

Supplementary Appendix for “Generalized Autoregressive Score Models with Applications”

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

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

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

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

^(d) *Tinbergen Institute, Amsterdam*

August 8, 2011

Abstract

In this Supplementary Appendix we present additional new material related to the main paper “Generalized Autoregressive Score Models with Applications”. We refer to the model as the GAS model. For reference purposes, we first give a short review of the relevant equations for the general GAS model. Appendix A presents more existing models that can be represented as special cases of GAS models. Appendix B formulates new models including unobserved components models, models with time-varying higher order moments, time-varying multinomial model and dynamic mixture models. In Appendix C we present the simulation results for the two illustration models of the main paper : the Gaussian copula model with time-varying correlations and the marked point process model.

Basic GAS model specification

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

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

We assume that y_t is generated by the observation density

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

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

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

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

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

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

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

Each different choice for the scaling matrix S_t results in a different GAS model. In many

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

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

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 the well-known observation driven GARCH model of Engle (1982) and Bollerslev (1986), the ACD model of Engle and Russell (1998), and the ACI model of Russell (2001) as well as most of the Poisson count models considered by Davis et al. (2003). Another possibility is the GAS model with scaling matrix

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

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

Appendix A : more special cases of GAS models

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 $\varepsilon_t \sim N(0, \sigma^2)$. 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

$$s_t = (x_t' x_t)^{-1} x_t (y_t - x_t' f_{t-1}), \quad (6)$$

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

$$f_t = \omega + A_0 (x_t' x_t)^{-1} x_t (y_t - x_t' f_{t-1}) + B_1 f_{t-1}. \quad (7)$$

In case $x_t \equiv 1$, the updating equation (7) for the time-varying intercept reduces to the exponentially weighted moving average (EWMA) recursion by setting $\omega = 0$ and $B_1 = 1$, that is

$$f_t = f_{t-1} + A_0 (y_t - f_{t-1}). \quad (8)$$

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

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

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 (8) 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 (8) can be extended by including σ^2 as a time-varying factor and recomputing the scaled score function in (6) for the new time-varying parameter vector $f_{t-1} = (\beta_{t-1}', \sigma_{t-1}^2)'$.

The GAS updating function (7) 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 $s_t = x_t(y_t - x_t'f_{t-1})$.

Dynamic exponential family models

Consider the exponential family of distributions represented by

$$\exp(\eta(\theta)'T(y_t) - C(\theta) + h(y_t)), \quad (9)$$

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

$$E_{t-1}[\dot{\eta}'T(y_t)] = \dot{C}, \quad (10)$$

and

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

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

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

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, Rigby, and Stanispoulos (2003) proposed observation-driven models for the subclass of natural exponential family members when $\eta(\theta)'T(y_t) = \theta'y_t$ in (9). Expression (10) then reduces to $E_{t-1}[y_t] = \partial C/\partial\eta = g(f_{t-1}, Y_1^{t-1}, X_1^t, F_1^{t-2})$ where $g(\cdot)$ is known as the link function. They then model the link function using explanatory variables and autoregressive/moving av-

erage 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.

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

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

The GAS model specification is given by the equations (1) and (2). We have defined ∇_t in (3) and \mathcal{I}_t in (4). The (i, j) element of \mathcal{I}_t is denoted by $\mathcal{I}_{t,ij}$. We further note that $\Psi(x, k) = \partial^k \ln \Gamma(x) / \partial x^k$.

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, Dunsmuir, and Streett (2003). The latter three models can be obtained by scaling the relevant score vector from Table 1 with an identity scaling matrix, $S_{t-1} = I$ or the matrix square root of $S_{t-1} = \mathcal{I}_{t-1}^{-1}$.

Appendix B : more new GAS model formulations

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, \dots, 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

$$\psi_t = f_{1,t-1} + \dots + f_{r,t-1} \quad (11)$$

with $f_t = (f_{1,t}, \dots, f_{r,t})'$. 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 $p(y_t|\psi_t; \theta) = N(f_{1,t-1} + f_{2,t-1}, \sigma^2)$ and updating equation

$$f_t = \begin{pmatrix} \omega \\ 0 \end{pmatrix} + \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} s_t + \begin{bmatrix} 1 & 0 \\ 0 & \phi_1 \end{bmatrix} f_{t-1} + \begin{bmatrix} 0 & 0 \\ 0 & \phi_2 \end{bmatrix} f_{t-2}. \quad (12)$$

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

$$s_t = y_t - \psi_t = y_t - f_{1,t-1} - f_{2,t-1} = \varepsilon_t, \quad (13)$$

and can be interpreted as the single source of error. The static parameter vector θ , consisting of coefficients $\omega, a_1, a_2, \phi_1, \phi_2$ 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.

Table 2: *Estimation results for the parameters in the trend-cycle GAS(1,2) decomposition model (11) with the updating equation (12) and the scaled scoring function (13) 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 (14)–(15) 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 (0.043)	0.723 (0.206)	0.563 (0.202)	1.328 (0.130)	-0.424 (0.142)	0.905 (0.041)	-324.51
UC	0.825 (0.040)	0.604 (0.098)	0.621 (0.112)	1.501 (0.102)	-0.573 (0.106)	– –	-324.06

The UC trend-cycle decomposition model is then given by $y_t = f_{1,t} + f_{2,t}$ with

$$f_{1,t} = \omega + f_{1,t-1} + a_1 \xi_{1,t}, \quad \xi_{1,t} \sim N(0, 1), \quad (14)$$

$$f_{2,t} = \phi_1 f_{2,t-1} + \phi_2 f_{2,t-2} + a_2 \xi_{2,t}, \quad \xi_{2,t} \sim N(0, 1), \quad (15)$$

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 2. 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 1. 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 2 and the one-step ahead predicted estimate of $f_{2,t}$ is plotted in Figure 1. 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 1 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

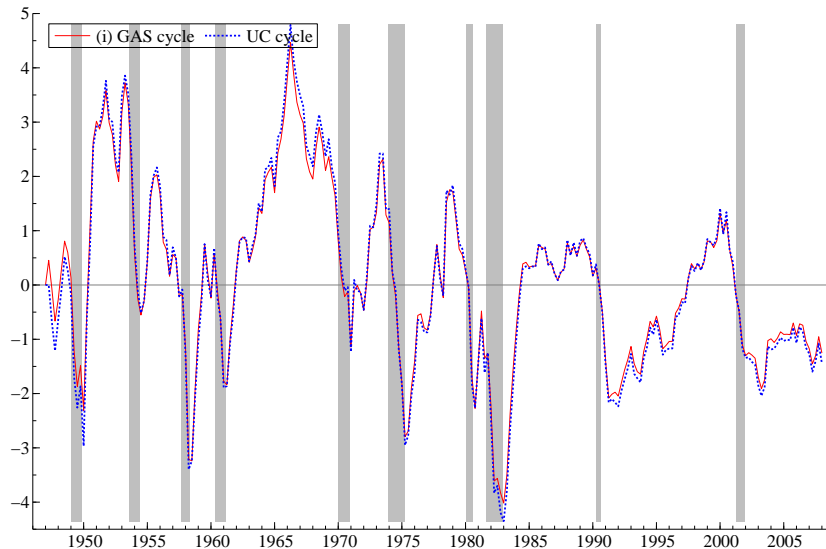


Figure 1: *Trend-cycle illustration: estimated cycles from the GAS and UC trend-cycle models based on quarterly log of U.S. real gdp from 1947(1) through 2008(1). NBER recession dates are indicated by the shaded regions.*

Harvey and Jaeger (1993). Regression and intervention effects can also be incorporated in the GAS specification. Since the resulting GAS models are equivalent to single source of error models, we refer to Ord, Koehler, and Snyder (1997) for a more detailed discussion on this class of models.

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, Burke, Heravi, and Persaud (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 $\varepsilon_t \sim t_{\nu_t}$. The error term is scaled to have unit variance such that σ_{t-1}^2 is the conditional variance while ν_t is the time-varying degrees of freedom parameter. Define the vector of factors as $f_t = (\sigma_t^2, -\ln \left\{ \frac{b-a}{\nu_t-a} - 1 \right\})$ 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_t^2 and ν_t , we obtain the score vector as given by

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

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

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

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. We label this model the tv-t-GAS(1,1) model.

We consider daily returns on the S&P 500 from February 1989 through April 2008 as an illustration. We compare the tv-t-GAS(1,1) model described above to a t-GAS(1,1) model with constant ν 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 3 and estimates of the conditional variance are plotted in panel (i) of Figure 2. 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 2. 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

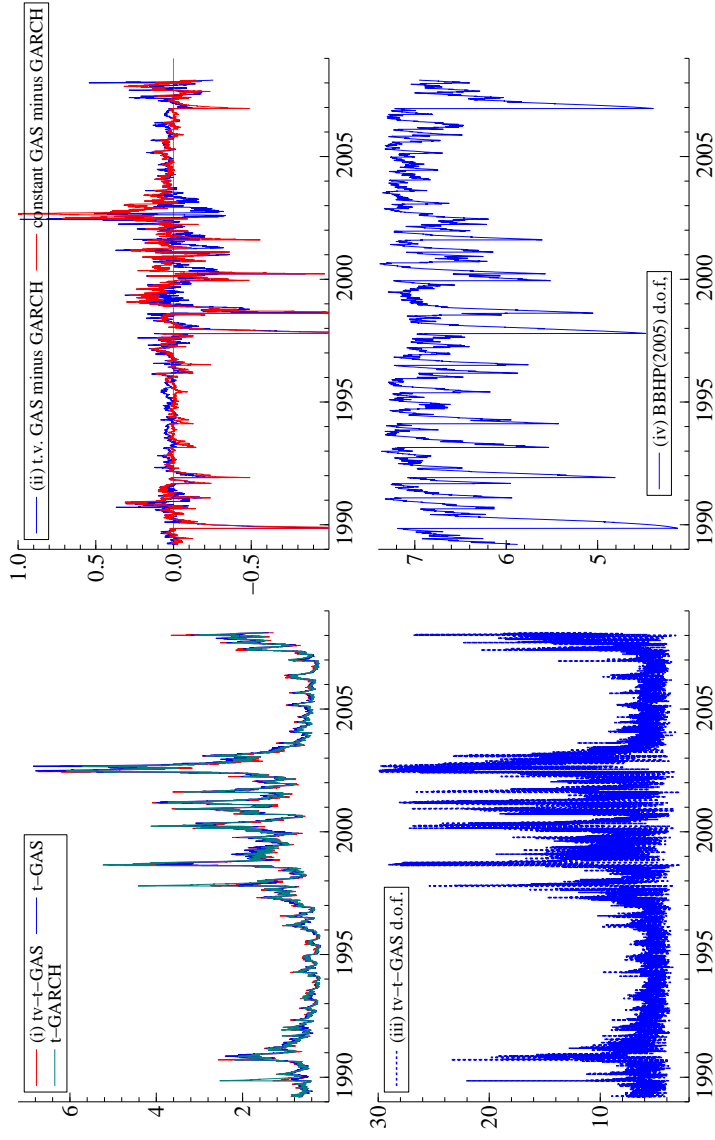


Figure 2: Time-varying degrees of freedom illustration: (i) estimated conditional variances from the t -GAS(1,1), t -GARCH(1,1), and tv - t -GAS(1,1) models; (ii) differences between the two GAS(1,1) models and the t -GARCH(1,1) model; (iii) estimated time-varying degrees of freedom from the tv - t -GAS(1,1) model; (iv) estimated time-varying degrees of freedom from the GARCH model of Brooks et al. (2005).

of freedom is due to the fact that the t-GAS model does not treat outliers like a standard t-GARCH model. From 1998-2003, volatility increases 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.

Table 3: *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 (0.005)	0.003 (0.001)	0.004 (0.001)	0.003 (0.001)	0.002 (0.001)	0.007 (0.003)
ω_2	-2.373 (0.310)	-	-	-	-	-
a_{11}	0.057 (0.007)	0.047 (0.007)	0.044 (0.006)	0.049 (0.007)	0.026 (0.006)	0.061 (0.009)
a_{12}	-0.128 (0.043)	-	-	-	-	-
a_{21}	-0.219 (0.033)	-	-	-	-	-
a_{22}	-1.498 (0.002)	-	-	0.005 (0.006)	-	-
b_{11}	0.994 (0.003)	0.951 (0.007)	0.997 (0.002)	0.949 (0.007)	0.997 (0.003)	0.995 (0.004)
b_{12}	0.000 (0.000)	-	-	-	-	-
b_{21}	0.982 (0.154)	-	-	-	-	-
b_{22}	0.026 (0.121)	-	-	0.965 (0.024)	-	-
ν	-	6.699 (0.622)	7.032 (0.677)	-	5.367 (0.610)	10.96 (2.074)
log-like	-6138.18	-6153.02	-6156.46	-6153.44	-2359.55	-3778.63

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 2 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, Burke, Heravi, and Persaud (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, Burke, Heravi, and Persaud (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 3 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 3. 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 ω_t in the variance. We leave this extension to future research.

Time-varying multinomial model

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 $t = 1, \dots, n$, has a J -dimensional multinomial distribution with vector of probabilities π_t and let $\pi_{j,t}$ be the j th 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 J th element omitted. The score with respect to $f_{j,t-1}$ is given by

$$\nabla_{jt} = \tilde{y}_{jt} - \tilde{\pi}_{j,t-1}, \quad (16)$$

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

$$\mathcal{I}_{ii,t-1} = \tilde{\pi}_{i,t-1}(1 - \tilde{\pi}_{i,t-1}), \quad (17)$$

$$\mathcal{I}_{ij,t-1} = -\tilde{\pi}_{i,t-1}\tilde{\pi}_{j,t-1}. \quad (18)$$

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

$$f_t = \omega + \sum_{i=0}^{q-1} A_i S_{t-i-1} (\tilde{y}_{t-i} - \tilde{\pi}_{t-i-1}) + \sum_{j=1}^p B_j f_{t-j}, \quad (19)$$

where the scale matrix $S_{t-1} = \mathcal{I}_{t-1}^{-1}$ can be constructed from (17) and (18). The ACM model of Russell and Engle (2005) can be obtained as a special case of the GAS model (19) 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 (19).

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 3 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 opening and closing of the market when volume is higher.

In our sample, 98% of the price changes fall within a ± 5 tick range of zero (see panel (iii)), where a tick is now 1 cent after decimilization of the market in 2001. Decimilization 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 $\dim(\tilde{f}_t) \ll \dim(f_t)$. The GAS(1,1) model reduces to

$$\tilde{f}_t = A_0 \Phi_1' \mathcal{I}_{t-1}^{-1} \Phi_1 \Phi_1' (\tilde{y}_t - \tilde{\pi}_{t-1}) + B_1 \tilde{f}_{t-1}, \quad (20)$$

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 (20) 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 3. 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.

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

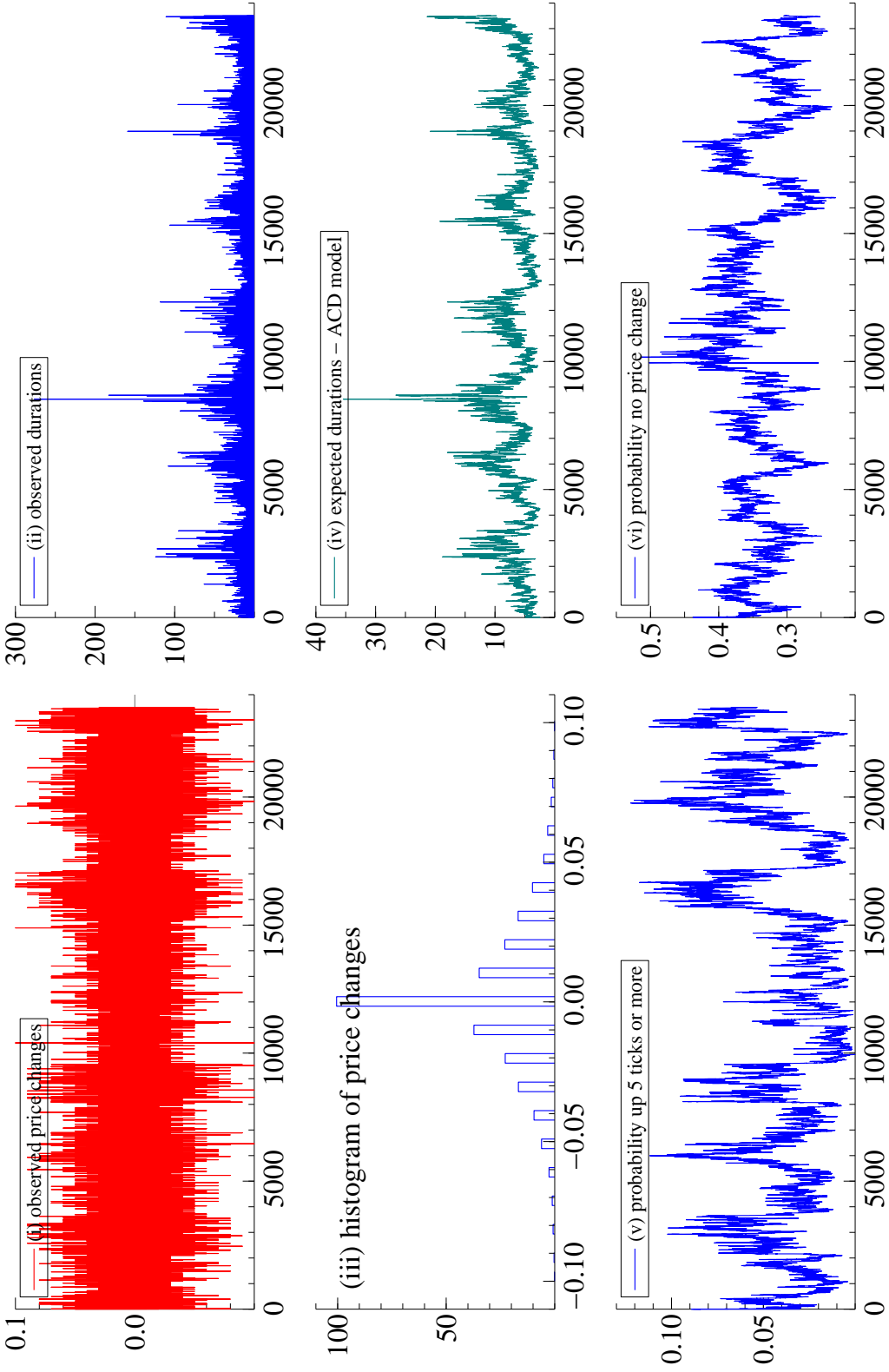


Figure 3: Time-varying multinomial $GAS(1,2)$ - $ACD(1,2)$ illustration: (i) observed price changes; (ii) observed durations; (iii) histogram of price changes; (iv) estimated expected duration from the ACD model; (v) estimated probability of an increase of 5 ticks or more; (vi) estimated probability of a trade with no change in price.

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

$$\mathcal{L}_t = \sum_{j=1}^J \pi_{jt} \mathcal{L}_{jt}. \quad (21)$$

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

$$\pi_{jt} = \frac{e^{f_{jt}}}{1 + \sum_{k=1}^{J-1} e^{f_{kt}}} \Leftrightarrow f_{jt} = \ln(\pi_{jt}) - \ln \left(1 - \sum_{k=1}^{J-1} \pi_{kt} \right), \quad (22)$$

for $j = 1, \dots, J-1$ 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

$$\frac{\partial \mathcal{L}_t}{\partial f_{j,t-1}} = \frac{\pi_{j,t-1} \mathcal{L}_{jt}}{\sum_{k=1}^J \pi_{k,t-1} \mathcal{L}_{kt}} - \pi_{j,t-1}, \quad (23)$$

for $j = 1, \dots, J-1$. The interpretation of (23) 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 (21) 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 scaling matrix S_t that is constructed using current and past $\mathcal{I}_{t|t-1}$ values which are weighted according some exponentially decaying scheme. The local smoothing for S_t is needed here to avoid that S_t becomes non-invertible. This GAS model provides an observation driven alternative to a hidden Markov model (HMM). We compare it to a simplified version of the model in Hamilton (1989) without autoregressive dynamics, that is

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

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.

Table 4: *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 (0.008)	1.127 (0.005)	0.869 (0.003)	0.360 (0.017)	2.333 (0.113)	0.672 (0.006)	-329.70
	μ_1	μ_2	σ	p_{11}	p_{22}	-	
HMM	-0.090 (0.019)	1.106 (0.007)	0.830 (0.003)	0.741 (0.007)	0.918 (0.003)		-333.17

Estimates of the parameters of both models are reported in Table 4. 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 4 presents the growth rate of log U.S. real GDP along with the estimated conditional mean $\pi_t \mu_1 + (1 - \pi_t) \mu_2$ from the GAS and HMM models. The GAS and HMM estimates nicely follow the changes in the mean of the series. The estimated probabilities of a recession from each model are plotted in panel (ii) of Figure 4. The estimated probabilities from the GAS model reflect the possibility of the model to rapidly adapt to

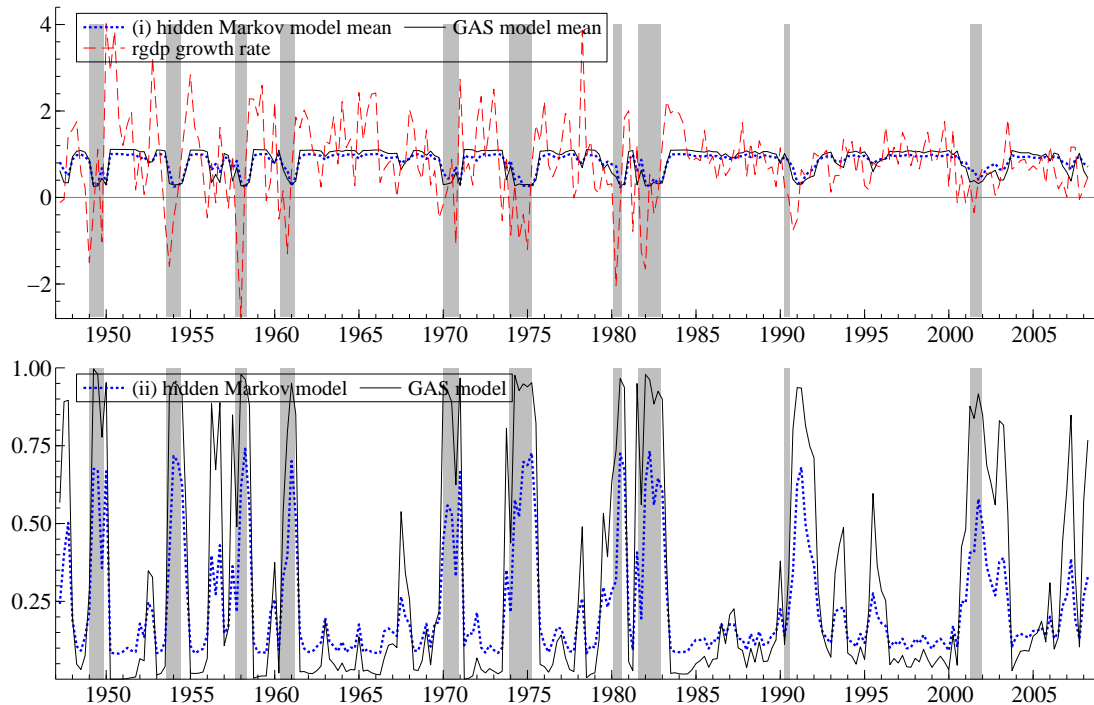


Figure 4: *Mixture model illustration: (i) growth rate of log U.S. real GDP from 1947(2)-2008(2) and the estimated conditional mean from the GAS(1,1) model and the hidden Markov model; (ii) one-step ahead predicted probability of a recession from each model. NBER recession dates are represented by the shaded regions.*

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.

Appendix C : Simulation experiments

In this section, we provide simulation evidence on the statistical properties of the GAS ML estimators. We present the simulation results for the two illustration models of the main paper : the bivariate Gaussian copula model with time-varying correlations and the marked point process model.

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 3.1 of the main paper. We consider the model specification in (23) of the main paper 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 5. 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 . The densities of the t-values appear slightly biased for the ω and B parameters. However, the bias diminishes as the sample size increases.

The pooled marked point process model

To investigate the statistical properties of the GAS model for the marked point processes of Section 4.2 in our main paper, 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 that is adopted in the application of Koopman, Lucas, and Monteiro (2008) and is given by

$$\lambda_{jk}(t) = \eta_j + \psi'_j f_{t^*}, \quad (24)$$

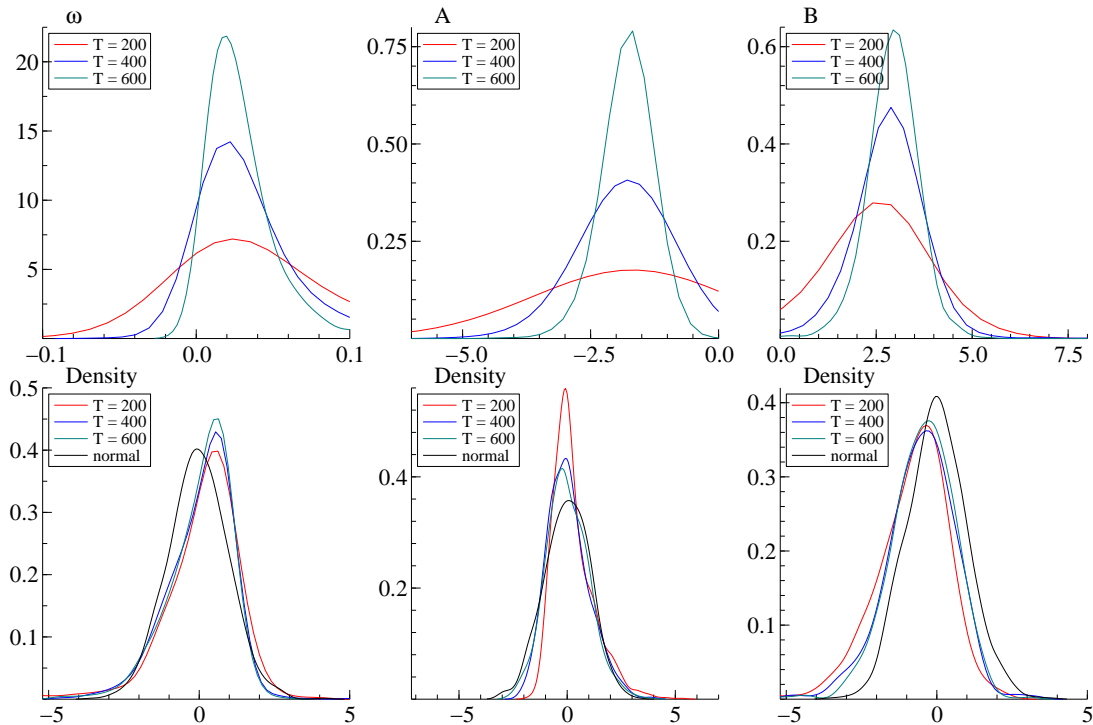


Figure 5: *Simulation densities over 1,000 simulations of a time-varying Gaussian copula model. The top panel contains the densities of the parameter estimates, the bottom panels contain the densities of t -values computed using the inverted second derivative of the Hessian at the optimum.*

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 $\omega = 0$. In particular, the GAS update equation which in our case are given by

$$\lambda_{1t} = \eta_1 + f_t, \quad \lambda_{2t} = \eta_2 + \alpha f_t, \quad f_t = A s_t + B f_{t-1},$$

where s_t is given by

$$s_t = \left[\sum_{j,k} w_{jk}(t) \psi_j \psi_j' \right]^{-1} \left(\sum_{j,k} y_{jk}(t) \psi_j - R_{jk}(t) \cdot (t - t^*) \cdot \exp(\lambda_{jk}(t)) \psi_j \right), \quad (25)$$

where $w_{jk}(t) = R_{jk}(t) \cdot \exp(\lambda_{jk}(t)) / \sum_{j,k} R_{jk}(t) \cdot \exp(\lambda_{jk}(t)) = \mathbb{P}[y_{jk}(t) = 1]$ is the probability

of the next event being of type j for company k . 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 of the main paper.

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.

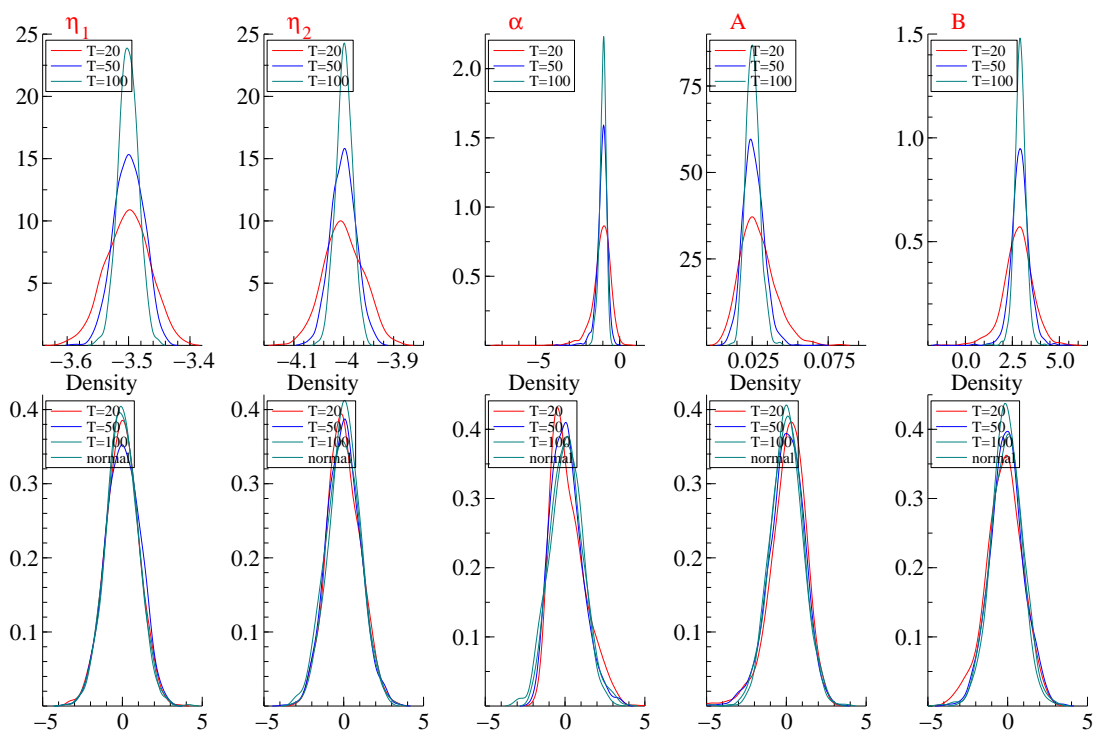


Figure 6: *Simulation densities over 1,000 simulations of a marked point process model. The top panel contains the densities of the parameter estimates, the bottom panels contain the densities of t -values computed using the inverted second derivative of the Hessian at the optimum.*

The Monte Carlo results are graphically presented in Figure 6. 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 $T = 20$.

References

- 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(3), 307–327.
- Bollerslev, T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. *The Review of Economics and Statistics* 69(3), 542–547.
- Brooks, C., S. P. Burke, S. Heravi, and G. Persaud (2005). Autoregressive conditional kurtosis. *Journal of Financial Econometrics* 3(3), 399–421.
- 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.
- Davis, R. A., W. T. M. Dunsmuir, and S. Streett (2003). Observation driven models for Poisson counts. *Biometrika* 90(4), 777–790.
- 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. and J. R. Russell (1998). Autoregressive conditional duration: a new model for irregularly spaced transaction data. *Econometrica* 66(5), 1127–1162.
- 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, C. R. and A. Siddique (1999). Autoregressive conditional skewness. *Journal of Financial and Quantitative Analysis* 34, 465–487.
- 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(1), 399–424.
- Lehmann, E. L. and G. Casella (1998). *Theory of Point Estimation*. New York, NY: Springer Press.
- 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).
- 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*.
- Watson, M. W. (1986). Univariate detrending methods and stochastic trends. *Journal of Monetary Economics* 18, 49–75.