

PREDICTING TIME-VARYING PARAMETERS WITH PARAMETER-DRIVEN AND OBSERVATION-DRIVEN MODELS

Siem Jan Koopman, André Lucas, and Marcel Scharth*

Abstract—We verify whether parameter-driven and observation-driven classes of dynamic models can outperform each other in predicting time-varying parameters. We consider existing and new dynamic models for counts and durations, but also for volatility, intensity, and dependence parameters. In an extended Monte Carlo study, we present evidence that observation-driven models based on the score of the predictive likelihood function have similar predictive accuracy compared to their correctly specified parameter-driven counterparts. Dynamic observation-driven models based on predictive score updating outperform models based on conditional moments updating. Our main findings are supported by the results from an extensive empirical study in volatility forecasting.

I. Introduction

IN this paper, we study the predictive ability of parameter-driven versus observation-driven model classes. We consider dynamic models for count, intensity, duration, volatility, and copula densities and present results for three specifications that have been widely used in the analysis and forecasting of economic and financial time series. The first specification is the nonlinear non-Gaussian state-space model as formulated, for example, in Shephard and Pitt (1997) and Durbin and Koopman (1997). This model belongs to the class of parameter-driven models. The second specification is an observation-driven model and is based on the score function of the predictive likelihood function. We refer to it as the generalized autoregressive score model. Creal et al. (2013) discuss this class of models in detail. The other observation-driven specification relies on the moment function of the time series. Typical examples are the generalized autoregressive conditional heteroskedasticity (GARCH) model of Engle (1982) and Bollerslev (1987), the autoregressive conditional duration (ACD) model of Engle and Russell (1998), and the multiplicative error models of Engle and Gallo (2006). For ease of reference, we group the latter set of models under the general heading of autoregressive conditional moment (ACM) models.

Cox (1981) classifies time-varying parameter models in two classes: observation-driven and parameter-driven specifications. In an observation-driven model, current parameter values are obtained as deterministic functions of lagged dependent variables, as well as contemporaneous and lagged

exogenous variables. In this approach, the parameters evolve randomly over time, but they are perfectly predictable one step ahead given past information. The likelihood function for observation-driven models is available in closed form via the prediction error decomposition. This feature leads to simple estimation procedures and has contributed to the popularity of these models in applied econometrics and statistics.

In parameter-driven models, parameters vary over time as dynamic processes with idiosyncratic innovations. Analytical expressions for the likelihood function are hardly ever available in closed form. Likelihood evaluation therefore becomes more involved for parameter-driven models, typically requiring the use of efficient simulation methods. Special cases of this class are stochastic volatility models as discussed by Tauchen and Pitts (1983), Taylor (1986), Melino and Turnbull (1990), and Ghysels, Harvey, and Renault (1996); the stochastic conditional duration model of Bauwens and Veredas (2004); the stochastic conditional intensity model of Bauwens and Hautsch (2006); the stochastic copula models of Hafner and Manner (2012); and the non-Gaussian unobserved components time series models of Durbin and Koopman (2000).

Given the different natures of observation-driven and parameter-driven models and the substantial efforts devoted to studying and applying a variety of these models, it is important to assess the relative merits of these two approaches from an out-of-sample perspective. A robust out-of-sample performance is key to the applicability of any time series model. However, three substantial problems have obstructed a systematic comparison between observation-driven and parameter-driven models across a range of data-generating processes (DGPs). Our aim is to contribute to the literature by providing solutions to each of these problems and, hence, enabling a wide-ranging comparison between the two classes of models.

First, parameter-driven models are flexible and easily applied in new settings. For any conditional observation density, we can make a specific parameter time-varying by turning it into a stochastic process subject to its own innovation. By contrast, observation-driven models have so far lacked a similarly flexible unifying framework. For a new observation density and parameterization, we need to construct a new function of the data to update the time-varying parameter. Whereas the appropriate function is (arguably) clear in some cases, such as volatility modeling, in many other settings, it may not be evident.

The generalized autoregressive score (GAS) model of Creal et al. (2013) is a class of observation-driven models with a similar degree of generality as obtained for nonlinear, non-Gaussian state-space models. The GAS model

Received for publication December 10, 2012. Revision accepted for publication November 17, 2014. Editor: Mark W. Watson.

* Koopman: VU University Amsterdam, Tinbergen Institute, and CREATES, Aarhus University; Lucas: VU University Amsterdam and Tinbergen Institute; Scharth: University of New South Wales.

We thank Andrew Harvey for comments on an earlier version of this paper. S.J.K. acknowledges support from CREATES, Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation. A.L. acknowledges the financial support of the Dutch National Science Foundation (NWO, grant VICI453-09-005). M.S. acknowledges partial support from the Australian Research Council grants DP0988579 and LP0774950.

A supplemental appendix is available online at http://www.mitpressjournals.org/doi/suppl/10.1162/REST_a_00533.

adopts the score vector of the predictive model density to update the time-varying parameters. It follows that GAS models can be based on any observation density. Creal et al. (2013) show that the GAS class encompasses well-known observation-driven models such as the GARCH model of Bollerslev (1986). At the same time it enables the development of new models such as the mixed measurement dynamic factor model of Creal et al. (2014). In the same way as parameter-driven models, GAS models can accommodate different parameterizations of the observation density in a straightforward manner. We believe that the GAS framework provides a natural observation-driven alternative to the state-space framework for a wide range of different DGPs.

Second, observation-driven and parameter-driven models are inherently hard to compare even if they are based on the same measurement density. The difficulty stems from the fact that the predictive distribution of a parameter-driven model is a mixture of measurement densities over the random time-varying parameter, whereas the predictive density of observation-driven models is simply the observation density given a perfectly predictable parameter. Parameter-driven models typically generate overdispersion, heavier tails, and other features that may directly put such models at an advantage over observation-driven models.

In order to develop a systematic comparison between the two classes of models, we need to control for this distinction. We therefore develop new observation-driven models that aim to accommodate similar degrees of overdispersion and fat tails as parameter-driven models. We introduce new generalized autoregressive score models based on exponential gamma, Weibull gamma, and double gamma mixtures. Beyond their role in the current analysis, these GAS model formulations are also of intrinsic interest as new duration and multiplicative error models that combine the flexibility of their mixture distributions with score-based updates.

Third, parameter estimation for nonlinear non-Gaussian state-space models is computationally intensive. As a result, large-scale comparative analyses such as the one in Hansen and Lunde (2005) often exclude parameter-driven models. To overcome this computational challenge, we turn to the recently developed numerically accelerated importance sampling method (NAIS) of Koopman, Lucas, and Scharth (2015). The NAIS algorithm leads to fast and numerically efficient parameter estimation for nonlinear non-Gaussian state-space models and requires no model-specific interventions other than the specification of the appropriate observation densities. We can therefore easily apply the NAIS algorithm repeatedly across the range of DGPs considered in our analysis. We also employ the method for the efficient computation of the out-of-sample forecasts, the prime focus of this study.

We obtain two main findings. First, when the DGP is a state-space model, the predictive accuracy of a (misspecified) GAS model is similar to that of a (correctly specified) state-space model. This holds in particular if the (conditional) observation density for the GAS specification allows

for heavy tails and overdispersion. For the nine model specifications in this paper, the loss in mean squared error from using a GAS model instead of the correct state-space specification is smaller than 1% most of the time and never higher than 2.5%. We conclude that we can obtain high predictive accuracy for many relevant time-varying parameter models without the need to specify and estimate a parameter-driven model. Second, we find that the GAS models outperform many of the familiar observation-driven models from the literature, which we have referred to as ACM models. By relying on the full density structure to update the time-varying parameters, GAS models capture additional information in the data that is not exploited by ACM models (see also Blasques, Koopman, and Lucas, 2015). GAS models are therefore effective new tools for forecasting that often lead to important forecasting gains over other classes of observation-driven models.

To illustrate the relevance of our findings for empirical applications, we analyze the performance of different parameter-driven and observation-driven models for predicting the volatility of twenty Dow Jones index stocks and five major stock indices over a period of several years. We find that stochastic volatility and GAS models based on the Student's- t distribution achieve a similar performance for all series, while ACM models in many cases lead to substantially larger forecasting errors. This conclusion is robust to the introduction of leverage effects, more complex two-factor specifications for the volatility dynamics, and different sample sizes.

We structure the rest of this paper as follows. In section II, we present our three econometric approaches for modeling time-varying parameters. section III introduces several new GAS models for continuous mixtures. Section IV discusses the estimation of parameters for the different model classes. Section V presents the Monte Carlo results. Section VI presents the empirical application. Section VII concludes.

II. Modeling Time-Varying Parameters

A. Dynamic Model Specifications

Let y_1, \dots, y_n denote a sequence of $p \times 1$ dependent variables of interest. In financial applications, for example, the variables may represent stock returns, the time between asset transactions, the number of firm defaults within a certain period, and so on. We are interested in modeling the mean, variance, or another relevant characteristic of the conditional distribution of y_t given all the data up to time $t-1$. We assume that y_t is generated by the observation density

$$y_t | \theta_t \sim p(y_t | \theta_t; \psi), \quad \theta_t = \Lambda(\alpha_t), \quad t = 1, \dots, n, \quad (1)$$

where θ_t is a time-varying parameter vector, $\Lambda(\cdot)$ is a possibly nonlinear link function, and α_t has a linear dynamic specification. In this paper, we focus on the case in which

α_t is a scalar variable. The static parameter vector ψ incorporates additional fixed and unknown coefficients from the density $p(y_t|\theta_t; \psi)$.

State-space models. In parameter-driven models, the state vector α_t evolves according to an idiosyncratic source of innovations. We model α_t as a Gaussian autoregressive process of order 1

$$\begin{aligned}\alpha_{t+1} &= \delta + \phi\alpha_t + \eta_t, & \alpha_1 &\sim N(a_1, P_1), \\ \eta_t &\sim N(0, \sigma_\eta^2),\end{aligned}\quad (2)$$

where δ is a constant and ϕ is the autoregressive coefficient. We assume that the initial state vector α_1 is normally distributed with mean $\delta/(1 - \phi)$ and variance $\sigma_\eta^2/(1 - \phi^2)$.

Equations (1) and (2) characterize a class of nonlinear non-Gaussian state-space models (see Durbin & Koopman, 2012, for a general discussion). More generally, the state vector α_t can also represent higher-order autoregressive moving average, random walk, cyclical, seasonal, and other dynamic processes, including an aggregation of those components. Shephard and Pitt (1997) and Durbin and Koopman (1997) develop simulation-based methods for the estimation of ψ , α_t , and θ_t . Liesenfeld and Richard (2003), Richard and Zhang (2007), Jungbacker and Koopman (2007), and Koopman et al. (2015) report recent developments on Monte Carlo methods for the analysis of general nonlinear non-Gaussian state-space models.

Generalized autoregressive score models. In observation-driven models, the time-varying vector α_t in equation (1) depends on lagged values of y_t and the model parameters in a deterministic way. We consider the autoregressive update

$$\alpha_{t+1} = d + a s_t + b \alpha_t, \quad (3)$$

where d , a , and b are fixed coefficients and $s_t = s_t(\alpha_t, \mathcal{F}_t; \psi)$ is the driving mechanism, with \mathcal{F}_t representing the information set consisting of all observations up to time t . We can also consider extensions of the dynamic specification (3), which are similar to the ones available for the state-space model. For example, the dynamic process for α_t can include more lags for s_t and α_t . Another example is to decompose α_t in different dynamic processes that are associated with seasonal and cycle effects as in Harvey and Luati (2014). Finally, the autoregressive update (3) can be replaced by a long memory process as in Janus, Koopman, and Lucas (2014).

Specific choices for the driving mechanism s_t lead to different classes of observation-driven models. In this paper, we focus on the generalized autoregressive score (GAS) class of Creal et al. (2013) as our main observation-driven model. The GAS framework is of similar generality as the state-space model, equations (1) and (2), in that it is applicable to any measurement density. In this framework, the updating step s_t in equation (3) is the scaled density score,

$$\begin{aligned}s_t &= S_t(\alpha_t) \cdot \nabla_t(\alpha_t), & \nabla_t(\alpha_t) &= \frac{\partial \ln p(y_t | \alpha_t, \mathcal{F}_{t-1}; \psi)}{\partial \alpha_t}, \\ S_t(\alpha_t) &= S(t, \alpha_t, \mathcal{F}_{t-1}; \theta),\end{aligned}\quad (4)$$

where $S(\cdot)$ is the scaling matrix. A GAS model updates the parameter α_{t+1} in the direction of steepest increase of the log density at time t given the current parameter α_t and data history \mathcal{F}_t . It follows from the properties of the score vector that $\mathbb{E}(s_t|\mathcal{F}_{t-1}) = 0$, and hence the GAS update is a martingale difference under the correct specification.

Creal et al. (2013) discuss appropriate choices for S_t based on the curvature of the log density at time t as summarized by the Fisher information matrix,

$$\mathcal{I}_t(\alpha_t) = \mathbb{E}[\nabla_t(\alpha_t)\nabla_t(\alpha_t)'|\mathcal{F}_{t-1}], \quad (5)$$

therefore linking the scaling matrix to the variance of the score. In this paper, we take the scaling matrix as $S_t(\alpha_t) = \mathcal{I}_t(\alpha_t)^{-1/2}$. For this choice of scaling, the updating step s_t has a constant unit variance and is invariant under any nondegenerate parameter transformation $\Lambda(\cdot)$. The constant unit variance property can be a useful device for detecting model misspecification in applications. Other choices for the scaling matrix such as $S_t = \mathcal{I}_t^{-1}$ are also possible and lead to different observation-driven models (see Creal et al., 2013). We can also specify the observation-driven dynamics directly for θ_t .

A useful feature of the GAS approach is the automatic treatment of the link function $\Lambda(\cdot)$. This characteristic facilitates comparisons between GAS and state-space models and is particularly helpful if the parameter of interest θ_t is subject to constraints. For example, if θ_t is a correlation parameter, $\theta_t = \tanh(\alpha_t)$ ensures that the correlation is between -1 and $+1$. Finally, for some models such as the time-varying conditional volatility, intensity, or duration models, using the transformation $\theta_t = \exp(\alpha_t)$ leads to an information matrix \mathcal{I}_t that does not depend on α_t , such that the choice of scaling matrix S_t becomes irrelevant. Here we report the results for GAS models with $S_t = \mathcal{I}_t^{-1/2}$, whether it depends on α_t or not. We have found that this choice of scaling leads to greater predictive accuracy across different DGPs.

Autoregressive conditional moment models. Many observation-driven models link the time-varying parameter directly to the data or some transformation of it. A common approach is to define s_t such that

$$\mathbb{E}[s_t(y_t, \theta_t; \psi)|\mathcal{F}_{t-1}] = \theta_t = \alpha_t. \quad (6)$$

We refer to this class as autoregressive conditional moment (ACM) models.

The ACM model adheres to the intuitive notion that the time-varying parameter should increase (or decrease) if the realized value for s_t is higher (or lower) than its conditional expectation. For example, if θ_t is the conditional mean $\theta_t = \mathbb{E}(y_t|\mathcal{F}_{t-1})$, the ACM update is $s_t = y_t$. Similarly, if

TABLE 1.—OBSERVATION DENSITIES

Model Type	Distribution	Density	Link Function
Count	Poisson	$\frac{\lambda_t^{y_t}}{y_t!} e^{-\lambda_t}$	$\lambda_t = \exp(\alpha_t)$
Count	Negative binomial	$\frac{\Gamma(k_1 + y_t)}{\Gamma(k_1)\Gamma(y_t + 1)} \left(\frac{k_1}{k_1 + \lambda_t}\right)^{k_1} \left(\frac{\lambda_t}{k_1 + \lambda_t}\right)^{y_t}$	$\lambda_t = \exp(\alpha_t)$
Intensity	Exponential	$\lambda_t e^{-\lambda_t y_t}$	$\lambda_t = \exp(\alpha_t)$
duration	gamma	$\frac{1}{\Gamma(k_1)\beta_t^{k_1}} y_t^{k_1-1} e^{-y_t/\beta_t}$	$\beta_t = \exp(\alpha_t)$
Duration	Weibull	$\frac{k_1}{\beta_t} \left(\frac{y_t}{\beta_t}\right)^{k_1-1} e^{-(y_t/\beta_t)^{k_1}}$	$\beta_t = \exp(\alpha_t)$
volatility	Gaussian	$\frac{1}{\sqrt{2\pi}\sigma_t} e^{-y_t^2/2\sigma_t^2}$	$\sigma_t^2 = \exp(\alpha_t)$
Volatility	Student's- <i>t</i>	$\frac{\Gamma(\frac{v+1}{2})}{\sqrt{(v-2)\pi}\Gamma(\frac{v}{2})\sigma_t} \left(1 + \frac{y_t^2}{(v-2)\sigma_t^2}\right)^{-\frac{v+1}{2}}$	$\sigma_t^2 = \exp(\alpha_t)$
Copula	Gaussian	$\frac{1}{2\pi\sqrt{1-\rho_t^2}} \exp\left[-\frac{z_{1t}^2 + z_{2t}^2 - 2\rho_t z_{1t}z_{2t}}{2(1-\rho_t^2)}\right]$ $\prod_{i=1}^2 \frac{1}{\sqrt{2\pi}} e^{-z_{it}^2/2}$	$\rho_t = \frac{1 - \exp(-\alpha_t)}{1 + \exp(-\alpha_t)}$
Copula	Student's- <i>t</i>	$\frac{\Gamma(\frac{v+2}{2})\Gamma(\frac{v}{2})}{\Gamma(\frac{v+1}{2})} \frac{1}{\sqrt{1-\rho_t^2}} \left[1 + \frac{z_{1t}^2 + z_{2t}^2 - 2\rho_t z_{1t}z_{2t}}{v(1-\rho_t^2)}\right]^{-\frac{v+2}{2}}$ $\prod_{i=1}^2 \frac{1}{\Gamma(\frac{v+1}{2})} (1 + z_{it}^2/v)^{-\frac{v+1}{2}}$	$\rho_t = \frac{1 - \exp(-\alpha_t)}{1 + \exp(-\alpha_t)}$

The table displays the dynamic densities that we consider in our simulation study. We write them as $p(y_t|\theta_t; \psi)$, where θ_t is the parameter of interest. We assume that $\theta_t = \Lambda(\alpha_t)$, where $\Lambda(\cdot)$ is a monotonically increasing transformation, and α_t has a linear dynamic specification. We denote the data by y_t . For the Gaussian copula model, $z_{it} = \Phi^{-1}(y_{it})$, where the observations y_{it} have uniform (0, 1) marginal distributions and $\Phi^{-1}(\cdot)$ denotes the inverse normal CDF. For the Student's-*t* copula, $z_{it} = T_v^{-1}(y_{it})$, where the observations y_{it} have uniform (0, 1) marginal distributions and $T_v^{-1}(y_{it})$ is the inverse CDF of a Student's-*t* with v degrees of freedom.

θ_t is the conditional variance of y_t , then $s_t = (y_t - \mu_{y,t-1})^2$ with $\mu_{y,t-1} = \mathbb{E}(y_t|\mathcal{F}_{t-1})$.

Key examples of ACM models include the generalized autoregressive conditional heteroskedasticity (GARCH) model of Bollerslev (1986), the autoregressive conditional duration (ACD) and intensity (ACI) models of Bollerslev (1986), the autoregressive conditional Poisson model of Rydberg and Shephard (2000), the dynamic conditional correlation model of Engle (2002), some autoregressive copulas in Patton (2006), and realized volatility models as in Shephard and Sheppard (2010) and Hansen, Huang, and Shek (2012). Due to their widespread use, the class of ACM models provides a useful benchmark in our analysis.

B. Observation Densities

In table 1 we present the observation densities $p(y_t|\theta_t; \psi)$ from equation (1) as used in our current simulation study. We consider a wide range of specifications, including densities for count, intensity, duration, volatility, and copula models. The combination of the dynamic gamma, Weibull, normal, Student's-*t*, and copula densities in table 1 with equation (2) directly leads to the stochastic conditional duration, stochastic volatility, and stochastic copula models as in Tauchen and Pitts (1983), Bauwens and Veredas (2004), Bauwens and Hautsch (2006), and Hafner and Manner (2012).

Table 2 completes the specifications of the considered observation-driven models. It presents the generalized autoregressive score and autoregressive conditional moment updates s_t for the densities in table 1. The ACM updates lead to the well-known autoregressive conditional Poisson, autoregressive conditional duration, autoregressive conditional intensity, GARCH, and autoregressive copula

specifications. The ACM model for the gamma distribution with $k = 1/\beta_t$ corresponds to the multiplicative error model of Engle and Gallo (2006). The GAS specifications for the exponential, gamma, normal, and Student's-*t* volatility models and the Gaussian copula have appeared in Creal et al. (2013). The Student's-*t* volatility model is also discussed in Harvey (2013). The GAS model for the Student's-*t* copula is obtained from Creal, Koopman, and Lucas (2011).

C. Parameter-Driven versus Observation-Driven Models

When considering parameter-driven models, $p(y_t|\mathcal{F}_{t-1}; \psi)$ is the mixture distribution

$$p(y_t|\mathcal{F}_{t-1}; \psi) = \int_{\theta} p(y_t|\theta_t; \psi) p(\theta_t|\mathcal{F}_{t-1}; \psi) d\theta_t. \quad (7)$$

In a parameter-driven framework, estimation of θ_t takes by construction into account the complete density structure of past observations. As we discuss below, the same is not necessarily true for observation-driven models. The mixture distribution, equation (7), may also describe relevant features of the data. It is typically the case that higher-order conditional moments of y_t , such as kurtosis, are at least as high for $p(y_t|\mathcal{F}_{t-1}; \psi)$ as for $p(y_t|\theta_t; \psi)$. For example, Carnero, Peña, and Ruiz (2004), among others, show that the Gaussian stochastic volatility model of table 1 with the dynamics as specified in equation (2) is conditionally leptokurtic. Similarly, the stochastic count and duration models we study below display conditional overdispersion.

A major obstacle for the application of the parameter-driven models in table 1 is that the conditional density $p(\theta_t; \psi|\mathcal{F}_{t-1})$ is typically not available in closed form. The

TABLE 2.—OBSERVATION-DRIVEN MODEL UPDATES

Model type	Distribution	GAS		ACM
		$\nabla_t(\theta_t)$	$\mathcal{I}_t(\theta_t)$	s_t
Count	Poisson	$\frac{y_t}{\lambda_t} - 1$	$\frac{1}{\lambda_t}$	y_t
Count	Negative binomial	$\frac{y_t}{\lambda_t} - \frac{k_1 + y_t}{k_1 + \lambda_t}$	$\frac{k_1}{\lambda_t(k_1 + \lambda_t)}$	y_t
Intensity	Exponential	$\frac{1}{\lambda_t} - y_t$	$\frac{1}{\lambda_t^2}$	y_t
Duration	Gamma	$\frac{y}{\theta_t^2} - \frac{k_1}{\beta_t}$	$\frac{k}{\beta_t^2}$	y_t/k_1
Duration	Weibull	$\frac{k_1}{\beta_t} \left[\left(\frac{y_t}{\beta_t} \right)^{k_1} - 1 \right]$	$\left(\frac{k_1}{\beta_t} \right)^2$	$\frac{y_t}{\Gamma(1+k_1^{-1})}$
Volatility	Gaussian	$\frac{1}{2\sigma_t^2} \left(\frac{y_t^2}{\sigma_t^2} - 1 \right)$	$\frac{1}{2\sigma_t^4}$	y_t^2
Volatility	Student's- t	$\frac{1}{2\sigma_t^2} \left(\frac{\omega_t y_t^2}{\sigma_t^2} - 1 \right)$ $\omega_t = \frac{v+1}{(v-2)+y_t^2/\sigma_t^2}$	$\frac{v}{2(v+3)\sigma_t^4}$	y_t^2
Copula	Gaussian	$\frac{(1+\rho^2)(\hat{z}_{1,t}-\rho_t)-\rho_t(\hat{z}_{2,t}-2)}{(1-\rho^2)^2}$	$\frac{1+\rho_t^2}{(1-\rho_t^2)^2}$	$z_{1,t}z_{2,t}$
Copula	Student's- t	$\frac{(1+\rho^2)(\omega_t \hat{z}_{1,t}-\rho_t)-\rho_t(\omega_t \hat{z}_{2,t}-2)}{(1-\rho^2)^2}$ $\omega_t = \frac{v+2}{v+\frac{\hat{z}_{2,t}-2\rho_t \hat{z}_{1,t}}{1-\rho^2}}$	$\frac{(v+2+v\rho_t^2)}{(v+4)(1-\rho_t^2)^2}$	$z_{1,t}z_{2,t}$

The table displays the score and information matrix for the models given in table 1. The GAS update $s_t = \nabla_t(\alpha_t)\mathcal{I}_t(\alpha_t)^{-1/2}$ we consider in this paper is invariant under nondegenerate parameter transformations, so that $s_t = \nabla_t(\theta_t)\mathcal{I}_t(\theta_t)^{-1/2}$. The equivalent ACM models are without parameter transformation; see table 1. The data are denoted by y_t . For the copula models, $y_{i,t}$ has a uniform $(0, 1)$ marginal distribution for $i = 1, 2$, $\hat{z}_{1,t} = z_{1,t}z_{2,t}$, and $\hat{z}_{2,t} = z_{1,t}^2 + z_{2,t}^2$, where $z_{i,t} = \Phi^{-1}(y_{i,t})$ for the Gaussian copula and $z_{i,t} = T_v^{-1}(y_{i,t})$ for the Student's- t copula, with $\Phi^{-1}(\cdot)$ and $T_v^{-1}(\cdot)$ denoting the inverse normal CDF and the inverse CDF of a Student's- t distribution with v degrees of freedom, respectively.

likelihood-based estimation of parameters in parameter-driven models therefore requires the use of computationally intensive simulation methods for evaluating the high-dimensional integral that characterizes the likelihood function of the model (see Durbin & Koopman, 2002). Simulation methods are also necessary for estimating and forecasting the time-varying parameter θ_t . We discuss parameter estimation in section IV.

In observation-driven models, the time-varying parameter θ_t is perfectly predictable one step ahead given past information. It implies that the likelihood functions for ACM and GAS models are available in closed form. Hence, the estimation of the parameters is straightforward for observation-driven models. It also contributes to their widespread use in applied econometrics and statistics. Within the observation-driven class of models, some key differences exist between GAS and ACM specifications. GAS models can handle parameter transformations and are applicable in cases where ACM updates are not readily available (see Creal et al., 2013). By making use of the observation density score, the GAS model update s_t takes the full density information into account (see Blasques et al., 2015). By contrast, ACM models rely exclusively on a specific moment of $p(y_t|\theta_t; \psi)$, such as its mean or variance.

We illustrate the difference using time-varying volatility models for the Gaussian and Student's- t distributions based on the GAS and ACM approaches. From table 2, we learn that a Gaussian GAS(1,1) volatility model with $\alpha_t = \sigma_t^2$ and update $s_t = \mathcal{I}_t^{-1}\nabla_t$ reduces to

$$\alpha_{t+1} = d + a(y_t^2 - \alpha_t) + b\alpha_t, \quad (8)$$

which is equivalent to the standard GARCH(1,1) model, the corresponding ACM model. If we replace the normal distribution by the Student's- t distribution with v degrees of freedom, the ACM model updating step remains the same, such that the ACM model reduces to the GARCH(1,1) model with Student's- t distributed errors. However, we learn from table 2 that the GAS update step becomes

$$s_t = \mathcal{I}_t^{-1}\nabla_t = (1 + 3v^{-1}) \cdot \left(\frac{(1 + v^{-1})}{(1 - 2v^{-1}) \{1 + v^{-1}y_t^2 / [(1 - 2v^{-1})\alpha_t]\}} y_t^2 - \alpha_t \right). \quad (9)$$

If $v^{-1} \rightarrow 0$, the GAS update step recovers the GARCH(1,1) model, as it should. If v is finite and observations are fat tailed, a large value y_t^2 receives less weight in the updating of α_{t+1} due to its presence in the denominator of equation (9). This feature is intuitively appealing. The GARCH update with y_t^2 becomes more volatile in the presence of fat tails. Large values of y_t^2 are then more likely to reflect noise caused by the excess kurtosis in the conditional distribution of the dependent variable rather than large increases in variance. Therefore, the GAS update step for the Student's- t distribution discounts large values of y_t^2 in comparison to the Gaussian case.

III. Observation-Driven Continuous Mixture Models

For a given observation density $p(y_t|\theta_t; \psi)$, parameter-driven and observation-driven specifications imply different models for the conditional density $p(y_t|\mathcal{F}_{t-1}; \psi)$. We need

to address this distinction in order to carry out a systematic comparison between these two approaches. In this section, we develop new observation-driven models based on exponential gamma, Weibull gamma, and double gamma mixtures. The new GAS models display overdispersion and fat-tail features that are comparable to those implied by parameter-driven models.

A. Weibull Gamma and Exponential Gamma Mixture Models

We consider the following parameterization of the Weibull distribution,

$$p(y_t | \gamma_t; k_1) = \gamma_t k_1 y_t^{k_1-1} \exp(-\gamma_t y_t^{k_1}), \quad (10)$$

where k_1 is a shape coefficient and γ_t is a time-varying scale variable. It follows that $\mathbb{E}(y_t | \gamma_t, k) = \gamma_t^{-1/k_1} \Gamma(1/k_1 + 1)$. Let $\gamma_t = \mu_t v_t$ where $\alpha_t = \log(\mu_t)$ follows a GAS updating step and v_t is an i.i.d. $\Gamma(k_2^{-1}, k_2)$ distributed random error with density function

$$p(v_t; k_2) = \frac{v_t^{k_2^{-1}-1} e^{-v_t/k_2}}{\Gamma(k_2^{-1}) k_2^{k_2^{-1}}}. \quad (11)$$

The multiplicative error v_t has mean 1 and variance $k_2 < \infty$. The Weibull gamma mixture or Burr density is given by

$$\begin{aligned} p(y_t | \mu_t; k) &= \int_0^\infty p(y_t | \mu_t, v_t; k_1) p(v_t) dv_t \\ &= \mu_t k_1 y_t^{k_1-1} \left(1 + k_2 \mu_t y_t^{k_1}\right)^{-(1+k_2^{-1})}. \end{aligned} \quad (12)$$

Lancaster (1979) and Das and Srinivasan (1997) illustrate the use of this distribution in econometrics. Grammig and Maurer (2000) propose an ACD model for a Weibull gamma mixture. They advocate the Burr model for the empirical analysis of price durations on the basis of its ability to account for nonmonotonic hazard functions (see also the discussions in Harvey, 2013).

We notice that

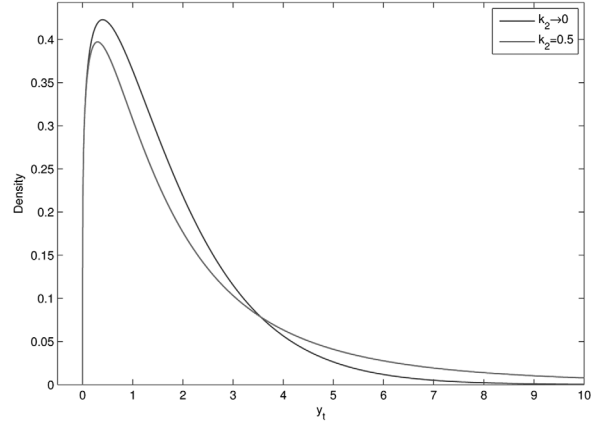
$$\begin{aligned} \mathbb{E}(y_t | \mu_t; k_1, k_2) &= \mu_t^{-k_1^{-1}} \Gamma(k_1^{-1} + 1) \mathbb{E}\left(v_t^{-k_1^{-1}}\right) \\ &= (\mu_t k_2)^{-k_1^{-1}} \frac{\Gamma(k_2^{-1} - k_1^{-1})}{\Gamma(k_2^{-1})}, \end{aligned} \quad (13)$$

and hence we need to impose $0 < k_2 < k_1$ so that $\Gamma(k_2^{-1} - k_1^{-1})$ exists. The score and the inverse of the Fisher information matrix with respect to μ_t are given by

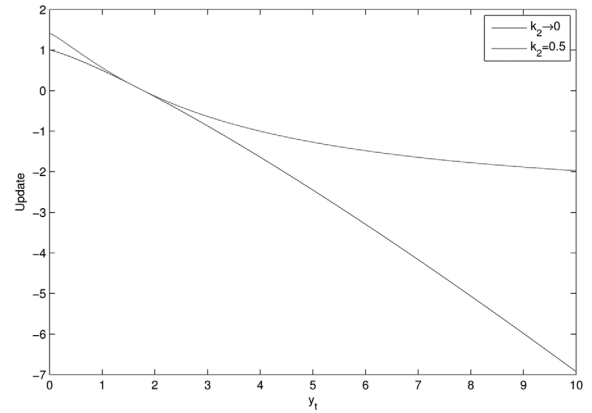
$$\nabla_t = \frac{1}{\mu_t} - (1 + k_2) \frac{y_t^{k_1}}{1 + k_2 \mu_t y_t^{k_1}}, \quad \mathcal{I}_t^{-1} = \mu_t^2 (1 + 2k_2), \quad (14)$$

respectively. This update recovers the Weibull model when $k_2 \rightarrow 0$. We base the GAS update step for $\alpha_t = \log(\mu_t)$ on the scaled score

FIGURE 1.—WEIBULL GAMMA OR BURR MIXTURE GAS MODEL WITH $k = 1.2$ AND $\mu_t = 0.5$



(a) Weibull-gamma mixture probability density



(b) Predictive score updates

$$s_t = \mathcal{I}_t^{-1/2} \nabla_t = \sqrt{1 + 2k_2} \left(1 - (1 + k_2) \frac{\mu_t y_t^{k_1}}{1 + k_2 \mu_t y_t^{k_1}} \right). \quad (15)$$

By setting $k_1 = 1$ above, we obtain the exponential gamma GAS model.

Figure 1 illustrates the probability density function and the GAS updates for the Weibull ($k_2 = 0$) and the Weibull gamma mixture model ($k_2 = 0.5$) for $k_1 = 1.2$ and $\mu_t = 0.5$. For $k_2 = 0$, the scaled score for the update step collapses to $s_t = 1 - \mu_t y_t^k$. Panel a in figure 1 shows that the mixture density function significantly stretches the right tail of the distribution. Hence, large values of y_t typically signal a large realization of v_t , containing little information about μ_t . Accordingly, panel b shows that realizations of y_t in the right tail of the distribution have a limited additional impact on s_t in the mixture model. This property contrasts sharply with the corresponding ACM model, in which the update step for the conditional mean is linear in y_t irrespective of the value of the mixture variance k_2 .

B. Double Gamma Mixture Models

We can follow a similar approach to obtain a gamma-gamma mixture model. Let y_t be i.i.d. $\Gamma(k_1, \gamma_t^{-1})$ random variables with shape coefficient k_1 , time-varying scale variable γ_t^{-1} , and density function

$$p(y_t|\gamma_t; k_1) = \frac{\gamma_t^{k_1} y_t^{k_1-1} e^{-\gamma_t y_t}}{\Gamma(k_1)}, \quad (16)$$

where $\gamma_t = \mu_t v_t$. The error v_t follows the $\Gamma(k_2^{-1}, k_2)$ distribution with density (11).

The mixture density is

$$p(y_t|\mu_t; k_1, k_2) = \frac{\Gamma(k_1 + k_2^{-1})}{\Gamma(k_1)\Gamma(k_2^{-1})} \frac{k_2^{k_1} \mu_t^{k_1} y_t^{k_1-1}}{(1 + k_2 \mu_t y_t)^{k_1+k_2^{-1}}}. \quad (17)$$

We have

$$\mathbb{E}(y_t|\mu_t; \delta) = \frac{1}{\mu_t(1 - k_2)}, \quad (18)$$

which leads to the requirement that $0 < k_2 < 1$. We obtain

$$\begin{aligned} \nabla_t &= \frac{k_1}{\mu_t} - (1 + k_1 k_2) \frac{y_t}{1 + k_2 \mu_t y_t} \quad \text{and} \\ \mathcal{I}_t &= \frac{k_1}{\mu_t^2(1 + k_2(k_1 + 1))}. \end{aligned} \quad (19)$$

Hence the intuition for the GAS updates of the gamma-gamma mixture distribution is similar to that of the Weibull-gamma model.

IV. Maximum Likelihood Estimation

We estimate the parameter vectors in the three classes of models (state space, GAS, and ACM) by the method of maximum likelihood. We maximize the log-likelihood function numerically with respect to the parameters using numerical gradient-based optimization methods. The evaluation of the log-likelihood function is straightforward for observation-driven models. For the parameter-driven models, we rely on simulation methods for the evaluation of the log-likelihood function. Recent developments in importance sampling have shown that fast and reliable simulated maximum likelihood estimation is feasible for nonlinear non-Gaussian state-space models.

A. Observation-Driven Models: Maximum Likelihood

Given an observed time series y_1, \dots, y_n , we use the standard prediction error decomposition to obtain the maximum likelihood estimates as

$$\hat{\psi} = \arg \max_{\psi} \sum_{t=1}^n \ell_t, \quad (20)$$

where $\ell_t = \ln p(y_t|\theta_t, \mathcal{F}_{t-1}; \psi)$. We deduce $p(y_t|\theta_t, \mathcal{F}_{t-1}; \psi)$ directly from equation (1) for a given model. We evaluate

the log-likelihood functions for the GAS and ACM models after implementing the GAS and ACM updating steps and calculating ℓ_t for particular values of ψ . We obtain estimates of θ_t by evaluating the GAS or ACM recursions with ψ set equal to the maximum likelihood estimate $\hat{\psi}$.

B. Parameter-Driven Models: Simulated Maximum Likelihood

The numerically accelerated importance sampling (NAIS) method of Koopman et al. (2015) is a computationally and numerically efficient method for obtaining an unbiased estimate of the likelihood function of a nonlinear non-Gaussian state-space model. The method is applicable to a wide class of observation densities and is able to treat all model specifications in table 1. The method requires only the specification of equation (1) and a linear state equation such as equation (2). The computation times for parameter estimation range from a few seconds to slightly less than a minute for the sample size of 2,000 observations we consider in section V.

The likelihood for the state-space model specified by equations (1) and (2) is given by the analytically intractable integral

$$\begin{aligned} L(y; \psi) &= \int p(\alpha, y; \psi) d\alpha \\ &= \int \prod_{t=1}^n p(y_t|\alpha_t; \psi) p(\alpha_t|\alpha_{t-1}; \psi) d\alpha_1 \dots d\alpha_n, \end{aligned} \quad (21)$$

where $\alpha' = (\alpha'_1, \dots, \alpha'_n)$, $y' = (y'_1, \dots, y'_n)$ and $p(\alpha, y; \psi)$ is the joint density of y and α . To evaluate the likelihood function by importance sampling, we consider a Gaussian importance density $g(\alpha, y; \psi) = g(y|\alpha; \psi)g(\alpha; \psi)$, where $g(y|\alpha; \psi)$ and $g(\alpha; \psi)$ are both Gaussian densities. We then express the likelihood function as

$$\begin{aligned} L(y; \psi) &= \int \frac{p(\alpha, y; \psi)}{g(\alpha, y; \psi)} g(\alpha, y; \psi) d\alpha \\ &= g(y; \psi) \int \omega(\alpha, y; \psi) g(\alpha|y; \psi) d\alpha, \end{aligned} \quad (22)$$

where $g(y; \psi)$ is the likelihood function of the Gaussian importance model and $\omega(\alpha, y; \psi)$ is the importance weight function:

$$\begin{aligned} \omega(\alpha, y; \psi) &= p(y, \alpha; \psi) / g(y, \alpha; \psi) \\ &= p(y|\alpha; \psi) / g(y|\alpha; \psi). \end{aligned} \quad (23)$$

By generating S independent trajectories $\alpha^{(1)}, \dots, \alpha^{(S)}$ from the importance density $g(\alpha|y; \psi)$, we can estimate the likelihood function by computing

$$\widehat{L}(y; \psi) = g(y; \psi) \cdot \bar{\omega}, \quad \bar{\omega} = \frac{1}{S} \sum_{s=1}^S \omega_s, \\ \omega_s = \omega(\alpha^{(s)}, y; \psi), \quad (24)$$

where ω_s is the realized importance weight function in equation (23) for $\alpha = \alpha^{(s)}$. We base our estimations in section V on $S = 100$ simulated trajectories.

The choice of the importance sampling density partly determines the accuracy of the likelihood, estimate equation (24). We follow the approach of Richard and Zhang (2007) and choose an importance sampling density that (approximately) minimizes the variance of equation (24). Koopman et al. (2015) develop a new method to obtain such an efficient importance sampler using a combination of numerical integration techniques and approximating linear state-space methods. We provide the details in the online appendix. To further improve the numerical efficiency of the likelihood estimate, equation (24), we also use the control variables proposed by Koopman et al. (2015). The NAIS method also facilitates the computation of the smoothed estimates of the state vector α_t (see the online appendix for details).

V. Predictive Analysis: A Monte Carlo Study

We conduct a large-scale Monte Carlo study to investigate the predictive performances of state-space, generalized autoregressive score (GAS), and autoregressive conditional moment (ACM) models. We simulate series of observations y_1, \dots, y_n from both parameter-driven and observation-driven data-generation processes (DGPs), estimate the parameters for the parameter-driven and observation-driven models, and analyze forecasts for the time-varying parameter θ_t generated by these different specifications. The one-step-ahead prediction generated by model j at time t is $\mathbb{E}_j(\theta_{t+1}|y_1, \dots, y_t; \widehat{\psi}_j)$, where \mathbb{E}_j denotes the expectation under model j and parameter vector estimate $\widehat{\psi}_j$. In a Monte Carlo study, one can take the simulated values of the time-varying parameter θ_t as the actual realizations of θ_t , which would otherwise be unobserved in an empirical study. Hence, we compare the predictions with the true θ_t that we have simulated. In this way, we are able to accurately measure which models perform best across a range of empirically relevant DGPs.

A. Design of the Monte Carlo Study

In our first experiment, we take different state-space model specifications as DGPs. We consider the nine observation densities listed in table 1. The autoregressive state equation (2) completes the specifications of all parameter-driven models. We draw 1,000 realizations of time series with length $n = 5,000$ for each DGP, where the parameter values for the different DGPs are in table 3. The parameter values are chosen to be typical for what is found in empirical work for these and related models.

TABLE 3.—THE STATE-SPACE AND GAS MODELS AS DGP

DGP Model		State Space, GAS Parameters			
Type	Distribution	δ, d	ϕ, b	σ_η, a	Other
Count	Poisson	0.00	0.98	0.15	
Count	Negative binomial	0.00	0.98	0.15	$k_1 = 4$
Intensity	Exponential	0.00	0.98	0.15	
Duration	Gamma	0.00	0.98	0.15	$k_1 = 1.5$
Duration	Weibull	0.00	0.98	0.15	$k_1 = 1.2$
Volatility	Gaussian	0.00	0.98	0.15	
Volatility	Student's- t	0.00	0.98	0.15	$\nu = 10$
Copula	Gaussian	0.02	0.98	0.10	
Copula	Student's- t	0.02	0.98	0.10	$\nu = 10$

We specify the state-space models as $y_t|\theta_t \sim p(y_t|\Lambda(\alpha_t); \psi)$, $t = 1, \dots, n$, $\alpha_{t+1} = \delta + \phi\alpha_t + \eta_t$, $\eta_t \sim N(0, \sigma_\eta^2)$, $\alpha_1 \sim N(\delta/(1-\phi), \sigma_\eta^2/(1-\phi^2))$. We parameterize the generalized autoregressive score models as $y_t|\theta_t \sim p(y_t|\Lambda(\alpha_t); \psi)$, $t = 1, \dots, n$, $\alpha_{t+1} = d + a s_t + b \alpha_t$, where the $s_t = \mathcal{I}_t^{-1/2} \nabla_t$ is the scaled score from table 2. Table 1 provides the specifications for the observation densities and the parameterizations used to simulate the “true” θ_t 's in our Monte Carlo study.

In each simulation, we use the first 2,500 observations to estimate the parameters for the following model specifications: (a) the correctly specified state-space model; (b) the GAS model based on the same conditional observation density as the DGP, with the appropriate parameter transformation and scaling $S_t = \mathcal{I}_t^{-1/2}$, and the functional form of the parameter update for the model in table 2; (c) the ACM model for the corresponding specification; and (d) in the case of the exponential, gamma, Weibull, and Gaussian models, the robust model variant of the GAS and ACM specifications. We base the robust specifications on the exponential gamma, Weibull gamma, double gamma (see section III), and Student's- t distributions, respectively. These alternative specifications allow for fatter tails and more overdispersion and are therefore more comparable to the parameter-driven specifications (see the discussion in section III). We consider different sample sizes for estimation to investigate the effect of estimation error in our results. For reference, we also compute the predictions for the state-space model using the DGP parameters.

We compute one-step-ahead predictions for the next 2,500 values of θ_t given the estimated model parameters. We therefore consider 2.5 million ($2,500 \times 1,000$) forecasts for each specification. In the case of the state-space model, we estimate the parameters and predict θ_t one step ahead using the NAIS method (see the online appendix for further details). For the gamma and Weibull models, we predict the means for $\theta_t k$ and $\theta_t \Gamma(1 + 1/k)$, respectively, and we compare the predictions with their respective true stimulated counterparts. We measure the accuracy by means of the mean squared error (MSE), in levels and relative to the MSE of the state-space model. We compute the MSE across the 2.5 million forecasts of θ_t .

For the second experiment, we adopt the GAS model as the DGP. We consider the nine observation densities in table 1 and the update step of equation (3). table 2 provides the scaled scores s_t . The other details of the second experiment are the same as those for the first experiment. The parameter values for the GAS models are in table 3.

TABLE 4.—RESULTS FOR STATE-SPACE MODEL AS DGP (RELATIVE MSEs)

DGP Model		State Space		GAS		ACM	
Type	Distribution	True	Estimated	(1)	(2)	(1)	(2)
Estimation Based on all 2,500 Observations							
Count	Poisson	0.989	1.000	—	1.000	—	1.081
Count	Negative binomial	0.985	1.000	—	1.009	—	1.017
Intensity	Exponential	0.978	1.000	1.019	1.203	1.115	1.268
Duration	Gamma	0.985	1.000	1.004	1.047	1.037	1.035
Duration	Weibull	0.987	1.000	1.005	1.050	1.047	1.027
Volatility	Gaussian	0.983	1.000	1.022	1.079	1.047	1.046
Volatility	Student's- t	0.977	1.000	—	1.020	—	1.105
Copula	Gaussian	0.967	1.000	—	1.011	—	1.304
Copula	Student's- t	0.958	1.000	—	1.002	—	1.390
Estimation Based on Last 500 Observations							
Count	Poisson	0.909	1.000	—	1.013	—	1.101
Count	Negative binomial	0.887	1.000	—	1.018	—	1.004
Intensity	Exponential	0.871	1.000	1.050	1.206	1.185	1.304
Duration	Gamma	0.911	1.000	1.000	1.052	0.986	0.984
Duration	Weibull	0.914	1.000	1.011	1.069	0.985	0.979
Volatility	Gaussian	0.871	1.000	1.028	1.123	1.017	1.020
Volatility	Student's- t	0.853	1.000	—	1.032	—	1.081
Copula	Gaussian	0.878	1.000	—	1.024	—	1.208
Copula	Student's- t	0.908	1.000	—	1.011	—	1.334

We draw 1,000 realizations of time series length $n = 5,000$ for the state-space DGPs of table 3. In the upper panel, we use the first 2,500 observations, and in the lower panel we use the last 500 of the first 2,500 observations, for the estimation of the parameters of the correct model specification, a robust GAS model, column 1, and only for some DGPs, a GAS model based on the same observation density as the DGP, column (2), a robust autoregressive conditional moment (ACM) model, column (1) and only for some DGPs, and an ACM model based on the same observation density as the DGP, column (2). The robust GAS and ACM models are based on the mixture models of section III for the exponential, gamma, and Weibull densities, and the Student's- t model for the Gaussian volatility model. We compute one-step-ahead out-of-sample predictions for the next 2,500 values of the time-varying parameter θ_t (or the means $\theta_t k$ and $\theta_t \Gamma(1 + 1/k)$ for the gamma and Weibull models, respectively) using the models with the true and estimated parameters. We report the MSEs relative to those of the state-space model with estimated parameters.

B. Results for the State-Space Model as DGP

Table 4 presents the results for the state-space DGPs. We focus on two key findings. First, we notice that the differences in forecasting accuracy between the estimated state-space models and the corresponding estimated GAS models are small. We find that when the estimation is based on 2,500 observations, the effect of estimation error is small and the MSE increase for the GAS specifications is around 1% or less for six of the models and less than 2.5% for the exponential and volatility specifications. The intuition for this result follows from our discussions in sections IIC and III. GAS models construct an update step that accounts for the full density information. The score-based observation-driven models are therefore able to generate accurate forecasts when the assumed observation densities approximate the conditional densities implied by the state-space models relatively well. Similar results hold the sample size for estimation at 500, even though we observe a small increase in the relative MSEs of the GAS model in this setting.

The second finding is that the GAS specifications lead to large gains in forecasting performance over ACM models for the exponential, Student's- t volatility, Gaussian copula, and Student's- t copula models. The GAS models also outperform the ACM specifications for the other models, but by a small margin. For the exponential and Student's- t volatility models, the result is due to the fact that the ACM updates are sensitive to realizations from the tails of their distributions. The heavy-tailed (robust) GAS models overcome these problems by incorporating the fat-tailed nature of the error distribution in the update step for α_t . Interestingly, we find

that the ACM models may have a better relative performance when the estimation sample size is small. In the case of the gamma and Weibull models, it outperforms the correct specification by a small margin.

We can also provide further insight for the copula models. We learn from table 2 that the GAS copula update is given by

$$\mathcal{I}_t^{-1/2} \nabla_t = \frac{(1 + \rho_t^2)(\Phi^{-1}(y_{1t})\Phi^{-1}(y_{2t}) - \rho_t) - \rho_t(\Phi^{-1}(y_{1t})^2 + \Phi^{-1}(y_{2t})^2 - 2)}{\sqrt{(1 + \rho_t^2)(1 - \rho_t^2)}} \quad (25)$$

(see Creal et al., 2013). The ACM update is $\Phi^{-1}(y_{1t})\Phi^{-1}(y_{2t})$, which is as a conditionally unbiased estimator of ρ_t (see Patton, 2006). Creal et al. (2013) argue that the ACM driver can be sensitive to large realizations of $\Phi^{-1}(y_{1t})$ or $\Phi^{-1}(y_{2t})$. They consider two possible scenarios for illustrative purposes: $\Phi^{-1}(y_{1t}) = 1$ and $\Phi^{-1}(y_{2t}) = 1$ or, alternatively, $\Phi^{-1}(y_{1t}) = 0.25$ and $\Phi^{-1}(y_{2t}) = 4$. The ACM update is the same for the two scenarios, although the second scenario provides a weaker justification for a large correlation update. In contrast, the GAS update is able to separate the two possibilities through the presence of the adjustment term $-\rho_t(\Phi^{-1}(y_{1t})^2 + \Phi^{-1}(y_{2t})^2 - 2)$, which discounts the large realization of $\Phi^{-1}(y_{2t})$.

We have used a rather long sample size for parameter estimation so that the reported results are not prone to small-sample variations. We therefore regard our findings as reliable and not dependent on random number-generating methods and nonaccurate parameter estimates. However, to

TABLE 5.—RESULTS FOR GAS MODEL AS DGP (RELATIVE MSEs AND MSEs)

DGP Model		Relative Mean Squared Error			Mean Squared Error		
Type	Distribution	State Space	GAS	ACM	State Space	GAS	ACM
Count	Poisson	2.888	1.000	9.187	0.012	0.004	0.038
Count	Negative binomial	1.192	1.000	3.838	0.008	0.006	0.024
Intensity	Exponential	5.849	1.000	4.959	0.048	0.008	0.041
Duration	Gamma	6.026	1.000	3.181	0.123	0.020	0.065
Duration	Weibull	7.614	1.000	5.217	0.050	0.007	0.034
Volatility	Gaussian	8.039	1.000	6.253	0.180	0.022	0.140
Volatility	Student's- <i>t</i>	1.994	1.000	3.426	0.057	0.029	0.098
Copula	Gaussian	1.540	1.000	3.812	0.002	0.002	0.006
Copula	Student's- <i>t</i>	1.175	1.000	5.490	0.002	0.002	0.010

We draw 1,000 realizations of time series length $n = 4,000$ for the GAS DGPs from table 3. We use the first 2,000 observations to estimate the parameters for three statistical models: the correct specification, the state-space specification with the same observation density as the DGP, and the autoregressive conditional moment (ACM) specification. The ACM updates are in table 2. We compute one-step-ahead out-of-sample predictions for the next 2,000 values of θ_t (or the means $\theta_t k_1$ and $\theta_t \Gamma(1 + k_1^{-1})$ for the gamma and Weibull models, respectively). We report the MSEs relative to those of the GAS model in the left-hand panel and the MSEs themselves in the right-hand panel.

provide some insight into the more realistic case that we have a smaller sample size for estimation, we report in the lower panel of table 4 the results when we use only the last 500 of our first 2,500 observations for estimation. The reported relative MSEs are somewhat more prone to small-sample estimation issues, but the overall set of results presents a similar picture. It has been expected that the use of the true parameters in the state-space model leads to a more pronounced reduction in MSEs when the sample size for estimation is smaller. The ACM duration models are able to produce smaller MSEs compared to the true state-space model but with estimated parameters. But apart from these specific findings, the earlier conclusions remain valid for an estimation sample of smaller length.

In the online appendix, we also report the associated model confidence sets (MCS) of Hansen, Lunde, and Nason (2011) for different sample sizes. For all model settings, the GAS model is a true competitor to the state-space model in the number of times that it is present in the 90% MCS. This finding applies especially to the count, intensity, and Student's-*t* volatility models for sample sizes $n \geq 500$. For both copula models, the GAS framework convincingly outperforms ACM. These results confirm our main conclusions presented here.

C. Results for GAS Model as DGP

Table 5 presents the results for the GAS DGPs. Since the time-varying parameters are perfectly predictable under the true DGP, the forecasting errors for the GAS specification are due only to parameter estimation error. Hence, the correct specification strongly outperforms the other two models in relative terms. We focus on the actual differences in mean squared errors since it is common practice; due to space limitations, we report results for other loss functions such as mean absolute errors in the online appendix. Our choices of parameters in table 3 imply that the transformed parameters α_t have the same unconditional means and variances and the same persistence as the states in the parameter-driven DGPs of table 3. The properties of the time-varying processes underlying the results in tables 4 and 5 are therefore similar.

A direct comparison between tables 4 and 5 is not justified given that the update step for the GAS and ACM models is (conditionally) a fixed function, while it is a stochastic function for the state-space models. We observe that the results vary substantially for different DGPs. The state-space models perform poorly overall for the gamma duration and Gaussian volatility models in terms of MSE. For the other specifications, and particularly in comparison with the ACM models, the differences are small. In some cases, the state-space models are producing smaller MSEs. For example, the results for the negative binomial and Student's-*t* models support our discussion in section IIC: the state-space models generate better or similar predictions if the GAS observation density is fat tailed, such as for the mixture models. Table 5 also shows that the forecasting performances of the ACM models are much less favorable compared to those of the GAS models. This result further stresses the distinction between GAS and ACM models. Most of these findings are confirmed in terms of model confidence sets and for different sample sizes (see the presented results in the online appendix).

VI. Empirical Example

To illustrate the relevance of our Monte Carlo findings for empirical applications, we analyze the performance of different parameter-driven and observation-driven models for predicting the log volatility of stock returns and stock index returns. We consider eight specifications with Student's-*t* observation densities: stochastic volatility and GAS models with and without leverage effects, the GARCH model, the GJR-GARCH model of Glosten, Jagannathan, and Runkle (1993), and the EGARCH model of Nelson (1991) with and without leverage terms.

A. Description of the Models

We specify the dynamics of the latent variable as a sum of two components in each of these models. In the case of the SV model, the log volatility follows a sum of two autoregressive processes of order 1. For the observation-driven models, we consider sums of two GAS(1,1), GARCH(1,1),

GJR-GARCH(1,1), and EGARCH(1,1) factors. Engle and Lee (1999) originally proposed the two-factor GARCH model; the extension for other observation-driven models is straightforward. We can regard the two-factor specifications as approximate long memory models that are able to account for the slow decays observed in the empirical autocorrelations of squared and absolute returns and realized volatilities.

The Student's- t observation density is given in table 1. The two-factor GAS model specifies the dynamic equations for the log volatility as

$$\begin{aligned}\log \sigma_t^2 &= d + \alpha_{1,t} + \alpha_{2,t}, \\ \alpha_{1,t+1} &= a_1 s_t + b_1 \alpha_{1,t} + c_1 (y_t/\sigma_t), \\ \alpha_{2,t+1} &= a_2 s_t + b_2 \alpha_{2,t} + c_2 (y_t/\sigma_t),\end{aligned}$$

with the restriction that $-1 < b_2 < b_1 < 1$ for identification. In the model without leverage, we impose the restriction that $c_1 = c_2 = 0$. The GAS update s_t for the Student's- t volatility model is with respect to $\alpha_t = \log \sigma_t^2$ and therefore the same as reported in table 2. Further details of estimation and forecasting are provided in section IVA.

The Student's- t two-factor stochastic volatility model is formulated with the same observation density given in table 1, but with the signal and the state update steps given by

$$\begin{aligned}\log \sigma_t^2 &= \delta + \alpha_{1,t} + \alpha_{2,t}, \\ \alpha_{1,t+1} &= \phi_1 \alpha_{1,t} + \eta_{1,t}, & \eta_{1,t} &\sim N(0, \sigma_{1,\eta}^2), \\ \alpha_{2,t+1} &= \phi_2 \alpha_{2,t} + \eta_{2,t}, & \eta_{2,t} &\sim N(0, \sigma_{2,\eta}^2),\end{aligned}$$

with $-1 < \phi_2 < \phi_1 < 1$. Let $\varepsilon_t = y_t/\sigma_t$ be the return innovation. We introduce leverage effects by letting ε_t be negatively correlated with the volatility disturbances $\eta_{1,t}$ and $\eta_{2,t}$. Define the vector of transformed disturbances $\zeta_t = (\Phi^{-1}[T(\varepsilon_t)], \eta_{1,t}/\sigma_{1,\eta}, \eta_{2,t}/\sigma_{2,\eta})'$, where $T_v(\cdot)$ denotes the CDF of a standardized Student's- t random variable with v degrees of freedom and $\Phi^{-1}(\cdot)$ the inverse standard normal CDF. We assume that

$$\zeta_t \sim N \left(0, \begin{bmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & 0 \\ \rho_2 & 0 & 1 \end{bmatrix} \right).$$

Based on this specification of the stochastic volatility model, we have adopted the importance sampling method NAIS for parameter estimation by maximum likelihood and for signal forecasting (some details and further references are given in section IVB).

B. Data Description and Forecast Precision Measures

Our data consist of daily and high-frequency prices for twenty stocks from the Dow Jones index between January 1993 and June 2012 and five major stock indexes between January 1996 and October 2012. We estimate the parameters for all eight models based on the daily close-to-close returns.

We compute one-step-ahead forecasts starting in 2001 and 2004 for the stocks and indexes, respectively. This choice implies that we use approximately 2,000 observations for initial estimation and 2,900 (2,200 for the indexes) observations for predictive evaluation. For each model, the parameters are reestimated on a daily basis in an expanding window including all previous daily returns. (In the online appendix, we replicate the forecasting exercise using a rolling window scheme and obtain similar results and conclusions.)

The precision of the forecasts from a model is evaluated by comparing the log volatility forecasts with the log of the daily realized volatilities as measured from high-frequency data. We focus here on the transformed parameters in order to make the model comparison less sensitive to short-term volatility peaks observed in the data, which make the prediction error distributions pronouncedly fat tailed and skewed to the right. We adopt the realized kernel of Barndorff-Nielsen et al. (2008) for the measurement of daily open-to-close variance of the stock and index returns. Since the model-implied volatility predictions are for daily volatility, while we compute the realized kernel over trading hours, we cannot compare these quantities directly. Hence we measure the forecast precision in terms of the variance of the residuals from standard Mincer-Zarnowitz regressions.

C. Empirical Results

Table 6 reports the Mincer-Zarnowitz residual variances as ratios with respect to those from the SV model with leverage effects, our choice of benchmark model. The results are consistent with those we have obtained in the Monte Carlo study. We find that SV and GAS models have very similar performances for all stocks and indexes except two, whether or not we consider leverage effects. When we consider leverage effects, the relative difference in variance between these models is 2%, or less for 21 out of 25 series. In contrast, the GARCH and GJR-GARCH models appear to be less robust, leading to markedly higher forecasting variances in some cases. The EGARCH model performs better than the other GARCH models, achieving a similar performance to the SV and GAS in a majority of cases. These results indicate that the exponential link function by itself can play a significant role in improving forecasting. However, we also find that the SV and GAS models lead to important gains over the EGARCH model in some other cases.

We present a summary of the results in the last rows of table 6. We report the proportion of series in which a model specification is among the best-performing models, meaning it appears in the 90% model confidence set (considering all models or comparing within the groups with and without leverage effects). The model confidence sets are proposed by Hansen et al. (2011), and their construction in our study is detailed in the online appendix. The models with leverage effects are clearly preferred over the ones without leverage, especially for the stock indexes. Focusing on leverage effects, the GAS model achieves the lowest forecasting

TABLE 6.—RESULTS FOR THE EMPIRICAL EXAMPLE

Stock/Index	No Leverage				Leverage			
	SV	GAS	GARCH	EGARCH	SV	GAS	GJR	EGARCH
Alcoa	1.08	1.09	1.08	1.08	1.00	0.98	1.00	1.02
American Express	1.10	1.10	1.11	1.10	1.00	0.99	0.99	0.99
Boeing	1.07	1.06	1.09	1.07	1.00	0.98	0.98	1.00
Caterpillar	1.12	1.12	1.22	1.13	1.00	0.98	1.19	1.06
Chevron	1.11	1.10	1.12	1.10	1.00	1.00	1.05	1.00
Walt Disney	1.12	1.15	1.16	1.16	1.00	1.02	1.03	1.07
General Electric	1.07	1.05	1.10	1.07	1.00	1.00	1.08	1.02
IBM	1.15	1.12	1.23	1.14	1.00	0.99	1.13	1.01
Intel	1.02	1.01	1.04	1.03	1.00	0.99	1.04	1.02
Johnson & Johnson	1.06	1.07	1.08	1.06	1.00	1.00	1.06	1.06
JPMorgan	1.06	1.07	1.05	1.08	1.00	1.00	0.98	1.02
Coca-Cola	1.04	1.03	1.06	1.04	1.00	0.97	1.00	0.98
McDonald's	1.06	1.09	1.10	1.10	1.00	1.03	1.05	1.13
Merck	1.04	1.02	1.31	1.11	1.00	1.06	1.30	1.20
Microsoft	1.07	1.06	1.16	1.09	1.00	0.99	1.12	1.04
Pfizer	1.03	1.00	1.10	1.03	1.00	1.00	1.13	1.11
Procter & Gamble	1.05	1.06	1.05	1.05	1.00	0.99	1.01	0.99
AT&T	1.05	1.05	1.07	1.06	1.00	1.01	1.02	1.03
Walmart	1.04	1.03	1.06	1.03	1.00	1.00	1.05	1.01
ExxonMobil	1.10	1.10	1.10	1.11	1.00	1.01	1.10	1.01
DAX	1.26	1.25	1.22	1.24	1.00	0.98	1.08	0.99
FTSE 100	1.22	1.18	1.17	1.18	1.00	1.12	1.03	1.14
NASDAQ	1.22	1.22	1.21	1.24	1.00	1.00	0.98	1.01
Nikkei 225	1.08	1.09	1.11	1.11	1.00	0.98	1.05	1.07
S&P 500	1.22	1.23	1.22	1.21	1.00	0.98	0.98	0.98
Average	1.10	1.09	1.13	1.10	1.00	1.00	1.06	1.04
Best model	0.00	0.04	0.00	0.00	0.32	0.48	0.12	0.04
In 90% MCS	0.16	0.20	0.00	0.12	0.88	0.88	0.56	0.48
In 90% MCS (by group)	0.68	0.84	0.60	0.60	0.88	0.88	0.44	0.40

The table presents the relative performance of different two-factor models with Student's- t return innovations. We have estimated the parameters using daily returns and have used these to forecast daily volatilities one step ahead. The performance is measured as the variance of the residuals of Mincer-Zarnowitz regressions of the realized volatilities on the daily volatility forecasts for each specification. We report the values relative to those of the SV model with leverage. In the last four rows, we report the average performance, the proportion in which a specification is the best-performing model, the proportion in which it appears in a 90% model confidence set (MCS) over all eight specifications, and the same proportion but then over all four specifications in each group ("No Leverage" and "Leverage" models).

errors for 48% of the series, closely followed by the SV model (32%). The 90% model confidence set includes the SV, GAS, and EGARCH models for 88%, 88%, and 48% of the series, respectively. The most common scenario is therefore one in which we cannot statistically select the SV or GAS model as the best model; they appear together in the MCS in nineteen series. At least one of them lies in the MCS for all series except one. The EGARCH model is often competitive but is rejected for half of the series in favor of the SV and/or GAS models.

We report additional results in the online appendix for unconditional forecasts and other loss functions. We also repeated our empirical study using parameter estimates and one-step-ahead forecasts from a rolling window of 2,000 observations instead of the expanding window of table 6. The results are very similar, and the overall conclusions remain. However, in the rolling window study, the GARCH model and the other volatility models without leverage appear more regularly in the 90% MCS. The GAS model, however, is still most present in the 90% MCS within the "No Leverage" group. The same comments apply to the results for the mean absolute error, presented in the online appendix. An additional difference with table 6 is that the SV model with leverage is best for 52% of the cases in terms of mean absolute error compared to 32% in terms of MSE. This finding may indicate that the SV with leverage model is more

robust to outlying observations than other volatility model specifications.

VII. Conclusion

We have studied the forecasting performance of three different classes of time-varying parameter models. We have considered nonlinear non-Gaussian state-space models as representatives of parameter-driven models, generalized autoregressive score models as flexible representatives of observation-driven models, and autoregressive conditional moment models such as the well-known GARCH and the autoregressive conditional duration models. Our results are applicable to a large range of specifications for count, intensity, duration, volatility, and dependence models.

The state-space and GAS specifications lead to similar predictive performances if the data-generating process is the state-space model. This holds particularly if the observation density in the GAS specification is sufficiently flexible to approximate the conditional distribution implied by the state-space model. If the DGP is the GAS model, the forecasting performance of state-space models sometimes decreases when compared to GAS models. We conclude that GAS models provide a competitive alternative to state-space models from a forecasting perspective at a reduced computational cost. The practical advantage for this model stems from

the fact that the likelihood function is available in closed form. The analysis based on this models therefore does not require the use of computational-intensive simulation methods.

We have also established that GAS models often lead to important forecasting gains over ACM models, including GARCH and dynamic conditional correlation models. ACM models are typically intuitively based on moment conditions derived from the conditional distribution of the observations. However, our results show that they possibly miss key information about the observation density when updating the time-varying parameters. Our evidence therefore shows that GAS models are a useful new tool for forecasting.

The findings rely mostly on our large-scale Monte Carlo study. We have verified these findings with an empirical study for twenty individual stocks and five major stock indices. The results are consistent with those obtained from our Monte Carlo study. The SV and GAS models have very similar forecasting performances for all series considered.

REFERENCES

- Barndorff-Nielsen, Ole E., Peter Reinhard Hansen, Asger Lunde, and Neil Shephard, "Designing Realised Kernels to Measure the Ex-Post Variation of Equity Prices in the Presence of Noise," *Econometrica* 76 (2008), 1481–1536.
- Bauwens, Luc, and Nikolaus Hautsch, "Stochastic Conditional Intensity Processes," *Journal of Financial Econometrics* 4 (2006), 450–493.
- Bauwens, Luc, and Davide Veredas, "The Stochastic Conditional Duration Model: A Latent Factor Model for the Analysis of Financial Durations," *Journal of Econometrics* 119 (2004), 381–412.
- Blasques, Francisco, Siem Jan Koopman, and André Lucas, "Information Theoretic Optimality of Observation Driven Time Series Models for Continuous Responses," *Biometrika* 102 (2015), 325–343.
- Bollerslev, Tim, "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics* 21 (1986), 307–328.
- , "A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return," this REVIEW 69 (1987), 542–547.
- Carnero, M. Angeles, Daniel Peña, and Esther Ruiz, "Persistence and Kurtosis in GARCH and Stochastic Volatility Models," *Journal of Financial Econometrics* 2 (2004), 319–342.
- Cox, David R, "Statistical Analysis of Time Series: Some Recent Developments," *Scandinavian Journal of Statistics* 8 (1981), 93–115.
- Creal, Drew D., Siem Jan Koopman, and André Lucas, "A Dynamic Multivariate Heavy-Tailed Model for Time-Varying Volatilities and Correlations," *Journal of Business and Economic Statistics* 29 (2011), 552–563.
- , "Generalized Autoregressive Score Models with Applications," *Journal of Applied Econometrics* 28 (2013), 777–795.
- , "Observation Driven Mixed-Measurement Dynamic Factor Models with an Application to Credit Risk," this REVIEW 96 (2014), 898–915.
- Das, Sanghamitra, and Krishna Srinivasan, "Duration of Firms in an Infant Industry: The Case of Indian Computer Hardware," *Journal of Development Economics* 53 (1997), 157–167.
- Durbin, James, and Siem Jan Koopman, "Monte Carlo Maximum Likelihood Estimation for Non-Gaussian State Space Models," *Biometrika* 84 (1997), 669–684.
- , "Time Series Analysis of Non-Gaussian Observations based on State Space Models from Both Classical and Bayesian Perspectives," *Journal of the Royal Statistical Society, Series B* 62 (2000), 3–56.
- , "A Simple and Efficient Simulation Smoother for State Space Time Series Analysis," *Biometrika* 89 (2002), 603–616.
- , *Time Series Analysis by State Space Methods*, 2nd ed. (Oxford: Oxford University Press, 2012).
- Engle, Robert F., "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica* 50 (1982), 987–1007.
- , "Dynamic Conditional Correlation," *Journal of Business and Economic Statistics* 20 (2002), 339–350.
- Engle, Robert F., and Giampiero Gallo, "A Multiple Indicators Model for Volatility Using Intra-Daily Data," *Journal of Econometrics* 131 (2006), 3–27.
- Engle, Robert F., and Gary J. Lee, "A Long-Run and Short-Run Component Model of Stock Return Volatility" (pp. 475–497), in Robert F. Engle and Halbert L. White, eds., *Cointegration, Causality, and Forecasting: A Festschrift in Honour of Clive W. J. Granger* (Oxford: Oxford University Press, 1999).
- Engle, Robert F., and Jeffrey R. Russell, "Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data," *Econometrica* 66 (1998), 1127–1162.
- Ghysels, Eric, Andrew C. Harvey, and Eric Renault, "Stochastic Volatility," in Gangadharrao S. Maddala and Calyampudi R. Rao, eds., *Handbook of Statistics*, vol. 14 (Amsterdam: Elsevier, 1996).
- Glosten, Lawrence R., Ravi Jagannathan, and David E. Runkle, "On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks," *Journal of Finance* 48 (1993), 1779–1801.
- Grammig, Joachim, and Kai-Oliver Maurer, "Non-Monotonic Hazard Functions and the Autoregressive Conditional Duration Model," *Econometrics Journal* 3 (2000), 16–38.
- Hafner, Christian M., and Hans Manner, "Dynamic Stochastic Copula Models: Estimation, Inference and Applications," *Journal of Applied Econometrics* 27 (2012), 269–295.
- Hansen, Peter Reinhard, Zhuo Huang, and Howard H. Shek, "Realized GARCH: A Joint Model for Returns and Realized Measures of Volatility," *Journal of Applied Econometrics* 27 (2012), 877–906.
- Hansen, Peter Reinhard, and Asger Lunde, "A Forecast Comparison of Volatility Models: Does Anything Beat a GARCH(1,1) Model?" *Journal of Applied Econometrics* 20 (2005), 873–889.
- Hansen, Peter Reinhard, Asger Lunde, and James M. Nason, "The Model Confidence Set," *Econometrica* 79 (2011), 453–497.
- Harvey, Andrew C., *Dynamic Models for Volatility and Heavy Tails* (Cambridge: Cambridge University Press, 2013).
- Harvey, Andrew C., and Alessandra Luati, "Filtering with Heavy Tails," *Journal of the American Statistical Association* 109 (2014), 1112–1122.
- Janus, Pawel, Siem Jan Koopman, and André Lucas, "Long Memory Dynamics for Multivariate Dependence under Heavy Tails," *Journal of Empirical Finance* 29 (2014), 187–206.
- Jungbacker, Borus, and Siem Jan Koopman, "Monte Carlo Estimation for Nonlinear Non-Gaussian State Space Models," *Biometrika* 94 (2007), 827–839.
- Koopman, Siem Jan, André Lucas, and Marcel Scharth, "Numerically Accelerated Importance Sampling for Nonlinear Non-Gaussian State-Space Models," *Journal of Business Economic Statistics* 33 (2015), 114–127.
- Lancaster, Tony, "Econometric Methods for the Duration of Unemployment," *Econometrica* 47 (1979), 939–956.
- Liesenfeld, Roman, and Jean-François Richard, "Univariate and Multivariate Stochastic Volatility Models: Estimation and Diagnostics," *Journal of Empirical Finance* 10 (2003), 505–531.
- Melino, Angelo, and Stuart M. Turnbull, "Pricing Foreign Currency Options with Stochastic Volatility," *Journal of Econometrics* 45 (1990), 239–265.
- Nelson, Daniel B., "Conditional Heteroskedasticity in Asset Returns: A New Approach," *Econometrica* 59 (1991), 347–370.
- Patton, Andrew J., "Modelling Asymmetric Exchange Rate Dependence," *International Economic Review* 47 (2006), 527–556.
- Richard, Jean-François, and Wei Zhang, "Efficient High-Dimensional Importance Sampling," *Journal of Econometrics* 141 (2007), 1385–1411.
- Rydberg, Tina H., and Neil Shephard, "A Modelling Framework for the Prices and Times of Trades Made on the NYSE" (pp. 217–246), in William J. Fitzgerald, Richard L. Smith, Andrew T. Walden, and Peter C. Young, eds., *Nonlinear and Nonstationary Signal Processing* (Cambridge: Cambridge University Press, 2000).

- Shephard, Neil, and Michael K. Pitt, "Likelihood Analysis of Non-Gaussian Measurement Time Series," *Biometrika* 84 (1997), 653–667.
- Shephard, Neil, and Kevin Sheppard, "Realising the Future: Forecasting with High-Frequency-Based Volatility (HEAVY) Models," *Journal of Applied Econometrics* 25 (2010), 197–231.
- Tauchen, George E., and Mark Pitts, "The Price Variability-Volume Relationship in Speculative Markets," *Econometrica* 51 (1983), 485–505.
- Taylor, Stephen J., *Modelling Financial Time Series* (Chichester, UK: Wiley, 1986).

Copyright of Review of Economics & Statistics is the property of MIT Press and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.