Generalized Autoregressive Score Models

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

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

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

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

(d) *Tinbergen Institute, Amsterdam*

March 1, 2012

1 Introduction

To capture the dynamic behavior of univariate and multivariate time series processes, we can allow parameters to be time-varying by having them as functions of lagged dependent variables as well as exogenous variables. Although other approaches of introducing time dependence exists, this particular approach have become popular in applied statistics and econometrics. Typical examples are the generalized autoregressive conditional heteroskedasticity (GARCH) models of Engle (1982) and Bollerslev (1986), the autoregressive conditional duration and intensity (ACD and ACI, respectively) models of Engle and Russell (1998) and the dynamic copula models of Patton (2006). Here we discuss a further development of Creal, Koopman, and Lucas (2012) which is based on the score function of the predictive model density at time t. They argue that the score function is an effective choice for introducing a driving mechanism for time-varying parameters. In particular, by scaling the score function appropriately, standard dynamic models such as the GARCH, ACD, and ACI models can be recovered. Application of this framework to other non-linear, non-Gaussian, possibly multivariate, models will lead to the formulation of new time-varying parameter models. They have labeled their model as the generalized autoregressive score (GAS) model. Here we aim to introduce the GAS model and to illustrate the approach for a class of multivariate point-process models that is used empirically for the modeling credit risk. We further aim to show that time-varying parameters in a multi-state model for pooled marked point-processes can be introduced naturally in our framework.

2 The GAS model

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

$$\mathcal{F}_t = \{Y^{t-1} \, , F^{t-1} \, , X^t\}, \qquad \text{for } t = 1, \ldots, n.$$

We assume that y_t is generated by the observation density

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

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

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

where ω is a vector of constants, coefficient matrices A_i and B_j have appropriate dimensions for $i = 1, \ldots, p$ and $j = 1, \ldots, q$, while s_t is an appropriate function of past data, $s_t = s_t(y_t, f_t, \mathcal{F}_t; \theta)$. The unknown coefficients in (2) are functions of θ, that is $\omega = \omega(\theta)$, $A_i = A_i(\theta)$, and $B_j = B_j(\theta)$ for $i = 1, \ldots, p$ and $j = 1, \ldots, q$.

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

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

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

The use of the score for updating f_t is intuitive. It defines a steepest ascent direction for improving the model's local fit in terms of the likelihood or density at time t given the current position of the parameter f_t. This provides the natural direction for updating the parameter. In addition, the score depends on the complete density, and not only on the first or second order moments of the observations y_t. Via its choice of the scaling matrix S_t, the GAS model allows for additional flexibility in how the score is used for updating f_t. In many situations, it is natural to consider a form of scaling that depends on the variance of the score. For example, we can define the scaling matrix as

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

where E_{t-1} is expectation with respect to the density $p(y_t|f_t, \mathcal{F}_t; \theta)$. For this choice of S_t, the GAS model encompasses well-known models such as GARCH, ACD and ACI. Another possibility for scaling is

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

where S_t is defined as the square root matrix of the (pseudo)-inverse information matrix for (1) with respect to f_t. An advantage of this specific choice for S_t is that the statistical properties of the corresponding GAS model become more tractable. In particular, the driver s_t becomes a martingale difference with unity variance.

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

$$\hat{\theta} = \arg\max_{\theta} \sum_{t=1}^{n} \ell_t, \tag{6}$$

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

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

$$f_{t+1} = \omega + A_1\left(y_t^2 - f_t\right) + B_1 f_t, \tag{7}$$

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

$$f_{t+1} = \alpha_0 + \alpha_1 y_t^2 + \beta_1 f_t, \qquad f_t = \sigma_t^2, \tag{8}$$

where coefficients $\alpha_0 = \omega$, $\alpha_1 = A_1$ and $\beta_1 = B_1 - A_1$ are unknown. When we assume that ε_t follows a Student's t distribution with ν degrees of freedom and unit variance, the GAS $(1, 1)$

specification for the conditional variance leads to the updating equation

$$f_{t+1} = \omega + A_1 \cdot \left(1 + 3\nu^{-1}\right) \cdot \left(\frac{(1+\nu^{-1})}{(1-2\nu^{-1})(1+\nu^{-1} y_t^2 \;/ (1-2\nu^{-1})\; f_t)} y_t^2 - f_t \right) + B_1 f_t. \qquad (9)$$

This model is clearly different compatered to the standard t-GARCH$(1,1)$ model which has the Student's t density in (1) with the updating equation (7). The denominator of the second term in the right-hand side of (9) causes a more moderate increase in the variance for a large realization of $|y_t|$ as long as ν is finite. The intuition is clear: if the errors are modeled by a fat-tailed distribution, a large absolute realization of y_t does not necessitate a substantial increase in the variance. Multivariate extensions of this approach are developed in Creal, Koopman, and Lucas (2011).

Example : Regression model The time-varying linear regression model $y_t = x_t'\beta_t + \varepsilon_t$ has a $k \times 1$ vector x_t of exogenous variables, a $k \times 1$ vector of time-varying regression coefficients β_t and normally independently distributed disturbances $\varepsilon_t \sim \mathrm{N}(0, \sigma^2)$. Let $f_t = \beta_t$. The scaled score function based on $S_t = \mathcal{J}_{t|t-1}$ in for this regression model is given by

$$s_t = (x_t' x_t)^{-1/2} x_t (y_t - x_t' f_t)/\sigma, \qquad (10)$$

where the inverse of $\mathcal{I}_{t|t-1}$ used to construct $\mathcal{J}_{t|t-1}$ is 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

$$f_{t+1} = \omega + A_1 \frac{x_t}{(x_t' x_t)^{1/2}} \cdot \frac{(y_t - x_t' f_t)}{\sigma} + B_1 f_t. \qquad (11)$$

The updating equation (11) can be extended by including σ^2 as a time-varying factor and by adjusting the scaled score function (10) for the time-varying parameter vector $f_t = (\beta_t' \;,\; \sigma_t^2)'$.

3 Illustration: dynamic pooled marked point process

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

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

$$\lambda_{k,t} = d + Z f_t + X_{k,t}\beta, \tag{12}$$

where d is a $J \times 1$ vector of baseline intensities, Z is a $J \times r$ matrix of factor loadings, and β is a $p \times 1$ vector of regression parameters for the exogenous covariates $X_{k,t}$. The $r \times 1$ vector of dynamic factors f_t is specified by the GAS $(1, 1)$ updating equation (2) with $\omega = 0$. Since f_t is not observed directly, we need to impose a sign restriction on Z to obtain economic interpretations for the time-varying parameters. We assume the model has a factor structure: intensities of all firms are driven by the same vector of time-varying systematic parameters f_t.

The log-likelihood specification using (12) is given by

$$\ell_t = \sum_{j=1}^{J}\sum_{k=1}^{N} y_{jk,t}\lambda_{jk,t} - R_{jk,t} \cdot (t - t^*) \cdot \exp\left(\lambda_{jk,t^*}\right), \tag{13}$$

where $R_{k,t} = (R_{1k,t}, \ldots, R_{Jk,t})'$ and $R_{jk,t}$ is a zero-one variable indicating whether company k is potentially subject to risk j at time t. Define P as a $J \times J$ diagonal matrix with jth diagonal element $p_{j,t} = \sum_k R_{jk,t} \cdot \exp[\lambda_{jk,t}] \;/\; \sum_{j,k} R_{jk,t} \cdot \exp[\lambda_{jk,t}] = \mathrm{P}[\sum_k y_{jk,t} = 1 \mid \sum_{j,k} y_{jk,t} = 1]$, i.e., the probability that the next event is of type j given that an event happens for firm k. Based on the first and second derivative of ℓ_t and setting $S_t = \mathcal{J}_{t|t-1}$, we obtain the score and scaling matrix

$$\nabla_t = Z' \left(\sum_{k=1}^{N} y_{k,t} - R_{k,t} \cdot (t - t^*) \cdot \exp(\lambda_{k,t^*}) \right), \qquad S_t = (Z'PZ)^{-\frac{1}{2}}. \tag{14}$$

By combining these basic elements into a GAS specification, we have obtained a new time-varying parameter model for credit rating transitions. In comparison with related models, parameter estimation for the current model is much easier.

Application to Moody's credit rating data

For our illustration, we adopt the marked point-process model (12), (13) and (2) with $\omega = 0$ and $s_t = S_t \nabla_t$ given by (14) for a data set which contains Moody's rating histories of all US corporates over the period January 1981 to March 2010. The initial credit ratings for each firm are known at the beginning of the sample and we observe the transitions from one rating category to another over time. Moody's ratings include 21 different categories, some of which are sparsely populated. For the sake of this illustration, therefore, we pool the ratings into a much smaller set of 2 credit classes: investment grade (IG) and sub-investment grade (SIG). Default is treated as an absorbing category: it makes for $J = 4$ possible events. It is often concluded in credit risk studies that default probabilities are countercyclical. We therefore allow the log-intensities (12) to depend upon the annual growth rate (standardized) of US industrial production as an exogenous variable. We only present the results for a single factor, $r = 1$. In order to identify the parameters in the 1×4 vector Z, we set its last element to unity so that our single factor is common to all transition types but is identified as the event representing a move from SIG to default.

For the one-factor model ($r = 1$), we perform a full benchmark analysis for the new GAS model in relation to our benchmark model of Koopman et al. (2008), hereafter referred to as KLM08 . The marked point process KLM08 model has the same observation density (13) as the GAS model. However, the time-varying parameter f_t follows an Ornstein-Uhlenbeck process driven by an independent stochastic process. Parameter estimation for the KLM08 model is more involved than for the GAS model due to the presence of a dynamic, non-predictable stochastic component.

Figure 1 compares the estimates of f_t obtained from the two model specifications. For each of the four possible rating transitions, we plot the intensity of the transition (in basis points on a log scale). These intensities, after dividing them by the number of days in a year, can approximately be interpreted as the daily transition probabilities for each rating transition type. We learn from Figure 1 that the estimates of the time-varying probabilities of the GAS model are almost identical to those of the KLM08 model. However, in our current GAS framework, the results can be obtained without the need of computationally intensive simulation methods required for models such as KLM08 . It underlines an attractive feature of our GAS approach.

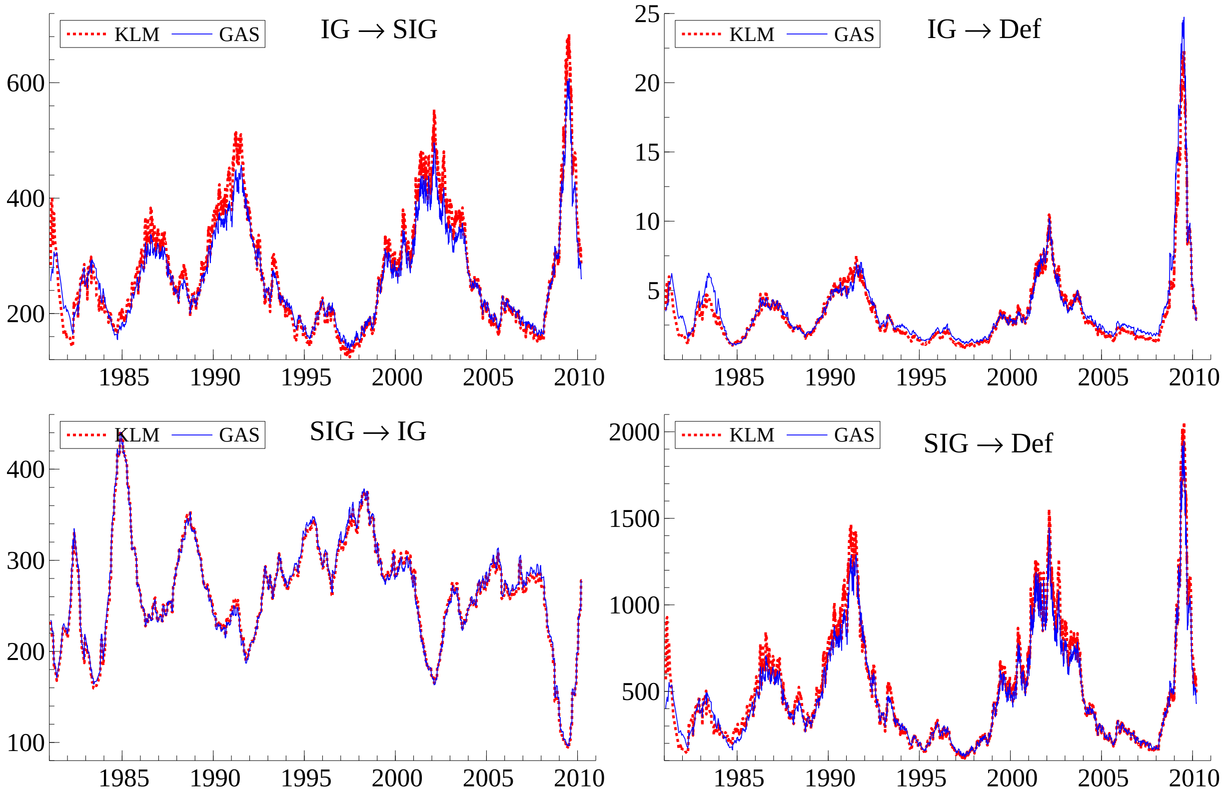

Figure 1: *The estimated intensities (in basis points) for each transition type for the one-factor marked point process model. Moody's rating histories are for all US corporates between January 1981 and March 2010.*

References

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics 31*(3), 307–327.

Creal, D. D., S. J. Koopman, and A. Lucas (2011). A dynamic multivariate heavy-tailed model for time-varying volatilities and correlations. *Journal of Business & Economic Statistics 29*, 552–563.

Creal, D. D., S. J. Koopman, and A. Lucas (2012). Generalized Autoregressive Score Models with Applications. *Journal of Applied Econometrics*. forthcoming.

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

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

Koopman, S. J., A. Lucas, and A. Monteiro (2008). The multi-state latent factor intensity model for credit rating transitions. *Journal of Econometrics 142*(1), 399–424.

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