, β_t, that are modeled as a vector autoregressive (VAR) process. Furthermore, the interest rates for all m maturities at time t are put into the observation vector $y_t = (y_{t,\tau_1}, \ldots, y_{t,\tau_m})'$ and modeled by $y_t = X_t\,\beta_t + \varepsilon_t$ where $X_t = [x_{\tau_1}(\lambda)', \ldots, x_{\tau_m}(\lambda)']'$, with serially independent $m \times 1$ disturbance vector $\varepsilon_t \sim \mathrm{N}(0, D)$ and $m \times m$ positive diagonal matrix D. The unknown parameters in the VAR model for β_t, as well as D and λ are then estimated by ML using the Kalman filter in this parameter driven approach.

In our observation driven approach, we take the GAS factor as $f_{t-1} = \beta_t$. For illustrative purposes we consider the Gaussian density for $y_t - X_t f_{t-1} \sim \mathrm{N}(0, D)$ and take the GAS$(1,1)$ updating equation as $f_t = \omega + A_0 s_t + B_1 f_{t-1}$ with

$$s_t = \left(X_t' D^{-1} X_t\right)^{-1} X_t' D^{-1}\left(y_t - X_t f_{t-1}\right),$$

assuming that $m > 3$ so that matrix $X_t' D^{-1} X_t$ is nonsingular. The static parameter vector is given by

$$\theta = \left(\text{diag}(D)' \,,\, \omega' \,,\, \vec{A_0}' \,,\, \vec{B_1}' \,,\, \lambda \right)' .$$

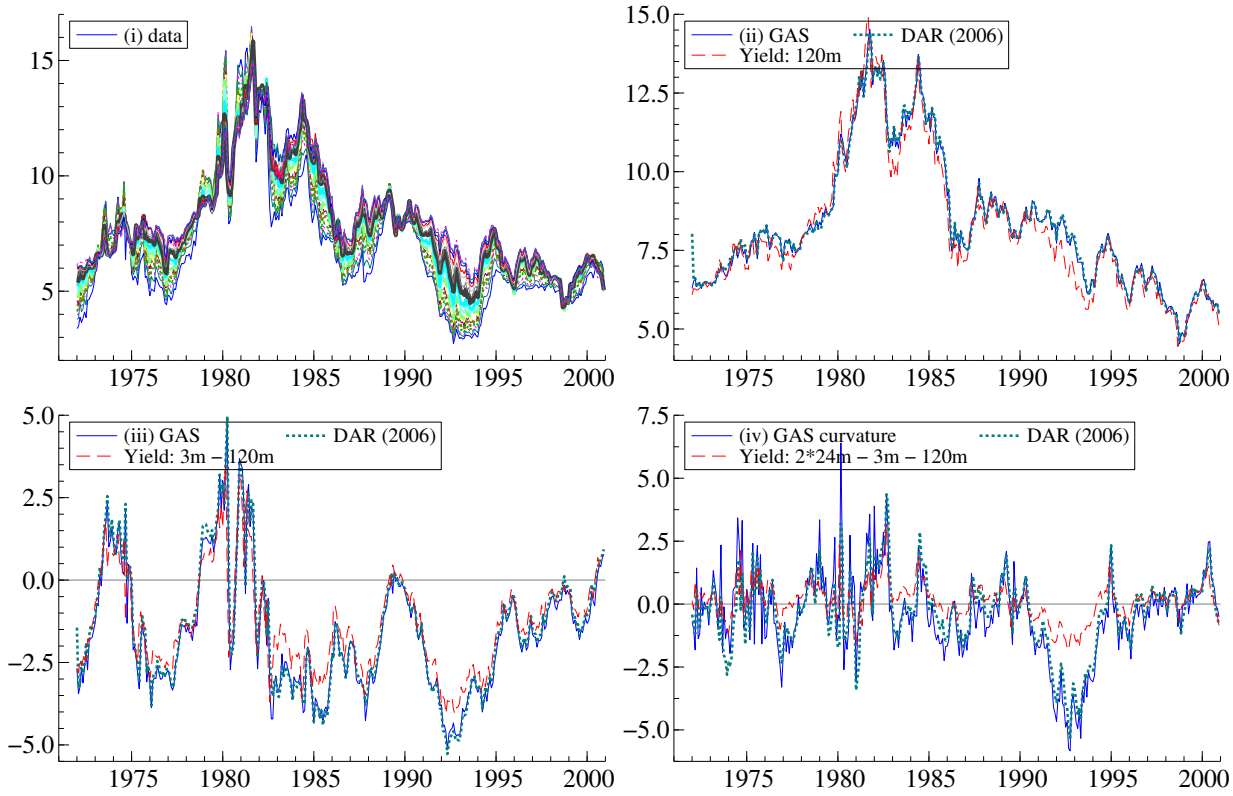

Figure 1: *Three factor dynamic Nelson-Siegel model. Panels (ii-iv) compare the estimated factor from the GAS model with the one-step ahead predicted estimate from the parameter driven DNS model. (i) term structure data; (ii) level factors and the 120 month yield; (iii) slope factors and the spread from the 3 month yield minus the 120 month yield; (iv) curvature factors and the 24 month yield minus the 3 and 120 month yield.*

4.1.2 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. 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 λ 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 estimates of the three factors in β_t are similar as we observe from the graphs in Figure 1. The three estimated factors in both models correspond closely to the empirical proxies of the yield curve factors over time.

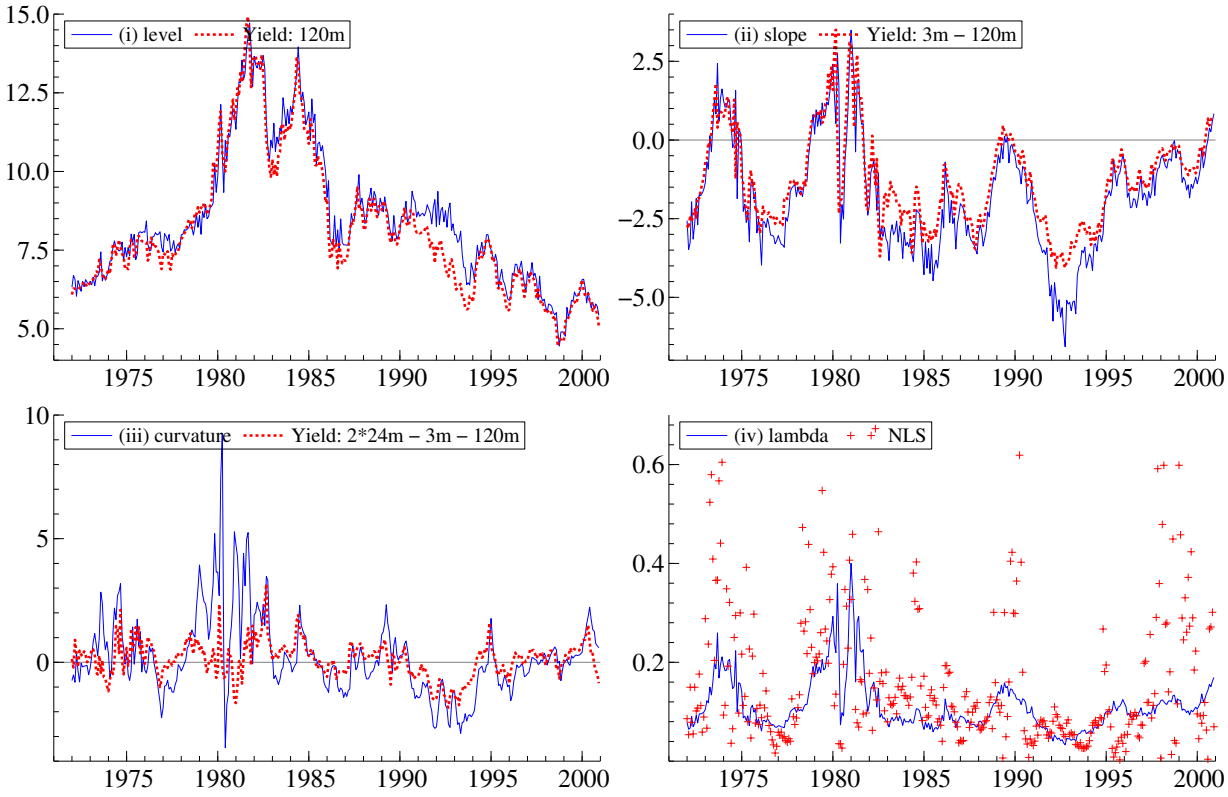

Figure 2: *Nonlinear dynamic Nelson-Siegel model: (i) level factor with the 120 month yield; (ii) slope factor with the spread from the 3 month yield minus the 120 month yield; (iii) curvature factor and the 24 month yield minus the 3 and 120 month yield; (iv) λ_t factor with the non-linear (cross-sectional) least squares estimates of λ in each period.*

4.1.3 Nonlinear extension

In case the coefficients of the Nelson-Siegel model are estimated for a given t by non-linear least squares, the estimate of λ (based on $m = 17$ observations) varies considerably over time for $t = 1, \ldots, n$. We follow Koopman, Mallee, and van der Wel (2009) in extending the DNS model by allowing λ to vary over time. However, here we adopt an observation driven approach using a GAS$(1, 1)$ model. We define 4×1 vector $f_t^+ = (\beta_t', \lambda_t)$ and specify it as $f_t^+ = \phi_0 + \Phi f_t$ where f_t is a 3×1 vector of factors. We thus impose a three-factor structure on the evolution

of (β_t, λ_t), such that we can also investigate the performance of the GAS model when there is non-linearity as well as restrictions on the parameter dynamics. To identify the model, we set the upper 3×3 matrix of the 4×3 matrix Φ equal to the identity matrix and the upper three elements of the 4×1 vector ϕ_0 equal to zero, see Section 3.2. As a result, λ_t is a linear function of β_t. The GAS$(1,1)$ recursion with information matrix scaling is $f_t = \omega + A_0 s_t + B_1 f_{t-1}$ where

$$s_t = \left(\dot{X}_t' D^{-1} \dot{X}_t\right)^{-1} \dot{X}_t' D^{-1} \left(y_t - X_t f_{t-1}\right),$$

and $\dot{X}_t = (X_t, (\partial X_t / \partial \lambda_{t-1}) \beta_{t-1})$. The results from estimating this model are shown in Figure 2. The first three factors are plotted with their empirical proxy from the data. Meanwhile the time-varying λ_t factor is plotted with the nonlinear least squares estimates of λ from the cross-section of yields in each period. The fourth factor λ_t varies considerably and roughly tracks the estimates of λ from the cross-section. By allowing λ 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

Models with time-varying intensities have received much attention in the finance and microeconometric 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 are playing 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, Lucas, and Monteiro (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.

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, Lucas, and Monteiro (2008), we model the log intensities of these processes by

$$\lambda_{jk}(t) = \eta_j + \psi_j' f_{t^*}, \tag{25}$$

where η_j is the baseline intensity and ψ_j is the vector of loadings for f_t, and t^* the last event time before t. The vector of dynamic factors f_t is specified by the GAS(1,1) updating equation (2) with $\omega = 0$. Since f_t is an unobserved process, we may impose sign restrictions for ψ_j to obtain economic interpretations for the factors. This GAS specification states that the intensities of all firms are driven by the same vector of time-varying systematic factors f_t. Model (25) nests the model of Russell (2001) when we set the dimension of f_t equal to the number of firms N. In a credit risk context, we typically have $\dim(f_t) << N$. 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 ψ_j's equal to unity.

The log-likelihood specification using (25) is

$$\ell_t = \sum_{j,k} y_{jk}(t)\lambda_{jk}(t) - R_{jk}(t) \cdot (t - t^*) \cdot \exp(\lambda_{jk}(t^*)), \tag{26}$$

where t^* is the last event time before t, and $R_{jk}(t)$ 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 ℓ_t, we obtain

$$s_{t+1} = \left[\sum_{j,k} w_{jk}(t)\psi_j\psi_j'\right]^{-1} \left(\sum_{j,k} y_{jk}(t)\psi_j - R_{jk}(t) \cdot (t - t^*) \cdot \exp(\lambda_{jk}(t))\psi_j\right), \tag{27}$$

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

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

Table 2: *Estimation results for the parameters in the one-factor GAS(1,1) intensity model* (25) *in a two-grade system, with* $\psi_4 = 1$*, with the scaled scoring function* (27) *and based on the S&P ratings of all US corporates between 1981 and 2005. The estimates are reported with asymptotic standard errors in parantheses below the estimates. Parameter* B_1 *is subject to a logistic transformation during estimation and* ψ_3 *is subject to a identifying restriction* $\psi_3 < 0$*.*

	IG → SIG	IG → DEF	SIG → IG	SIG → DEF
j	1	2	3	4
η	-3.920	-7.360	-3.360	-3.330
	(0.118)	(0.353)	(0.109)	(0.217)
ψ	0.520	1.190	-0.470	1.000
	(0.076)	(0.330)	(0.086)	___
	A_0	logit(B_1)	B_1	
	0.024	6.415	0.998	
	(0.003)	(0.537)		

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 f_t and the updating equation has the scaled score function (27). The resulting model is estimated under the restrictions $\psi_3 < 0$ and $\psi_4 = 1$. The estimation results are presented in Table 2. The GAS parameter B_1 is estimated close to unity which implies a persistent dynamic process for f_t. Given the estimates of ψ_j, the downgrades appear to be most sensitive to the common factor f_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 f_t with a loading estimate of 1.19. Interestingly, the estimated pattern (not shown) of the systematic intensity factor f_t is close to the estimated pattern of the parameter driven model of Koopman, Lucas, and Monteiro (2008). However, in a GAS framework we do not require computationally intensive methods such as importance sampling for parameter and factor estimation.

It is straightforward in our GAS framework to generalize the model to a three-factor model. In this case, A_0 and B_1 in the GAS updating equation become 3×3 matrices. To obtain identification, we set the loading vector ψ_j equal to the jth column of a 4×4 identity matrix for $j = 1, 2, 3$ while $\psi_4 = \psi_3$. This parsimonious specification implies that upgrades and downgrades between IG and SIG have different factors while transitions to default also have a distinct factor. The static parameter vector θ contains the elements of A_0 and B_1 together

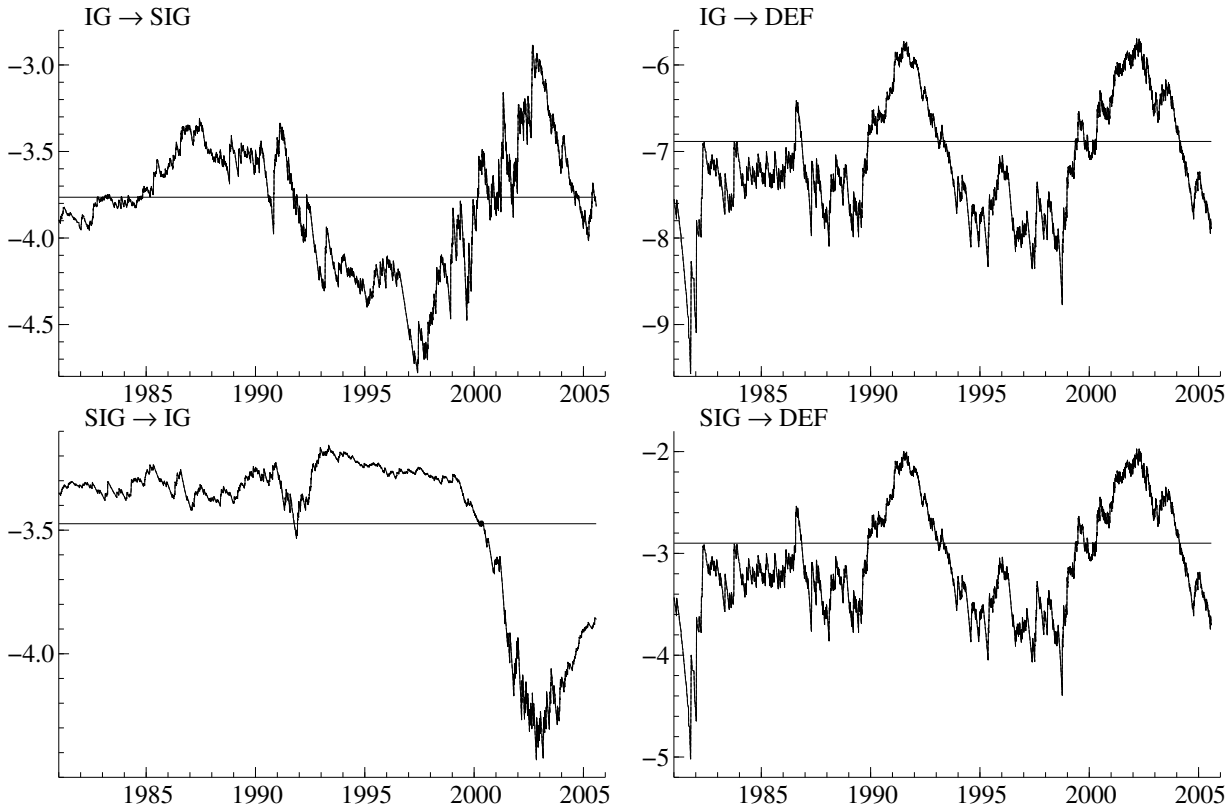

Figure 3: *Marked point-process illustration: the estimated intensities in a two-grade system of the GAS(1,1) model with three credit risk factors, based on the scaled score function* (27) *and using the S&P rating histories of US corporates for the period 1981–2005.*

with the baseline intensities η_j. These parameters can be estimated in the usual way by ML. After parameter estimation, we obtain similar estimated patterns for the three factors in f_t as for the more involved parameter driven model of Koopman, Lucas, and Monteiro (2008). In particular, we corroborate the finding that the dynamics of upgrades are substantially different from those of downgrades and defaults as can be clearly viewed in Figure 3 where the estimated intensities $\eta_j + \psi_j' f_{t-1}$ are displayed.

To conclude this example, note that by reparameterization this GAS model can also be extended to incorporate the time-varying multinomial model of Russell and Engle (2005). To see this for the case with information matrix scaling, consider as an example a setting with two competing risks characterized by the log intensities $f_{1t} = \ln(\lambda_{1t})$ and $f_{2t} = \ln(\lambda_{2t})$. The multinomial model is characterized by the log intensity of the pooled process, $\tilde{f}_{1t} = \ln(\lambda_{1t} + \lambda_{2t})$, and the logit transform of the probability of observing state 1 if an event occurs, $\tilde{f}_{2t} =$

$\ln(\lambda_{1t}/\lambda_{2t})$. Following (21), the GAS driver s_t then only has to be pre-multiplied by the matrix

$$\begin{pmatrix} \frac{\lambda_{1,t-1}}{\lambda_{1,t-1}+\lambda_{2,t-1}} & \frac{\lambda_{2,t-1}}{\lambda_{1,t-1}+\lambda_{2,t-1}} \\ 1 & -1 \end{pmatrix}.$$

4.3 Unobserved component models with a single source of error

Unobserved components or structural time series models are a popular class of parameter driven models where the unobserved components (UC) have a direct interpretation, see Harvey (1989). In this section, we describe observation-driven analogues to UC models. For a univariate time series $y_1, \ldots, y_n$, a univariate signal ψ_t can be extracted. The dynamic properties of ψ_t can be broken into a vector of factors f_{t-1} that are specified by the updating equation (2). For example, we can specify the signal as the sum of r factors, that is

$$\psi_t = f_{1,t-1} + \ldots + f_{r,t-1} \tag{28}$$

with $f_t = (f_{1,t}, \ldots, f_{r,t})'$. In the case $r = 2$, we can specify the first factor as a time-varying trend component (random walk plus drift) and the second factor as a second-order autoregressive process with possibly cyclical dynamics. For this decomposition we obtain the GAS(1,2) model with observation model $y_t = \psi_t + \varepsilon_t = f_{1,t-1} + f_{2,t-1} + \varepsilon_t$, observation density $p(y_t|\psi_t;\theta) = \mathrm{N}(f_{1,t-1} + f_{2,t-1}, \sigma^2)$ and updating equation

$$f_t = \begin{pmatrix} \omega \\ 0 \end{pmatrix} + \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} s_t + \begin{bmatrix} 1 & 0 \\ 0 & \phi_1 \end{bmatrix} f_{t-1} + \begin{bmatrix} 0 & 0 \\ 0 & \phi_2 \end{bmatrix} f_{t-2}. \tag{29}$$

The constant ω is the drift of the random walk trend factor $f_{1,t}$ and the autoregressive coefficients ϕ_1 and ϕ_2 impose a stationary process for the second factor $f_{2,t}$. The scaled score function is given by

$$s_t = y_t - \psi_t = y_t - f_{1,t-1} - f_{2,t-1} = \varepsilon_t, \tag{30}$$

and can be interpreted as the single source of error. The static parameter vector θ, consisting of coeffients ω, a_1, a_2, ϕ_1, ϕ_2 and σ, can be estimated straightforwardly by ML. The estimates of f_t result in a decomposition of y_t into trend, cycle, and noise. This GAS decomposition can be regarded as the observation driven equivalent of the UC models of Watson (1986) and Clark (1989), who also aim to decompose macroeconomic time series into trend and cycle factors. The UC trend-cycle decomposition model is then given by $y_t = f_{1,t} + f_{2,t}$ with

$$f_{1,t} = \omega + f_{1,t-1} + a_1\xi_{1,t}, \qquad \xi_{1,t} \sim \mathrm{N}(0,1), \tag{31}$$

$$f_{2,t} = \phi_1 f_{2,t-1} + \phi_2 f_{2,t-2} + a_2\xi_{2,t}, \qquad \xi_{2,t} \sim \mathrm{N}(0,1), \tag{32}$$

Table 3: *Estimation results for the parameters in the trend-cycle GAS(1,2) decomposition model* (28) *with the updating equation* (29) *and the scaled scoring function* (30) *based on quarterly log U.S. real GDP from 1947(1) to 2008(2). The estimates are obtained by ML and reported with asymptotic standard errors in parantheses below the estimates. Furthermore, the ML estimates of parameters in the parameter driven trend-cycle UC model* (31)–(32) *are reported which are based on the same data set.*

	ω	a_1	a_2	ϕ_1	ϕ_2	σ	log-like
GAS	0.825	0.723	0.563	1.328	-0.424	0.905	-324.51
	(0.043)	(0.206)	(0.202)	(0.130)	(0.142)	(0.041)	
UC	0.825	0.604	0.621	1.501	-0.573	–	-324.06
	(0.040)	(0.098)	(0.112)	(0.102)	(0.106)	_	

where the disturbances $\xi_{1,t}$ and $\xi_{2,t}$ are mutually and serially independent.

To illustrate the GAS trend-cycle decomposition model, we consider the time series of quarterly log U.S. real GDP from 1947(1) to 2008(2) obtained from the Federal Reserve Bank of St. Louis. The vector of static coefficients θ is estimated by ML and the results are reported in Table 3. The estimated autoregressive polynomial for factor $f_{2,t}$ has roots in the complex range and therefore factor $f_{2,t}$ has cyclical properties. We may interpret $f_{2,t}$ as a real-time business cycle indicator for time t which is displayed in Figure 4. To compare this indicator with the indicator produced by the Watson (1986) model, we also report the ML estimates of the corresponding coefficients in an UC trend-cycle model. These estimates are obtained by using the Kalman filter for likelihood evaluation. Parameter estimates for the UC model are reported in Table 3 and the one-step ahead predicted estimate of $f_{2,t}$ is plotted in Figure 4. We find that the parameter estimates from each model correspond closely. The second factor from each model exhibits cyclical behavior and the growth rate of the trend is estimated to be the same. Estimates of the GAS and UC cycle factors in Figure 4 are almost indistinguishable.

The GAS framework is sufficiently general to provide an observation driven alternative for the decomposition of univariate and multivariate time series based on UC models including models with trend, seasonal, cycle and irregular components. For example, the GAS updating equation can also be designed to incorporate the trend and cycle dynamics as formulated by Harvey and Jaeger (1993). Regression and intervention effects can also be incorporated in the GAS specification, see the discussion in Subsection 2.1. Since the resulting GAS models are equivalent to single source of error models, we refer to Ord et al. (1997) for a more detailed discussion on this class of models.

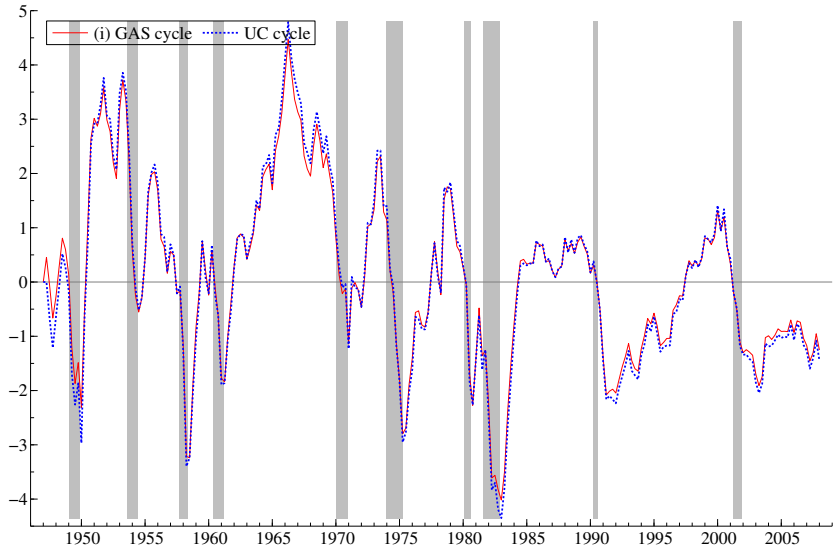

Figure 4: *Trend-cycle illustration: estimated cycles from the GAS and UC trend-cycle models based on quarterly log of U.S. real gdp from 1947(1) through 2008(1). NBER recession dates are indicated by the shaded regions.*

4.4 State space models with time-varying GAS parameters

The GAS framework can also be adopted to let the static parameters of linear Gaussian state space models vary over time. To illustrate its relevance, we consider the local level model as specified by

$$y_t = \mu_t + \varepsilon_t, \qquad \mu_t = \mu_{t-1} + \xi_t, \qquad \varepsilon_t \sim \mathrm{N}(0, \sigma_\varepsilon^2), \quad \xi_t \sim \mathrm{N}(0, \sigma_\xi^2), \quad t = 1, \ldots, n. \tag{33}$$

We treat the variances of the irregular and level disturbances, σ_ε^2 and σ_ξ^2 respectively, as GAS factors in order to obtain a time-varying UC model. The resulting model is similar in spirit to the model considered by Stock and Watson (2007) for forecasting quarterly U.S. inflation. The details of our time-varying local level model are given below. A related idea is to combine state space models with ARCH disturbances; see, e.g. Harvey, Ruiz, and Sentana (1992).

The variances σ_ε^2 and σ_ξ^2 of the local level model (33) can be replaced by two GAS factors to obtain a model with time-varying variances. Since variances must remain positive, we specify the two GAS factors as log-variances. Replacing the constant variances in (33) with the GAS factors, new disturbances for the local level model are given by

$$\varepsilon_t \sim \mathrm{N}\left\{0, \exp(f_{1,t-1})\right\}, \qquad \xi_t \sim \mathrm{N}\left\{0, \exp(f_{2,t-1})\right\}, \qquad t = 1, \ldots, n, \tag{34}$$

with $f_t = (f_{1,t}\,,\, f_{2,t})'$. Conditional on f_{t-1}, the unobserved level μ_t in (33) remains linear in the observations $y_1, \ldots, y_t$ and therefore the Kalman filter can be adopted to produce the optimal

estimate of μ_t which is given by $a_t = \mathrm{E}(\mu_t | f_{t-1}, Y_1^{t-1}, F_1^{t-2}; \theta)$ with its mean square error p_t. The log-likelihood function for the local level model (33) is given by

$$\ell(\theta) = \sum_{t=1}^{n} \ell_t(\theta; y_t, Y_1^{t-1}), \qquad \ell_t = \ell_t(\theta; y_t, Y_1^{t-1}) = -\frac{1}{2}\ln 2\pi - \frac{1}{2}\ln d_t - \frac{1}{2} v_t^2 \,/\, d_t, \tag{35}$$

where the prediction error v_t and its variance d_t are evaluated by the Kalman filter as given by

$$\begin{aligned} v_t &= y_t - a_t, & d_t &= p_t + \exp(f_{1,t-1}), & k_t &= p_t \,/\, d_t, & (36)\\ a_{t+1} &= a_t + k_t v_t, & p_{t+1} &= (1-k_t)p_t + \exp(f_{2,t-1}), & & & (37) \end{aligned}$$

for $t = 1, \ldots, n$ with diffuse initializations $a_1 = 0$ and $p_1 = \kappa$ while $\kappa \to \infty$, see Durbin and Koopman (2001). The Kalman filter update can be carried out simultaneously with the GAS updating equation (2) for f_t and with the scaled score function defined by (3). In this case, we have

$$\nabla_{i,t} = \frac{\partial \ell_t}{\partial \exp(f_{i,t-1})} \times \frac{\partial \exp(f_{i,t-1})}{\partial f_{i,t-1}} = \exp(f_{i,t-1}) \, \frac{\partial \ell_t}{\partial \exp(f_{i,t-1})} \tag{38}$$

for $i = 1, 2$. Given the Kalman filter equations (36)–(37), the latter term of (38) is evaluated by

$$\frac{\partial \ell_t}{\partial \exp(f_{i,t-1})} = -\frac{1}{2}\left(\dot{d}_{it} + 2 v_t \dot{v}_{it}\right) / \, d_t + \frac{1}{2}\dot{d}_{it} \left(v_t \,/\, d_t\right)^2, \tag{39}$$

where $\dot{v}_{it} = \partial v_t \;/\; \partial \exp(f_{i,t-1})$ and $\dot{d}_{it} = \partial d_t \;/\; \partial \exp(f_{i,t-1})$ are evaluated by the additional recursions

$$\begin{aligned} \dot{v}_{it} &= -\dot{a}_{it}, & \dot{d}_{it} &= \dot{p}_{it} + 1(i=1), & \dot{k}_{it} &= (\dot{p}_{it} - k_t \dot{d}_{it}) \;/\; d_t, & (40)\\ \dot{a}_{i,t+1} &= (1-k_t)\dot{a}_{it} + \dot{k}_{it} v_t, & \dot{p}_{i,t+1} &= -\dot{k}_{it} p_t + (1-k_t)\dot{p}_{it} + 1(i=2), & & & (41) \end{aligned}$$

where $1(i=j)$ equals one if $i = j$ and zero otherwise, with initializations $\dot{a}_{i1} = 0$ and $\dot{p}_{i1} = 0$, for $t = 1, \ldots, n$ and $i = 1, 2$. For the local level model, we therefore obtain two additional recursions to evaluate the score and they can be carried out simultaneously with the Kalman filter. By following Harvey (1989, p 140–2), we approximate the information matrix by

$$\mathcal{I}_{t-1}(i,j) = -\mathrm{E}_{t-1}\left(\frac{\partial^2 \ell_t}{\partial f_{i,t-1} \partial f'_{j,t-1}}\right) \approx \exp(f_{i,t-1}) \exp(f_{j,t-1}) \times \left(\frac{1}{2}\dot{d}_{it}\dot{d}_{jt} \,/\, d_t^2 + \dot{v}_{it}\dot{v}_{jt} \,/\, d_t\right), \tag{42}$$

where $\mathcal{I}_{t-1}(i,j)$ is the (i,j) element of $\mathcal{I}_{t-1}$ for $i, j = 1, 2$. The computation of this approximation is feasible given the additional recursions (40)–(41). However, in practice, the information matrix may become singular or close to singular. We therefore have adopted the EWMA smoothing scheme for the information matrix to obtain the scaled score function, see the discussion in Subsection 2.3.

The Kalman filter, the additional recursions for the score, and the GAS updating equation for f_t are carried out simultaneously. The parameter vector θ consists of the GAS updating

coefficients in ω, $A_0, \ldots, A_{p-1}$, $B_1, \ldots, B_q$ and, possibly, the smoothing coefficient α of the EWMA smoothing recursion for the information matrix. The estimation of θ is done by ML via the maximization of $\ell(\theta)$ in (35) with respect to θ. Given a value of θ, the recursions in real-time provide estimates of μ_t via the Kalman filter and estimates of f_t via the score recursions and the GAS updating equation simultaneously.

To illustrate the new GAS model, we consider a time series of quarterly U.S. Consumer Price Index inflation from 1959(1) to 2007(2) obtained from the FRED database. The local level model with a GAS(1,1) updating equation for the log-variances is adopted and the methodology of estimation as discussed above is implemented. The EWMA smoothing scheme for the information matrix depends on α which is estimated as part of θ. The GAS coefficient matrices A_0 and B_1 are chosen to be diagonal so that we need to estimate a total of seven coefficients. The estimation results are given by

$$\hat{\omega} = \begin{pmatrix} 0.122 \\ 0.003 \end{pmatrix}, \qquad \hat{A}_0 = \begin{bmatrix} 0.426 & 0 \\ 0 & 0.081 \end{bmatrix}, \qquad \hat{B}_1 = \begin{bmatrix} 0.569 & 0 \\ 0 & 0.916 \end{bmatrix}, \qquad \hat{\alpha} = 0.674,$$

with the maximized loglikelihood value given by -372.97. The loglikelihood value for a standard local level model is given by -394.19 indicating a substantial improvement in fit with the GAS model specification.

Panels (i) and (ii) of Figure 5 present the estimated factors $\exp(f_{1,t}/2)$ and $\exp(f_{2,t}/2)$, which are the standard deviations for ε_t and ξ_t, respectively. The standard deviation of the observation disturbance is moderate until the end of 2006 onwards at which time it has become relatively high. The standard deviation of the level disturbance increased in the 1970's during the periods of higher inflation and then decreased steadily over the remaining sample. The signal to noise ratio in our GAS framework is defined by the ratio $q_t = \exp(f_{2,t}) \,/\, \exp(f_{1,t})$. When it is low, the estimate of μ_t is based on a long range of past observations. When it is high, μ_t is estimated using only a small set of recent observations. The third graph in Figure 5 displays q_t based on the estimate of f_t. As the properties of the model suggested, the estimate of μ_t is only based on recent observations during the years of the oil-crisis, 1974–1976. From 1980 onwards, the level of inflation is more stable and a longer stretch of past observations are used in the estimator of the trend. Finally, the (filtered) estimate of μ_t is displayed in graph (iv) of Figure 5. The GAS framework captures the overall development of U.S. inflation effectively. The estimated patterns of the time-varying standard deviations are similar to those obtained by Stock and Watson (2007) who used Markov chain Monte Carlo to estimate the time-varying variances in their parameter driven model.

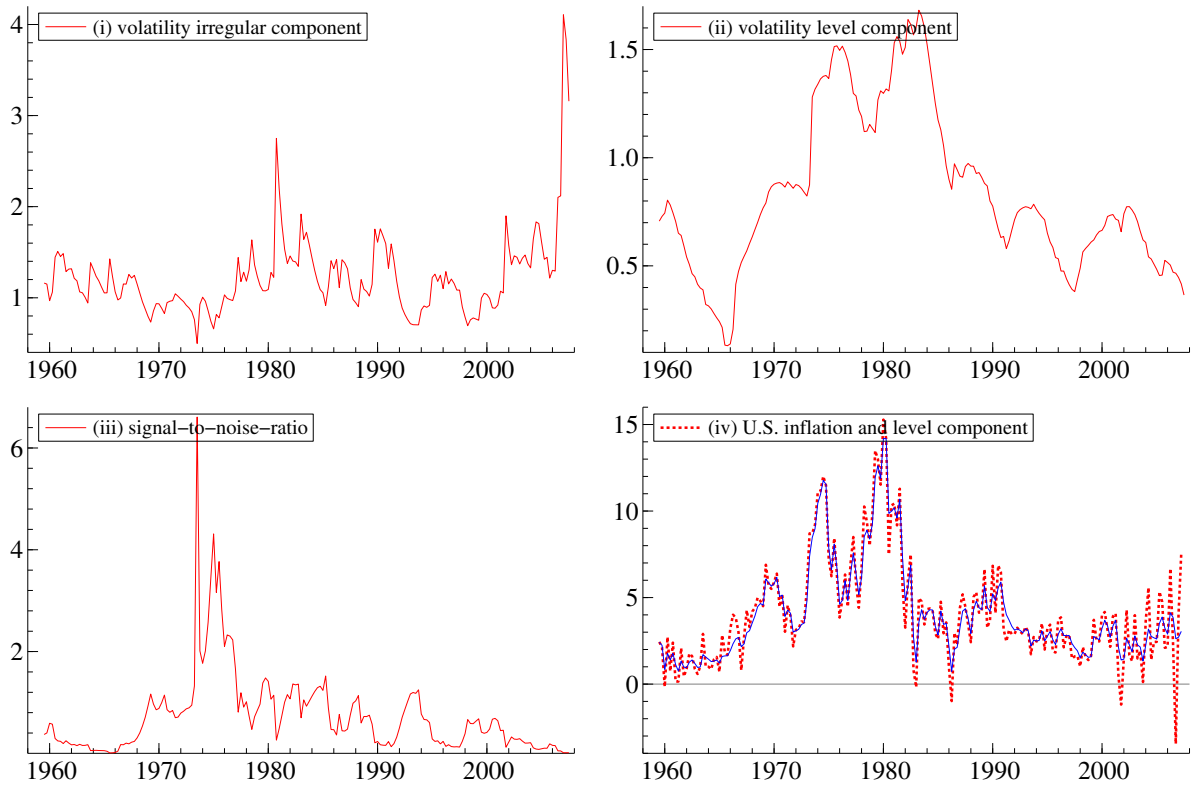

Figure 5: *A local level model illustration: estimation results for the time-varying variances. Results are for the quarterly U.S. inflation from 1959(1) to 2007(2): (i) estimated volatility* $\exp(f_{1,t}/2)$ *of the irregular component; (ii) estimated volatility* $\exp(f_{2,t}/2)$ *of the trend component; (iii) estimated signal-to-noise ratio* q_t*; (iv) U.S. inflation and estimated trend* μ_t*.*

4.5 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 subsection, 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.

4.5.1 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). Patton (2006) models[1] the driving mechanism for the dynamic bivariate Gaussian copula as

$$f_t = \omega + A \sum_{i=0}^{m-1} \Phi^{-1}(u_{1,t-i})\Phi^{-1}(u_{2,t-i}) + B f_{t-1}, \tag{43}$$

where Φ^{-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 smoothing parameter. The (Gaussian) correlation parameter ρ_t is obtained via the transformation $\rho_t = (1-\exp(-f_t))/(1+\exp(-f_t))$. Equation (43) 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 (43) then reduces to

$$f_t = \omega + A \cdot y_t + B \cdot f_{t-1}. \tag{44}$$

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

$$f_t = \omega + A\frac{2(y_t - \rho_{t-1} - \rho_{t-1}(1+\rho_{t-1}^2)^{-1}(x_t - 2))}{(1-\rho_{t-1}^2)} + B f_{t-1}. \tag{45}$$

The similarities and differences between (44) and (45) are clear. Both models are driven by y_t as positively clustered transformed marginals should increase the correlation parameter. The additional scaling factor $2/(1-\rho_{t-1}^2)$ in (45) is a consequence of modeling the transformed correlation parameter f_t rather than ρ_t directly. The additional ρ_{t-1} term in the numerator of the second term in (45) enforces $y_t - \rho_{t-1}$ to be a martingale difference. The most interesting difference between the two model specification is the final term involving x_t. The term $x_t - 2$ 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 ρ_{t-1}. If the correlation is positive, the impact on the value of x_{t-1} is negative. In this case, the x_t term offsets part of the effect of y_t if the latter has a positive

[1]We adapt Patton's notation here slightly to correspond with the timing convention used in the current paper, i.e., using f_{t-1} in the copula at time t rather than f_t.

value, i.e., if y_t corresponds with the current positive estimate of ρ_{t-1}. If y_t has a negative value, however, the x_t term reinforces the magnitude of the GAS step triggered by y_t.

The effects are visualized in Figure 6 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 peresented. Note that each pair of top and bottom graphs has the same scale on the vertical axis. If we consider the plot for $\rho = 0.9$, 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 $\rho = 0.2$.

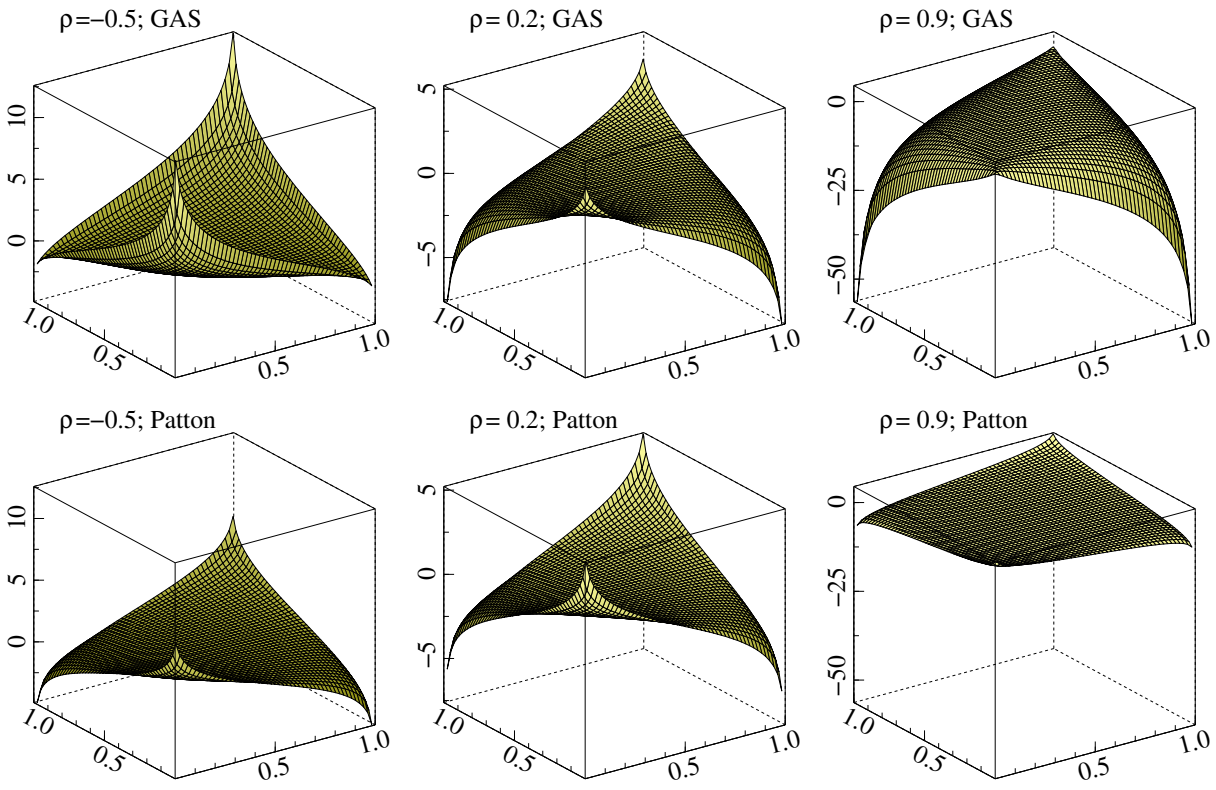

Figure 6: *A bivariate Gaussian copula illustration: comparisons between the GAS and Patton drivers as a function of the uniforms* (u_{1t}, u_{2t}). *The top panels contain the graphs for the GAS step in* (45) *for* $\rho_{t-1} = -0.5, 0.2, 0.9$ *(left, middle, right). The lower graphs contain the (re-centered) steps* $y_t - \rho_{t-1}$ *of the Patton model,* (44). *The vertical axes have the same scale for each column of graphs.*

A more striking feature, however, is the increased sensitivity along the off-diagonal areas for positive ρ. If the current estimate of ρ 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 ρ downwards compared to the Patton step. For negative values of ρ, 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.

4.5.2 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 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 (44) and (45), 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

$$\rho_t = \omega + A \cdot y_t + B \cdot \rho_{t-1}. \tag{46}$$

To enforce the stationarity property of this process, we estimate the logit transform of B. The results are presented in Table 4 and Figure 7.

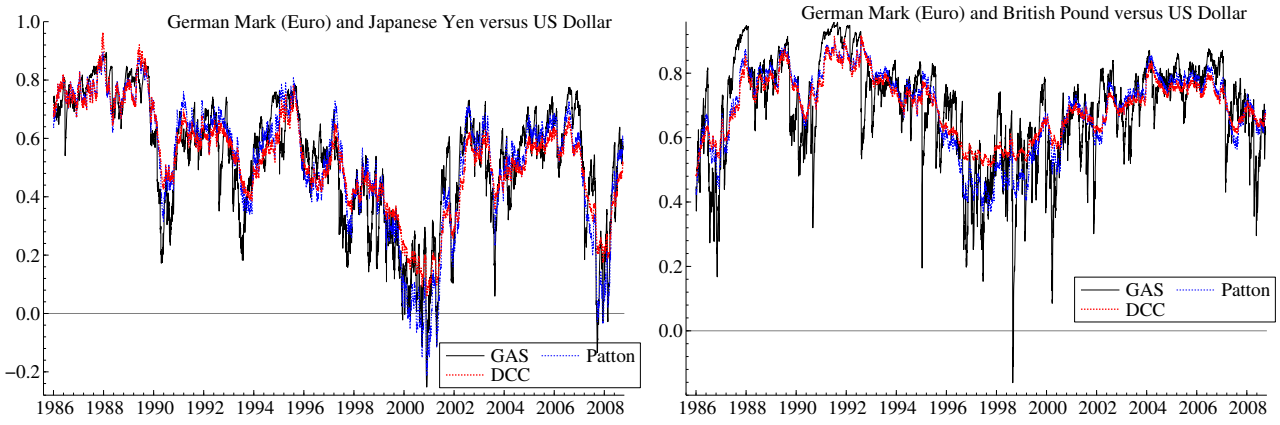

Figure 7: *A copula illustration: comparisons of the correlation parameter estimates for the GAS, Patton, and DCC drivers in* (44)–(46). *The data are the marginal AR-GARCH transforms of log exchange rates for the German Mark-US dollar and Japanese Yen-US dollar (left panel) and for the German Mark-US dollar and British Pound-US dollar (right panel). The sample period is January 1986–August 2008.*

Table 4 shows that the GAS specification increases the log-likelihood value 25 to 125 points for the same number of parameters. The figures show the empirical estimates of the time-varying

Table 4: *Parameter estimates for the GAS, Patton, and DCC drivers in* (44)–(46). *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, otherwise standard errors in parentheses.*

	$10^3\omega$	A	$\ln(\frac{B}{1-B})$	B	log-like
	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)	
DCC	1.03	0.008	4.65	0.991	1184.13
	(0.29)	(0.001)	(0.09)	(0.989,0.992)	
	German Mark (Euro)–US \$, British Pound–US \$				
GAS	12.55	0.082	4.97	0.993	2218.82
	(3.55)	(0.008)	(0.26)	(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)	
DCC	2.64	0.004	4.84	0.992	2060.43
	(0.39)	(0.000)	(0.11)	(0.990,0.994)	

correlation. Based on the estimates of the parameter B, 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 sharpe decline at the edges as visualized in Figure 6, 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.

4.5.3 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 as given by

$$f_t = \omega - m^{-1} A \sum_{i=0}^{m-1} |u_{1,t-i} - u_{2,t-i}| + B f_{t-1}, \tag{47}$$

where f_t captures the dependence between the coordinates. The intuition for (47) is clear. If the most recent u_{1t} and u_{2t} are close together, this is a signal of strong dependence and, therefore, f_t is increased. Similarly f_t is decreased if u_{1t} and u_{2t} are far apart.

Though the driving mechanism in (47) is intuitively straightforward, two issues are less clear. First, (47) 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 (47) provides an easy updating scheme for the bivariate case, the extension to the multivariate case is less obvious. In particular, if one has an Archimedian 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 (47) 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

$$C(u_1, \ldots, u_d) = \left(1 - d + \sum_{i=1}^{d} u_i^{-\alpha}\right)^{-1/\alpha}. \tag{48}$$

The Clayton copula is characterized by the dependence parameter α. Low values of α 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 specify $\alpha = f_{t-1}$ and define $S(\alpha) = \sum_{i=1}^{d} u_i^{-\alpha}$. The Clayton copula has pdf

$$c(u_1, \ldots, u_d) = (1 - d + S(\alpha))^{-1/\alpha - d} \cdot \prod_{i=0}^{d-1} \left((1 - i \cdot \alpha) u_i^{-\alpha-1}\right). \tag{49}$$

We obtain the score vector

$$\nabla_t = -\sum_{i=0}^{d-1} \left(\frac{i}{1 - i \cdot \alpha} - \ln(u_i)\right) + \frac{1}{\alpha^2} \ln(1 - d + S(\alpha)) + \tag{50}$$

$$\left(\frac{1}{\alpha} + d\right) \frac{\sum_{i=1}^{d} u_i^{-\alpha} \ln(u_i)}{1 - d + S(\alpha)}.$$

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

$$\mathcal{I}_{t-1} = \mathrm{E}_{t-1}\left[(\nabla_t)^2\right] \equiv h(f_{t-1}), \tag{51}$$

with the score vector ∇_t as defined in (50). Note that the function $h(\cdot)$ in (51) does not depend on time or on any parameter other than f_{t-1}. We can therefore construct a grid of values $f^{(0)} < \ldots < f^{(n)}$ and compute the function value $h(f^{(j)})$ 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 θ and set the starting value f_0. Using interpolation, compute $h(f_0)$ and use it to scale the score step $s_1 = \nabla_1/h(f_0)$. Compute the new parameter value f_1 through the GAS recursion, and again use interpolation to obtain $h(f_1)$. This process is repeated for the complete sample. Finally, the likelihood can be computed.

4.5.4 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,

$$C(u_1, \ldots, u_d) = \sum_{i=1}^{r} p_i C_i(u_1, \ldots, u_d), \tag{52}$$

with copula functions C_i and corresponding pdf c_i. 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

$$\frac{\partial \ln c}{\partial \theta} = \sum_{i=1}^{r} w_i \cdot \frac{\partial \ln c_i}{\partial \theta}, \tag{53}$$

and

$$\frac{\partial^2 \ln c}{\partial \theta \partial \theta'} = \sum_{i=1}^{r} w_i \cdot \left(\frac{\partial^2 \ln c_i}{\partial \theta \partial \theta'} + \frac{\partial \ln c_i}{\partial \theta} \frac{\partial \ln c_i}{\partial \theta'} \right) - \left(\sum_{i=1}^{r} w_i \cdot \frac{\partial \ln c_i}{\partial \theta} \right) \left(\sum_{i=1}^{r} w_i \cdot \frac{\partial \ln c_i}{\partial \theta} \right)', \tag{54}$$

and thus,

$$\mathrm{E}_{t-1}\left[\frac{\partial^2 \ln c}{\partial \theta \partial \theta'} \right] = -\mathrm{E}_{t-1}\left[\left(\sum_{i=1}^{r} w_i \cdot \frac{\partial \ln c_i}{\partial \theta} \right) \left(\sum_{i=1}^{r} w_i \cdot \frac{\partial \ln c_i}{\partial \theta} \right)' \right],$$

such that the scores of the individual copulas can be used directly to build the 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 α_L that accounts for lower tail dependence. The second component of the mixture is the survival Clayton copula, characterized by the parameter α_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 α_L or α_U, i.e., to either the upper or lower tail dependence. 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 α_U, and similarly in the lower tail for α_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.

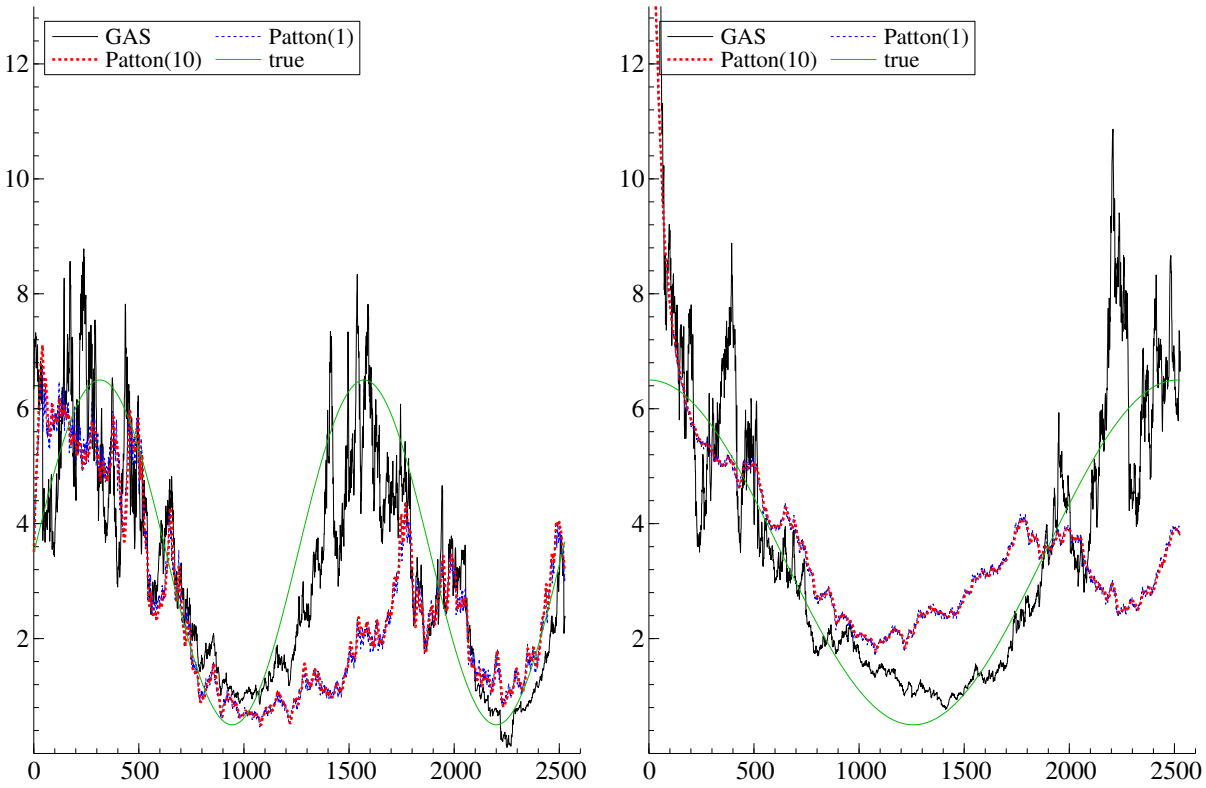

Figure 8: *Symmetrized Clayton copula illustration: comparisons between the correlation parameter estimates from the GAS framework and the Patton model based on a simulated data set.*

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 8 for smoothing parameter values $m = 1$ and $m = 10$.

It is clear that the driving mechanism based only on averages of $|u_{it} - u_{jt}|$ does capture 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 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.

4.6 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) proposed to allow the degrees of freedom parameter to be time-varying. Harvey and Siddique (1999), Jondeau and Rockinger (2003) and Brooks et al. (2005) consider models with time-varying skewness and kurtosis. We develop a t-GAS$(1,1)$ model for $y_t = \sigma_{t-1}\varepsilon_t$ where $\varepsilon_t \sim \mathrm{t}_{\nu_t}$. The error term is scaled to have unit variance such that σ^2_{t-1} is the conditional variance while ν_t is the time-varying degrees of freedom parameter. Define the vector of factors as $f_t = (\sigma^2_t, -\ln\left\{\frac{b-a}{\nu_t-a} - 1\right\})$ where the latter factor is the inverse of the logit transformation which is used to keep ν_t in the interval $[a, b]$. In our empirical work, we select the interval $[2.01, 30]$ to ensure that the conditional variance exists, i.e. $\nu_t > 2$. We note that it is possible to select the conditional kurtosis as a factor instead of ν_t but for some time series the conditional kurtosis may not exist.

Taking derivatives of the observation density with respect to σ^2_t and ν_t, we obtain the score vector as given by

$$\nabla_t = \begin{bmatrix} -\frac{1}{2\sigma^2_t} + \frac{(\nu_t+1)}{2}\left(1 + \frac{y^2_t}{(\nu_t-2)\sigma^2_t}\right)^{-1}\frac{y^2_t}{(\nu_t-2)\sigma^4_t}, \\ \frac{1}{2}\left\{\Gamma'(\frac{\nu_t+1}{2}) - \Gamma'(\frac{\nu_t}{2})\right\} - \frac{1}{2\nu_t} - \frac{1}{2}\ln\left(1 + \frac{y^2_t}{(\nu_t-2)\sigma^2_t}\right) + \frac{(\nu_t+1)}{2}\left(1 + \frac{y^2_t}{(\nu_t-2)\sigma^4_t}\right)^{-1}\frac{y^2_t}{(\nu_t-2)^2\sigma^4_t}. \end{bmatrix},$$

and with some additional derivations the elements of the information matrix are given by

$$E_{t-1}[\nabla_t\nabla'_t] = \begin{bmatrix} -\frac{\nu_t}{2\sigma^4_t(\nu_t+3)} & -\frac{3}{2\sigma^2_t(\nu_t+1)(\nu_t+3)(\nu_t-2)} \\ -\frac{3}{2\sigma^2_t(\nu_t+1)(\nu_t+3)(\nu_t-2)} & \frac{1}{4}\left\{\Gamma''\left(\frac{\nu_t+1}{2}\right) - \Gamma''\left(\frac{\nu_t}{2}\right)\right\} + \frac{(\nu_t+4)(\nu_t-3)}{2(\nu_t-2)^2(\nu_t+1)(\nu_t+3)} \end{bmatrix},$$

where the functions Γ' and Γ'' are the digamma and trigamma functions which can be evaluated in any matrix programming software. 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 (21). We label this model the tv-t-GAS(1,1) model.

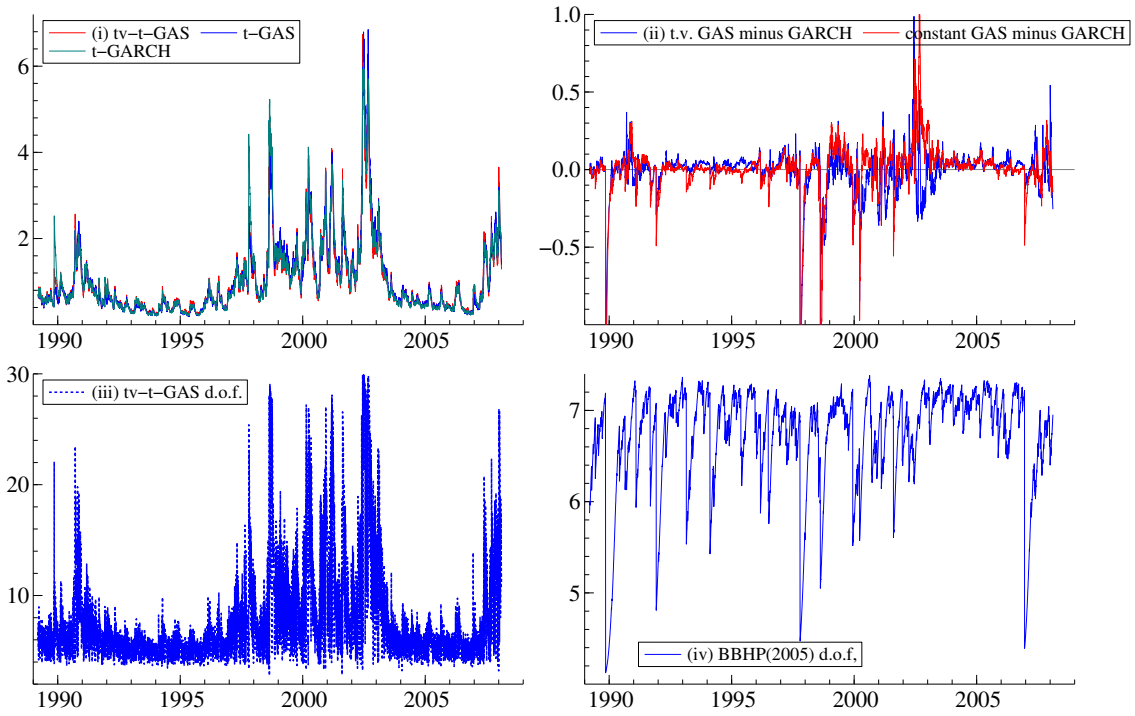

Figure 9: *Time-varying degrees of freedom illustration: (i) estimated conditional variances from the t-GAS(1,1), t-GARCH(1,1), and tv-t-GAS(1,1) models; (ii) differences between the two GAS(1,1) models and the t-GARCH(1,1) model; (iii) estimated time-varying degrees of freedom from the tv-t-GAS(1,1) model; (iv) estimated time-varying degrees of freedom from the GARCH model of Brooks et al. (2005).*

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 ν, that is equation (7), and a standard t-GARCH(1,1) model with constant ν as in Bollerslev (1987). Parameter estimates from each of these models are reported in Table 5 and estimates of the conditional variance are plotted in panel (i) of Figure 9. Focusing on the t-GAS(1,1) model versus the t-GARCH(1,1) model, we see that the log-likelihood values are close. Both the persistence parameter b_{11} and degrees of freedom are estimated to be larger for the t-GAS(1,1) model than for the t-GARCH(1,1) model. Estimates of the conditional variance in panel (i) are hard to distinguish from one another with the exception of those periods when there are outliers. To see this more clearly, we also plot the differences between the estimates from the two GAS models minus the GARCH model in panel (ii) of Figure 9. 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 than

from both GAS models. These are the large negative values in panel (ii). The difference in estimated degrees 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 and, relative to this level, large returns are not outliers. Estimates of the conditional variance from the GAS and GARCH models are still significantly different and economically meaningful during this period.

Turning our attention to the tv-t-GAS(1,1) model, the estimated time-varying degrees of freedom from this model is plotted in panel (iii) of Figure 9 and these estimates demonstrate significant variability. The log-likelihood 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 conditional kurtosis evolves independently from σ_t^2 according to its own GARCH(1,1) recursion. The implied estimates of ν_t can be calculated straightforwardly.

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 the literature on time-varying higher-order moments, the factors are typically forced to evolve independently by imposing zero restrictions on b_{12} and b_{21}. The estimated autoregressive coefficients b_{21} and b_{22} reported in Table 5 for the GAS model imply that both σ_t^2 and ν_t are driven by the same factor because b_{22} is close to zero. Accordingly, the estimates of ν_t in panel (iii) exhibit a similar pattern with the conditional variance in panel (i). Estimates of ν_t from the tv-t-GARCH(1,1) model, which imposes these restrictions, result in a different behavior for the time-varying degrees of freedom. The parameter b_{22} is estimated to be significant and persistent in this model.

To investigate this result further, we split the sample in half before and after 1998 and estimated ν using the t-GAS(1,1) model with constant degrees of freedom on the two sub-samples. Estimates from this model on the two sub-samples are reported in the right-hand columns of Table 5. The degrees of freedom parameter and its standard error clearly increase in the second half of the sample. Estimates of ν on the two sub-samples from the t-GARCH(1,1) model (not reported) are similar. Although this result may seem counterintuitive initially, the reason is that large returns during this period are no longer extreme outliers because the conditional volatility σ_t^2 is higher. This provides support for estimates of ν_t from our model and some evidence that modeling higher-order moments independently of the conditional variance may be inappropriate. The models described in this section might be improved further by linking the time-varying behavior of the degrees of freedom with a time-varying level parameter

Table 5: *Estimates from the 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. The tv-t-GARCH(1,1) model is from Brooks et. al. (2005). The full sample results are on the left. Split sample results for the t-GAS(1,1) model are on the right.*

	tv-t-GAS	t-GARCH	t-GAS	tv-t-GARCH	t-GAS pre-1998	t-GAS post-1998
ω_1	0.006	0.003	0.004	0.003	0.002	0.007
	(0.005)	(0.001)	(0.001)	(0.001)	(0.001)	(0.003)
ω_2	-2.373	-	-		-	-
	(0.310)					
a_{11}	0.057	0.047	0.044	0.049	0.026	0.061
	(0.007)	(0.007)	(0.006)	(0.007)	(0.006)	(0.009)
a_{12}	-0.128	-	-	-	-	-
	(0.043)					
a_{21}	-0.219	-	-	-	-	-
	(0.033)					
a_{22}	-1.498	-	-	0.005	-	-
	(0.002)			(0.006)		
b_{11}	0.994	0.951	0.997	0.949	0.997	0.995
	(0.003)	(0.007)	(0.002)	(0.007)	(0.003)	(0.004)
b_{12}	0.000	-	-	-	-	-
	(0.000)					
b_{21}	0.982	-	-	-	-	-
	(0.154)					
b_{22}	0.026	-	-	0.965	-	-
	(0.121)			(0.024)		
ν	-	6.699	7.032	-	5.367	10.96
		(0.622)	(0.677)		(0.610)	(2.074)
log-like	-6138.18	-6153.02	-6156.46	-6153.44	-2359.55	-3778.63

ω_t in the variance. We leave this extension to future research.

4.7 Time-varying multinomial

Trade by trade financial transaction prices lie on a discrete grid with most price changes taking only a small number of values. Russell and Engle (2005) proposed modeling this behavior using a conditional multinomial distribution with time-varying probabilities in conjunction with their ACD model. We construct a GAS version of their model. Consider the case where the observed series y_t, for $t = 1, \ldots, n$, has a J-dimensional multinomial distribution with vector of probabilities π_t and let $\pi_{j,t}$ be the jth element of this vector. The vector of factors f_t will have dimension $J-1$ with elements $f_{jt} = \ln \pi_{jt} - \ln(1 - \sum_{j=1}^{J-1} \pi_{jt})$ where the final probability $\pi_{J,t}$ is determined by the constraint that they sum to one. Denote $\tilde{y}_t$ and $\tilde{\pi}_t$ as the corresponding $J-1$ dimensional vectors with the Jth elemented omitted. The score with respect to $f_{j,t-1}$ is given by

$$\nabla_{jt} = \tilde{y}_{jt} - \tilde{\pi}_{j,t-1}, \tag{55}$$

while the diagonal and off-diagonal elements of the information matrix are given by

$$\mathcal{I}_{ii,t-1} = \tilde{\pi}_{i,t-1}(1 - \tilde{\pi}_{i,t-1}), \tag{56}$$

$$\mathcal{I}_{ij,t-1} = -\tilde{\pi}_{i,t-1}\tilde{\pi}_{j,t-1}. \tag{57}$$

Combining these results, a GAS(p, q) model for the multinomial distribution reduces to

$$f_t = \omega + \sum_{i=0}^{q-1} A_i S_{t-i-1}(\tilde{y}_{t-i} - \tilde{\pi}_{t-i-1}) + \sum_{j=1}^{p} B_i f_{t-j}, \tag{58}$$

where the scale matrix $S_{t-1} = \mathcal{I}_{t-1}^{-1}$ can be constructed from (56) and (57). The ACM model of Russell and Engle (2005) can be obtained as a special case of the GAS model (58) by selecting the scale matrix S_{t-1} to be the identity matrix. They also add the expected durations from an ACD model as explanatory variables in (58).

As an empirical illustration, we use transaction data from the NYSE TAQ database on Royal Dutch Shell A (RDSA) for the month of November 2007. After retaining trades between 9:30 and 4:00, there are 61,690 trades remaining. Panels (i)-(ii) of Figure 10 contain the observed price changes and observed durations for the first 23,500 trades, while panel (iii) is a histogram of all the trades. The observed durations give evidence of diurnal patterns that are typical of transactions data. In addition, the observed price changes indicate that the probabilities should contain a similar diurnal pattern, as trades with large tick sizes are less likely during openning and closing of the market when volume is higher.

In our sample, 98% of the price changes fall within a $\pm$ 5 tick range of zero (see panel (iii)), where a tick is now 1 cent after decimilization of the market in 2001. Decimilization

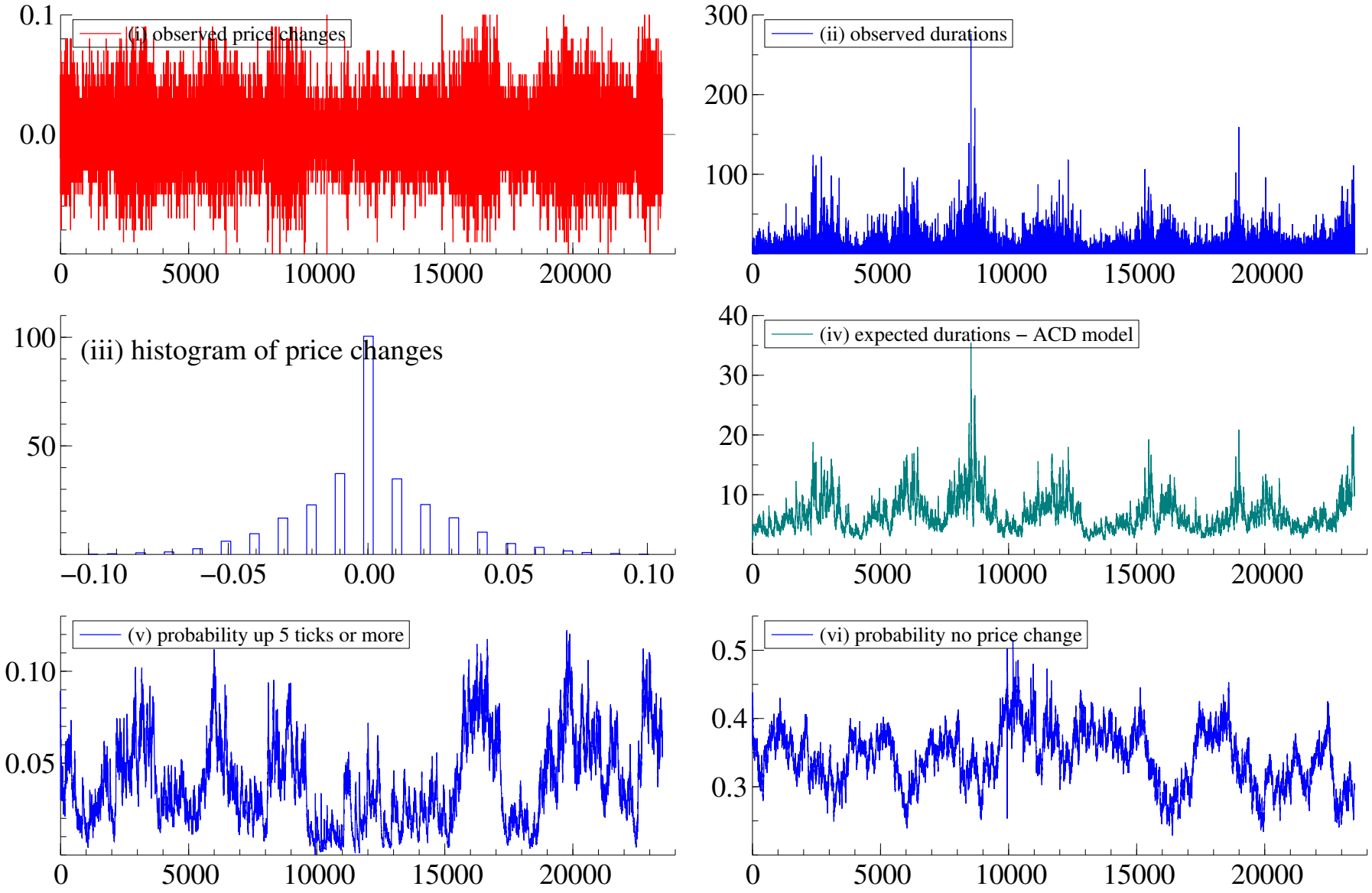

Figure 10: *Time-varying multinomial GAS(1,2)-ACD(1,2) illustration: (i) observed price changes; (ii) observed durations; (iii) histogram of price changes; (iv) estimated expected duration from the ACD model; (v) estimated probability of an increase of 5 ticks or more; (vi) estimated probability of a trade with no change in price.*

unfortunately causes an increase in the required dimension of the factor f_t and a corresponding increase in the number of parameters to estimate. For this example, f_t will have a minimum of 10 dimensions meaning that the A_0 matrix in an ACM(1,1) model will have 100 parameters. Our solution to this problem is to define new factors $\tilde{f}_t$ as $f_t = \Phi_0 + \Phi_1 \tilde{f}_t$ where $\tilde{f}_t$ has $\dim(\tilde{f}_t) << \dim(f_t)$. The GAS(1,1) model reduces to

$$\tilde{f}_t = A_0 \Phi_1' \mathcal{I}_{t-1}^{-1} \Phi_1 \Phi_1' (\tilde{y}_t - \tilde{\pi}_{t-1}) + B_1 \tilde{f}_{t-1}, \tag{59}$$

where the matrix Φ_1 must be restricted to identify the model. For illustration purposes, we selected $\dim(\tilde{f}_t) = 3$ and set the upper 3×3 elements of Φ_1 equal to the identity matrix for identification. Following Russell and Engle (2005), we include expected durations in (59) and jointly estimate the ACD model. We also restrict the matrices B_j to be diagonal. Specifying a multinomial-GAS(1,2)-ACD(1,2) model for this series, some of the estimated time-varying probabilities for the first third of the data set are shown in panels (v) and (vi) of Figure 10. Panel (v) is a plot of the probability of a price increase of 5 ticks or more while panel (vi) plots the probability of no price movement. The model picks up the diurnal dynamics of the price changes reasonably well with the probability of an increase of 5 ticks or more changing considerably throughout the day. An alternative observation driven model for trade-by-trade data has been proposed by Rydberg and Shephard (2003) using the GLAR methodology of Shephard (1995). We note that a GAS version of their model will be slightly different but close to their specification.

4.8 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 or sub-model has a likelihood $\mathcal{L}_{jt}$. Define the vector of GAS factors as the time-varying mixture probabilities π_{jt}, which defines a new mixture model

$$\mathcal{L}_t = \sum_{j=1}^{J} \pi_{jt} \mathcal{L}_{jt}. \tag{60}$$

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

$$\pi_{jt} = \frac{e^{f_i}}{1 + \sum_{k=1}^{J-1} e^{f_k}} \Leftrightarrow f_{jt} = \ln(\pi_{jt}) - \ln\left(1 - \sum_{k=1}^{J-1} \pi_{kt}\right), \tag{61}$$

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 $f_{j,t-1}$, we obtain

the elements of the score vector

$$\frac{\partial \mathcal{L}_t}{\partial f_{j,t-1}} = \frac{\pi_{j,t-1}\mathcal{L}_{jt}}{\sum_{k=1}^{J} \pi_{k,t-1}\mathcal{L}_{kt}} - \pi_{j,t-1}, \tag{62}$$

for $j = 1, \ldots, J-1$. The interpretation of (62) is intuitive. The probability of model j is increased if the relative likelihood of model j is above its expectation $\pi_{j,t-1}$. 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 $\phi_j(y)$ for $j = 1, 2$ implying an information matrix of the form

$$E_{t-1}[\nabla_t \nabla_t'] = \pi_{1,t}(1-\pi_{1,t}) E_{t-1}\left[\left(\frac{\phi_1(y) - \phi_2(y)}{\pi_{1,t}\phi_1(y) + (1-\pi_{1,t})\phi_2(y)}\right)^2\right],$$

where the expectation is taken with respect to the mixture distribution. We use numerical integration to compute the information matrix, which is feasible when the mixture model (60) 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 1947(2) to 2008(2) obtained from the Federal Reserve Bank of St. Louis. The GAS model is a mixture of two normals with different means μ_i for $i = 1, 2$ and a common variance σ^2. The GAS factor is the probability that the data comes from the normal distribution with low mean indicating the probability of a recession. The GAS(1,1) updating equation is adopted with an information smoothed scaling matrix S_t as in (15) with $\alpha = 0.05$. This GAS model provides an observation driven alternative to a hidden Markov model (HMM). We compare it to a simplied version of the model in Hamilton (1989) without autoregressive dynamics, that is

$$\begin{aligned}
y_t &= \mu_t + \varepsilon_t, \qquad \varepsilon_t \sim \mathcal{N}(0, \sigma^2), \\
\mu_t &= \begin{cases} \mu_1 & \text{if } S_t = 0 \\ \mu_2 & \text{if } S_t = 1 \end{cases} \\
p_{ij} &= P(S_t = j | S_{t-1} = i), \qquad i = 0, 1 \qquad j = 0, 1
\end{aligned}$$

In this model, the latent variable S_t is a regime-switching variable indicating whether the economy is in a recession or expansion. We base our comparison on the one-step ahead predicted estimates produced by the hidden Markov model because the GAS factor is effectively a one-step ahead predictor.

Estimates of the parameters of both models are reported in Table 6. The estimated values for each mean are reasonably close. The recession parameter μ_1 for the HMM model is slightly smaller and negative. Panel (i) of Figure 11 presents the growth rate of log U.S. real GDP along

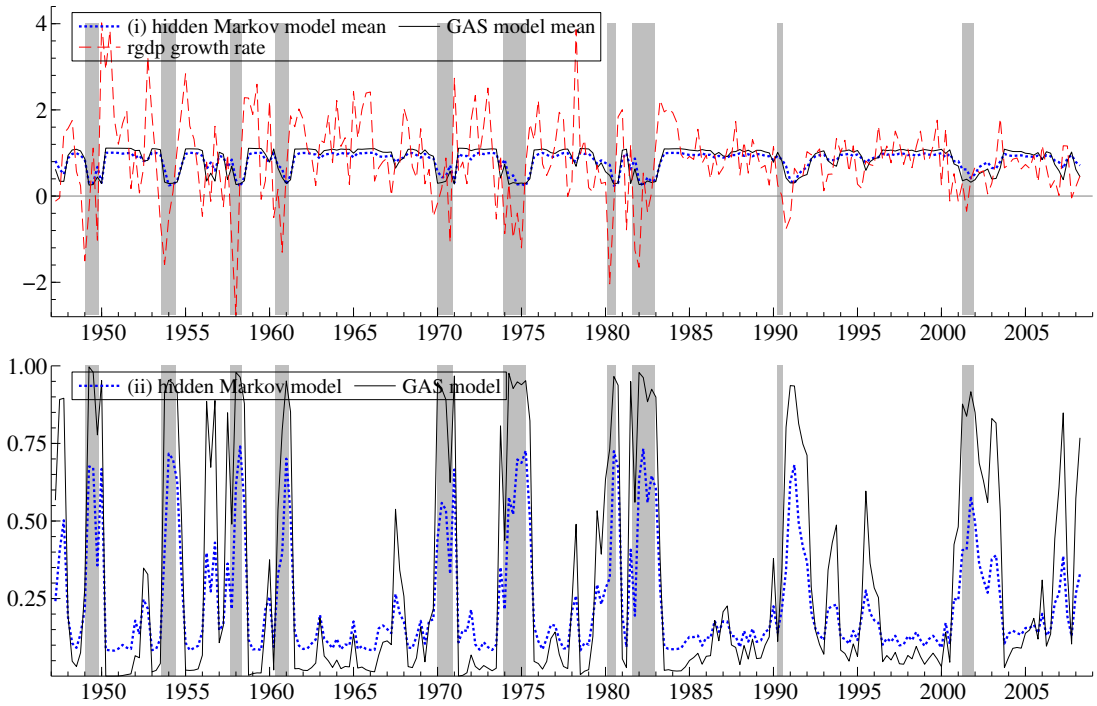

Figure 11: *Mixture model illustration: (i) growth rate of log U.S. real GDP from 1947(2)-2008(2) and the estimated conditional mean from the GAS(1,1) model and the hidden Markov model; (ii) one-step ahead predicted probability of a recession from each model. NBER recession dates are represented by the shaded regions.*

Table 6: *Estimates 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 parenthesis.*

	μ_1	μ_2	σ	ω	A	B	log-like
GAS	0.208	1.127	0.869	0.360	2.333	0.672	-329.70
	(0.008)	(0.005)	(0.003)	(0.017)	(0.113)	(0.006)	
	μ_1	μ_2	σ	p_{11}	p_{22}	-	
HMM	-0.090	1.106	0.830	0.741	0.918		-333.17
	(0.019)	(0.007)	(0.003)	(0.007)	(0.003)		

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 11. 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. In contrast, the one-step ahead predicted probabilities produced by the hidden Markov model do not change as rapidly and are not as clear. The GAS model offers a convenient method for forecasting economic downturns. A multivariate model incorporating leading economic variables would be an interesting extension of the GAS model presented here.

5 Simulation experiments

In this section, we provide simulation evidence on the statistical properties of the GAS ML estimators for a selected set of three examples. We concentrate on the marked point process model, the linear state space model with time varying variances, and the Gaussian copula model with time-varying correlation.

5.1 The pooled marked point process model

To investigate the statistical properties of the GAS model for the marked point processes of Section 4.2, we consider a simplified version of this model. We consider a cross-section of firms with two possible ratings, R_1 and R_2, and possible transitions between them. Neither of the states are absorbing so that no attrition of the panel of firms over time takes place. We consider panel sizes of $N = 250$ and $N = 2,500$ firms. Since the simulation results for both panel sizes are similar, we only present the graphs for $N = 2,500$.

The Monte Carlo study is based on the log intensity equation (25) and the GAS update equation which in our case are given by

$$\lambda_{1t} = \eta_1 + f_t, \qquad \lambda_{2t} = \eta_2 + \alpha f_t, \qquad f_t = As_t + Bf_{t-1},$$

where s_t is given by (27). The intensities λ_{1t} and λ_{2t} are for a R_1 firm becoming a R_2 firm and for a R_2 firm becoming a R_1 firm, respectively. The Monte Carlo data generation process is based on the parameter values $\eta_1 = -3.5$, $\eta_2 = -4.0$, $\alpha = -1$, $A = 0.025$ and $B = 0.95$. The parameter values are roughly in line with the empirical estimates for the levels of intensities and the magnitude of the systematic factor as reported in Table 2.

We consider the sample sizes $T = 20, 50, 100$ for the time series dimension in our data simulations. We generate 1,000 data sets for the Monte Carlo study. For each simulated data set, we compute the ML estimates as well as their t-values based on the numerical second derivative of the likelihood at the optimum. As in the empirical application, we enforce stationarity by parameterizing and estimating the logit transform of B in the GAS equation.

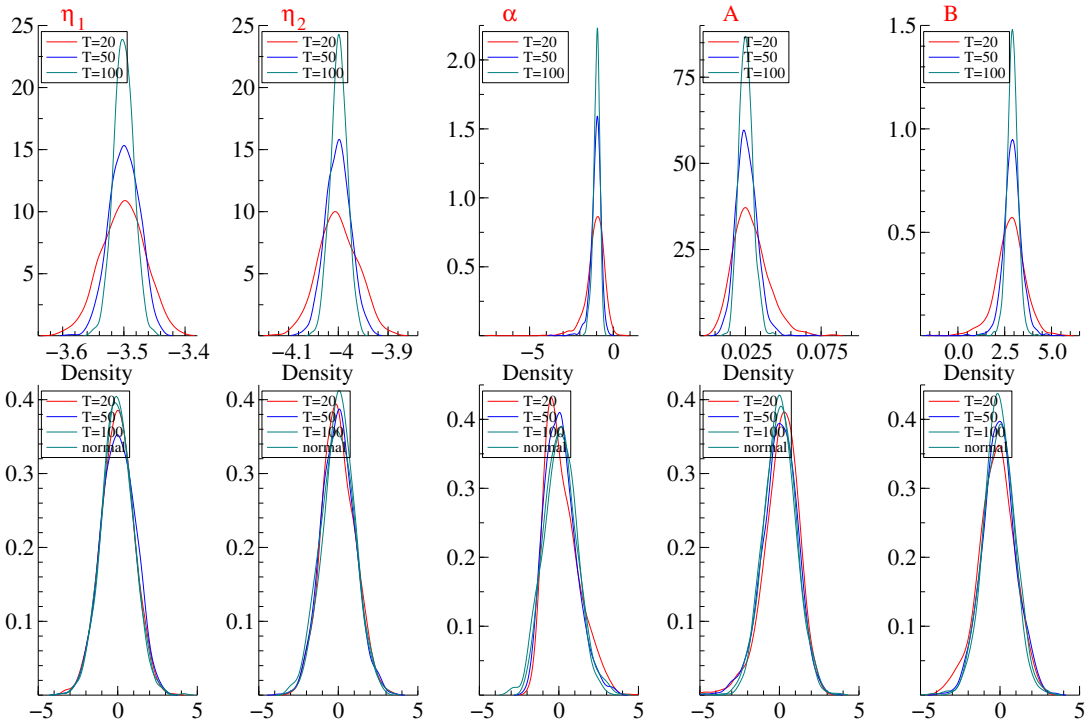

Figure 12: *Simulation densities over 1,000 simulations of a marked point process model. The top panel contains the densities of the parameter estimates, the bottom panels contain the densities of t-values computed using the inverted second derivative of the Hessian at the optimum.*

The Monte Carlo results are graphically presented in Figure 12. The densities of the parameter estimates reveal that for increasing sample sizes T, the estimates peak more at their true values. There is some skewness in the densities for the estimates of α and A, particularly for smaller sample sizes. If we consider the t-values, however, it appears that the approximation by the normal distribution for purposes of inference is reasonable, even for sample sizes as small as $T = 20$.

5.2 State space models with time-varying variances

The finite sample properties are also investigated for the UC model with time-varying variances, see Section 4.4 for the details of this model. In particular, the local level model (33) is adopted where we treat the log-variances of the irregular and the level disturbances, $\ln \sigma_\varepsilon^2$ and $\ln \sigma_\xi^2$ respectively, as GAS factors. The model for the two log-variance factors is given by $f_t = \omega + As_{t-1} + Bf_{t-1}$ where f_t is a 2×1 vector and the 2×2 matrices A and B are both diagonal. The true parameter values in the Monte Carlo study below are chosen to be close to those

obtained in the illustration of Section 4.4. To enforce a stable behaviour of the factors over time, the GAS parameters $\omega = (\omega_1, \omega_2)'$, $A = \text{diag}(A_1, A_2)$ and $B = \text{diag}(B_1, B_2)$ are subject to transformations. In the Monte Carlo simulations for generating data samples, we have adopted the following values for the coefficients:

$$\ln \omega_1 = -4.5, \quad \ln \omega_2 = 5.0, \quad \text{logit}\, A_1 = 0.1, \quad \text{logit}\, A_2 = 1.0,$$

$$\text{logit}\, B_1 = 3.5, \quad \text{logit}\, B_2 = 3.2, \quad \text{logit}\, \alpha = -1.0,$$

where α is the parameter of the EWMA smoothing scheme for the information matrix, see the discussion in Section 2.1. For each simulated series, the coefficients are estimated using the methods described in Section 4.4.

The Monte Carlo design is similar to the one presented in the previous section. The number of simulations equals $1,000$ while the time series dimension is set to $T = 100$, 250 and 1000. Smaller values for T are not of interest since the time-variation of the variances of both ε_t and η_t cannot be detected in a small time interval. This is certainly the case when dealing with nonstationary time series. The results of the Monte Carlo study are given as the simulation densities for parameter estimates in Figure 13 and as the corresponding densities for the t-values in Figure 14.

It is clear that the GAS parameter estimates for our current model vary much more compared to those obtained in the previous subsection for the marked point process model. It provides evidence of the difficulty in empirically identifying the correct parameters that control the time-variation of the variances in the nonstationary local level model. Given the flexibility of the local level model, we are encouraged that the modes of the simulation densities are close to the corresponding true values of the parameters. Although we may need relatively large time series dimensions, the estimation methodology is able to detect the correct location of the parameters in the majority of cases. In particular this applies to the coefficients of the GAS factor for the time-varying variance of η_t (ω_2, A_2 and B_2) and to the information matrix smoothing parameter α. The latter is surprising given its peripheral role in our estimation framework. We may expect even more encouraging results when we restrict ourselves to classes of stationary time series models. The densities for the t-values of the estimated parameters are presented in Figure 14 and they confirm that better estimation performance is obtained when the sample size T increases.

5.3 Time-varying Gaussian copula model

In our final simulation study, we focus on the finite sample properties of the time-varying Gaussian copula model described in Section 4.5. We consider the model specification in (45)

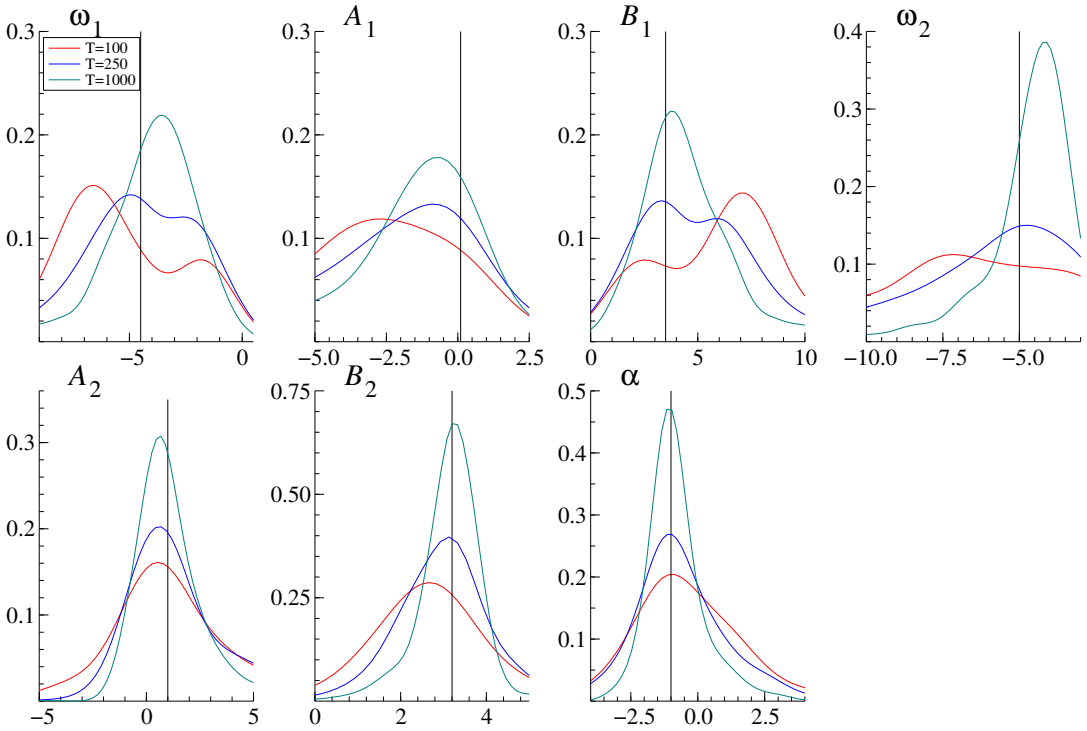

Figure 13: *Simulation densities for the estimated GAS parameters over 1,000 simulations of a local level model with time-varying log variances.*

with a GAS$(1,1)$ factor. The parameter settings for the model that generate the Monte Carlo data-sets are given by $\omega = 0.02$, $A = 0.15$, and $B = 0.96$. The simulation sample sizes are $T = 200, 400, 600$. To ensure stationarity of the factor f_t and for numerical stability, we carry out logit transformations for both A and B.

The results from the Monte Carlo experiment using $1,000$ simulations are presented in Figure 15. The density of the parameter estimates are converging toward their true values as T increases. The rate of convergence appears to be slower for this model than for the marked point process model in Subsection 5.1. The densities of the t-values appear slightly biased for the ω and B parameters. However, the bias diminishes as the sample size increases.

6 Conclusions

We have introduced Generalized Autoregressive Score (GAS) models. A GAS model is a uniformly applicable observation driven model specification to capture time variation in parameters. We have shown how GAS models encompass other well-known models, such as generalized autoregressive conditional heteroskedasticty models and autoregressive conditional

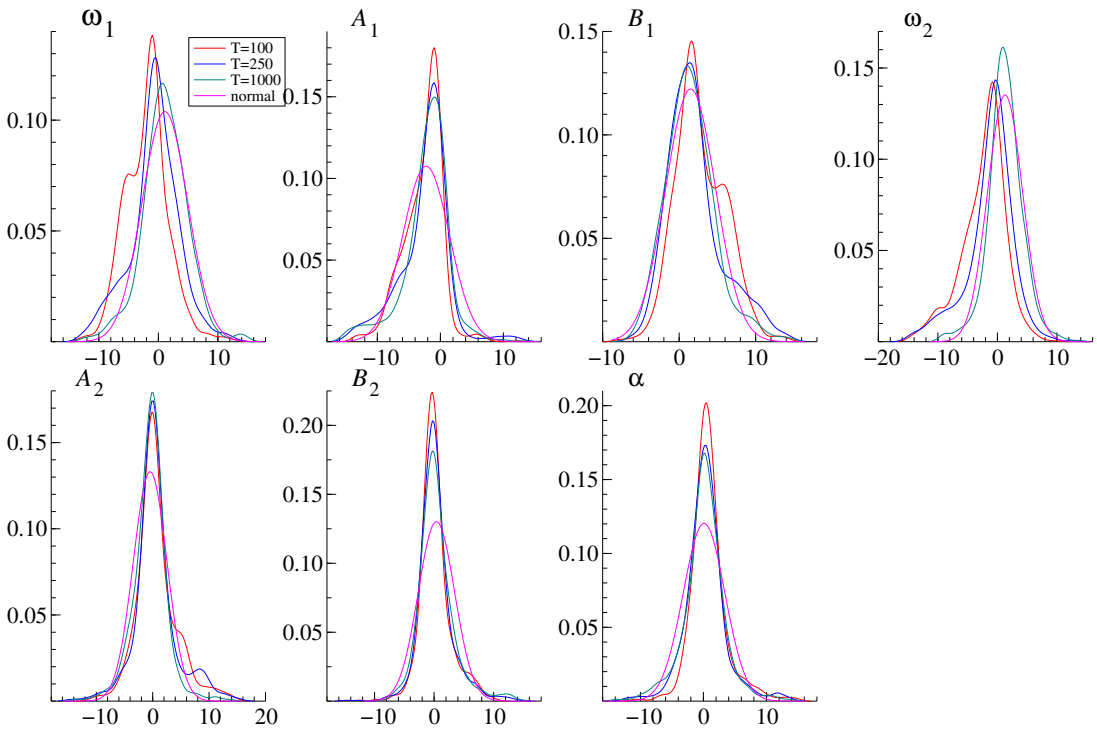

Figure 14: *Simulation densities for the t-values of the estimated GAS parameters over 1,000 simulations of a local level model with time-varying log variances. The t-values are computed using the inverted second derivative of the Hessian at the optimum.*

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.

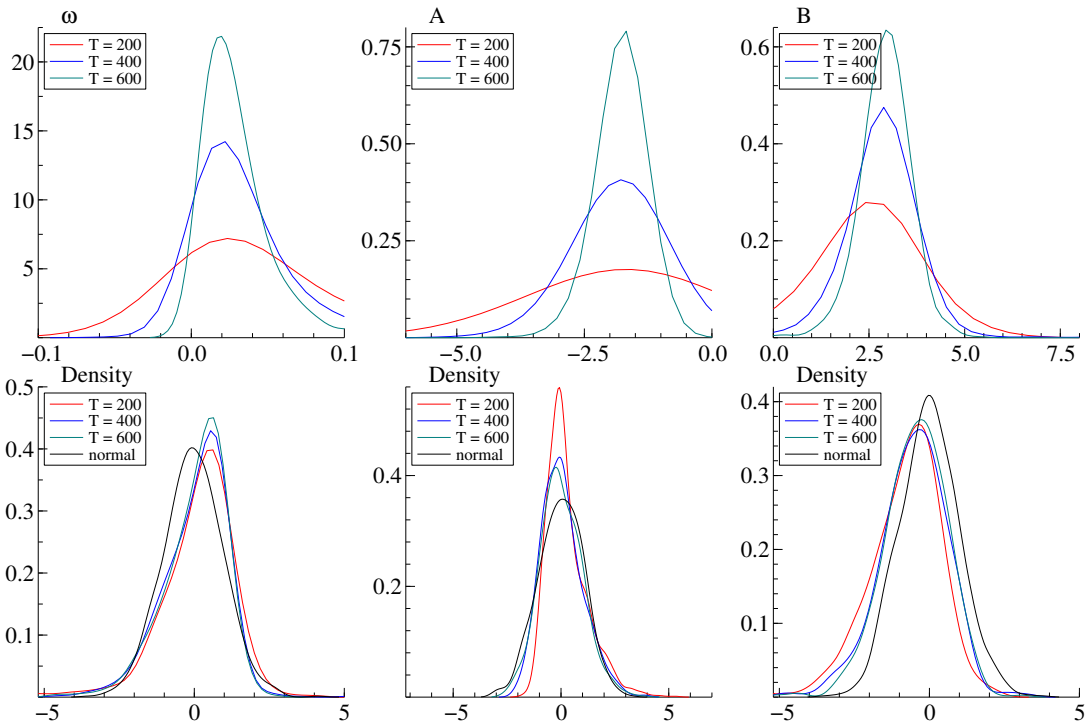

Figure 15: *Simulation densities over 1,000 simulations of a time-varying Gaussian copula model. The top panel contains the densities of the parameter estimates, the bottom panels contain the densities of t-values computed using the inverted second derivative of the Hessian at the optimum.*

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. To investigate the 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 basis for dynamic misspecification tests. A similar approach to test for the presence of possible ARCH effects is already 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 large 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 (1996) 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. We anticipate that the posterior densities for the latent GAS factors f_t are straightforward to formulate. Simulations for the latent factors can easily be generated. The coefficients of the GAS model can then be estimated as part of a Markov chain Monte Carlo analysis.

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 Subsection 4.5, 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

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

Benjamin, M. A., R. A. Rigby, and M. Stanispoulos (2003). Generalized autoregressive moving average models. *Journal of the American Statistical Association 98*(461), 214–223.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics 31*, 307–327.

Bollerslev, T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. *The Review of Economics and Statistics 69*(3), 542–547.

Brooks, C., S. P. Burke, S. Heravi, and G. Persand (2005). Autoregressive conditional kurtosis. *Journal of Financial Econometrics 3*(3), 399–421.

Cipollini, F., R. F. Engle, and G. M. Gallo (2006). Vector multiplicative error models: representation and inference. *NBER Working Paper*.

Clark, P. K. (1989). Trend reversion in real output and unemployment. *Journal of Econometrics 40*, 15–32.

Cox, D. R. (1958). The regression analysis of binary sequences (with discussion). *Journal of the Royal Statistical Society, Series B 20*(2), 215–242.

Cox, D. R. (1981). Statistical analysis of time series: some recent developments. *Scandinavian Journal of Statistics 8*, 93–115.

Davis, R. A., W. T. M. Dunsmuir, and S. Streett (2003). Observation driven models for Poisson counts. *Biometrika 90*(4), 777–790.

Dias, A. and P. Embrechts (2004). Dynamic copula models for multivariate high-frequency data in finance. *Manuscript, ETH Zurich*.

Diebold, F. X. and C. Li (2006). Forecasting the term structure of government bond yields. *Journal of Econometrics 130*, 337–364.

Diebold, F. X., G. Rudebusch, and S. Aruoba (2006). The macroeconomy and the yield curve. *Journal of Econometrics 131*, 309–338.

Doucet, A., N. de Freitas, and N. Gordon (2001). *Sequential Monte Carlo Methods in Practice*. New York, NY: Springer-Verlag.

Duffie, D., A. Eckner, G. Horel, and L. Saita (2006). Frailty correlated default. *manuscript Stanford University*.

Durbin, J. and S. J. Koopman (2001). *Time Series Analysis by State Space Methods*. Oxford: Oxford University Press.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica 50*(4), 987–1007.

Engle, R. F. (2002a). Dynamic conditional correlation: a simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics 20*(3), 339–350.

Engle, R. F. (2002b). New frontiers for ARCH models. *Journal of Applied Econometrics 17*, 425–446.

Engle, R. F. and T. Bollerslev (1986). Modelling the persistence of conditional variances. *Econometric Reviews 5*(1), 1–50.

Engle, R. F. and G. M. Gallo (2006). A multiple indicators model for volatility using intra-daily data. *Journal of Econometrics 131*, 3–27.

Engle, R. F. and S. Manganelli (2004). CAViaR: conditional autoregressive value at risk by regression quantiles. *Journal of Business & Economic Statistics 22*(4), 367–382.

Engle, R. F. and J. R. Russell (1998). Autoregressive conditional duration: a new model for irregularly spaced transaction data. *Econometrica 66*(5), 1127–1162.

Fiorentini, G., G. Calzolari, and L. Panattoni (1996). Analytic derivatives and the computation of GARCH estimates. *Journal of Applied Econometrics 11*, 399–417.

Hamilton, J. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica 57*(2), 357–384.

Hansen, B. E. (1994). Autoregressive conditional density estimation. *International Economic Review 35*(3), 705–730.

Harvey, A. C. (1989). *Forecasting, structural time series models and the Kalman filter*. Cambridge, UK: Cambridge University Press.

Harvey, A. C. and A. Jaeger (1993). Detrending, stylised facts and the business cycle. *Journal of Applied Econometrics 8*, 231–247.

Harvey, A. C., E. Ruiz, and E. Sentana (1992). Unobserved component time series models with ARCH disturbances. *Journal of Econometrics 52*(1-2), 129–157.

Harvey, C. R. and A. Siddique (1999). Autoregressive conditional skewness. *Journal of Financial and Quantitative Analysis 34*, 465–487.

Hosking, J. R. M. (1981). Fractional differencing. *Biometrika 68*(1), 165–176.

Jondeau, E. and M. Rockinger (2003). Conditional volatility, skewness, and kurtosis: existence, persistence, and comovements. *Journal of Economic Dynamics and Control 27*(10), 1699–1737.

Koopman, S. J., A. Lucas, and A. Monteiro (2008). The multi-state latent factor intensity model for credit rating transitions. *Journal of Econometrics 142*, 399–424.

Koopman, S. J., A. Lucas, and B. Schwaab (2008). Forecasting cross-sections of frailty-correlated default. *08-029/4*.

Koopman, S. J., M. I. Mallee, and M. van der Wel (2009). Analyzing the term structure of interest rates using the dynamic Nelson-Siegel model with time varying parameters. *Journal of Business & Economic Statistics 27*, forthcoming.

Lehmann, E. L. and G. Casella (1998). *Theory of Point Estimation*. New York, NY: Springer Press.

Nelson, C. R. and A. Siegel (1987). Parsimonious modelling of yield curves. *Journal of Business 60*(4), 473–489.

Nelson, D. B. (1996). ARCH Models as diffusion approximations. *Modelling Stock Market Volatility: Bridging the Gap to Continuous Time*.

Ord, J. K., A. B. Koehler, and R. D. Snyder (1997). Estimation and prediction for a class of dynamic nonlinear statistical models. *Journal of the American Statistical Association 92*(440).

Patton, A. J. (2006). Modelling asymmetric exchange rate dependence. *International Economic Review 47*(2), 527–556.

Russell, J. R. (2001). Econometric modeling of multivariate irregularly-spaced high-frequency data. *Unpublished manuscript, University of Chicago, Graduate School of Business*.

Russell, J. R. and R. F. Engle (2005). A discrete-state continuous-time model of financial transactions prices and times: the autoregressive conditional multinomial-autoregressive conditional duration model. *Journal of Business & Economic Statistics 23*(2), 166–180.

Rydberg, T. H. and N. Shephard (2003). Dynamics of trade-by-trade price movements: decomposition and models. *Journal of Financial Econometrics 1*, 2–25.

Shephard, N. (1995). Generalized linear autoregressions. *Unpublished manuscript, Nuffield College, University of Oxford*.

Shephard, N. (2005). *Stochastic Volatility: Selected Readings*. Oxford: Oxford University Press.

Shephard, N. and M. K. Pitt (1997). Likelihood analysis of non-Gaussian measurement time series. *Biometrika 84*, 653–67.

Stock, J. H. and M. W. Watson (2007). Why Has U.S. Inflation Become Harder to Forecast? *Journal of Money, Credit, and Banking 39*, 3–33.

van den Goorbergh, R. W. J., C. Genest, and B. J. M. Werker (2005). Bivariate option pricing using dynamic copula models. *Insurance Mathematics and Economics 37*(1), 101–114.

Watson, M. W. (1986). Univariate detrending methods and stochastic trends. *Journal of Monetary Economics 18*, 49–75.