Predicting Time-Varying Parameters with Parameter-Driven and Observation-Driven Models

Koopman, S.J.; Lucas, A.; Scharth, M.

2012

TI 2012-020/4
Tinbergen Institute Discussion Paper

Predicting Time-Varying Parameters with Parameter-Driven and Observation-Driven Models

Siem Jan Koopman[a,b]
André Lucas[a,c,b]
Marcel Scharth[a,b]

[a] *VU University Amsterdam, The Netherlands;*
[b] *Tinbergen Institute, The Netherlands;*
[c] *Duisenberg school of finance, The Netherlands.*

Tinbergen Institute is the graduate school and research institute in economics of Erasmus University Rotterdam, the University of Amsterdam and VU University Amsterdam.

More TI discussion papers can be downloaded at http://www.tinbergen.nl

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 525 1600

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

Duisenberg school of finance is a collaboration of the Dutch financial sector and universities, with the ambition to support innovative research and offer top quality academic education in core areas of finance.

DSF research papers can be downloaded at: http://www.dsf.nl/

Duisenberg school of finance
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 525 8579

Predicting time-varying parameters with parameter-driven and observation-driven models*

Siem Jan Koopman[(a,b)] André Lucas[(a,c,b)] Marcel Scharth[(a,b)]

[(a)] VU University Amsterdam, The Netherlands

[(b)] Tinbergen Institute, The Netherlands

[(c)] Duisenberg school of finance, The Netherlands

March 2012

Abstract

We study whether and when parameter-driven time-varying parameter models lead to forecasting gains over observation-driven models. We consider dynamic count, intensity, duration, volatility and copula models, including new specifications that have not been studied earlier in the literature. In an extensive Monte Carlo study, we find that observation-driven generalised autoregressive score (GAS) models have similar predictive accuracy to correctly specified parameter-driven models. In most cases, differences in mean squared errors are smaller than 1% and model confidence sets have low power when comparing these two alternatives. We also find that GAS models outperform many familiar observation-driven models in terms of forecasting accuracy. The results point to a class of observation-driven models with comparable forecasting ability to parameter-driven models, but lower computational complexity.

Keywords : Generalised autoregressive score model, Importance sampling, Model confidence set, Nonlinear state space model, Weibull-gamma mixture.

JEL: C53, C58, C22.

*André Lucas acknowledges the financial support of the Dutch Science Foundation (NWO). Contact author: Marcel Scharth (`m.scharth@vu.nl`).

1 Introduction

In this paper we study the predictive ability of parameter-driven versus observation-driven time-varying parameter models. We consider dynamic count, intensity, duration, volatility and copula densities and focus on three approaches for modelling the time-varying parameters of interest: nonlinear non-Gaussian state space models as representatives of parameter-driven specifications; the flexible observation-driven generalised autoregressive score (GAS) class of Creal, Koopman, and Lucas (2012); and standard observation-driven models based on moments of the data, such as the generalised autoregressive conditional heteroscedasticity (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 this latter set of models under the general heading of autoregressive conditional parameter (ACP) models.

Cox (1981) classifies time-varying parameter models into two classes: observation-driven and parameter-driven specifications. In an observation-driven model, current parameters are deterministic functions of lagged dependent variables as well as contemporaneous and lagged exogenous variables. In this setting, parameters evolve randomly over time but are perfectly predictable one-step-ahead given past information. The likelihood function for observation-driven models is available in closed-form through the prediction error decomposition. This feature leads to simple estimation procedures and has contributed to the popularity of this class of 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 not available in closed-form for these models. 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) and Ghysels, Harvey, and Renault (1996), the stochastic conditional duration model of Bauwens and Veredas (2004), and the stochastic copula models of Hafner and Manner (2011).

Given the different nature of observation-driven and parameter-driven models and the large amount of effort devoted to studying and applying a variety of these specifi-

cations, 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). We contribute to the literature by providing solutions to each of these problems, thus enabling a full-scale 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 parametrisation, 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 modelling, in many other settings it may not be evident.

The generalised autoregressive score (GAS) model of Creal, Koopman, and Lucas (2012) is a class of observation-driven models with similar degree of generalisability as nonlinear non-Gaussian state space models. The GAS framework uses the scaled score vector of the predictive model density to update time-varying parameters. We can thus apply the GAS model to any observation density. Creal, Koopman, and Lucas (2012) show that the GAS class encompasses well-known observation-driven models such as the GARCH model of Bollerslev (1986), while at the same time enabling the development of completely new models such as the mixed measurement dynamic factor model of Creal, Schwaab, Koopman, and Lucas (2011). The GAS framework therefore provides a natural observation-driven alternative for the state space framework across a wide range of different DGPs.

Second, observation-driven and parameter-driven models are inherently difficult 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 observation 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 generalised 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, the estimation of parameters in 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 (2011). 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, which is the prime focus of our current study.

We obtain two main findings. First, when the DGP is a state space model, the predictive accuracy of an (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 square error from using a GAS model instead of the correct state space specification is most of the time inferior to 1% and never higher than 2.5%. We extend our analysis by considering the model confidence sets of Hansen, Lunde, and Nason (2011). For the state space DGPs, the GAS model lies in the 90% model confidence set for at least 60% of the samples with as many as $2,000$ observations. 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. In most cases, an observation-driven alternative is available that is both accurate and considerably easier to estimate.

Second, we find that the GAS models outperform many of the familiar observation-driven models from the literature which we have referred to as autoregressive conditional parameter (ACP) 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 ACP models. GAS models are therefore effective new tools for forecasting that often lead to important forecasting gains over other classes of observation-driven models.

We structure the rest of this paper as follows. In Section 2 we present our three econometric approaches for modelling time-varying parameters. Section 3 introduce several new GAS models for continuous mixtures. Section 4 discusses the estimation of parameters for the different model classes. Section 5 presents the results. Section 6 concludes.

2 Modelling time-varying parameters

2.1 Dynamic model specifications

Let $y_1, \ldots, y_n$ 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 modelling 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

$$y_t|\theta_t \sim p(y_t|\theta_t;\psi), \qquad \theta_t = \Lambda(\alpha_t), \qquad t = 1, \ldots, n, \tag{1}$$

where θ_t is a time-varying parameter vector, $\Lambda(\cdot)$ is a possibly nonlinear 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 $p(y_t|\theta_t;\psi)$.

2.1.1 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 one

$$\alpha_{t+1} = \delta + \phi\alpha_t + \eta_t, \qquad \alpha_1 \sim \mathrm{N}(a_1, P_1), \qquad \eta_t \sim \mathrm{N}(0, \sigma_\eta^2), \tag{2}$$

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) characterise a class of nonlinear non-Gaussian state space models; see Durbin and Koopman (2001) for a general discussion. More generally, α_t can also follow higher order autoregressive moving average, random walk, cyclical, seasonal and other processes or be 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, Lucas, and Scharth (2011) report recent developments on Monte Carlo methods for the analysis of general nonlinear non-Gaussian state space models.

Examples of specifications within this framework include stochastic volatility models as in Tauchen and Pitts (1983), Taylor (1986), Melino and Turnbull (1990) and Ghysels, Harvey, and Renault (1996), stochastic conditional duration models as in Bauwens and Veredas (2004), stochastic conditional intensity models as in Bauwens and Hautsch (2006), stochastic copulas as in Hafner and Manner (2011), and non-Gaussian unobserved components time series models as in Durbin and Koopman (2000).

2.1.2 Generalised autoregressive score models

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

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

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 for the dynamic specification (3) which are similar to the ones available for the state space model.

Specific choices for the driving mechanism s_t lead to different classes of observation-driven models. In this paper we focus on the generalised autoregressive score (GAS) class of Creal, Koopman, and Lucas (2012) as our central observation-driven model. The GAS framework is of similar generality as the state space model (1)–(2) in that it is applicable to any measurement density. In this framework, the updating step s_t in (3) is the scaled density score

$$s_t = S_t \cdot \nabla_t, \qquad \nabla_t = \frac{\partial \ln p(y_t \,|\, \alpha_t \,,\, \mathcal{F}_t \,;\, \psi)}{\partial \alpha_t}, \qquad S_t = S(t \,,\, \alpha_t \,,\, \mathcal{F}_t \,;\, \theta), \tag{4}$$

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, Koopman, and Lucas (2012) discuss appropriate choices for S_t based on the curvature of the log-density at time t as summarised by the Fisher information matrix

$$\mathcal{I}_t = \mathbb{E}\left[\nabla_t \nabla_t' | \mathcal{F}_{t-1}\right], \tag{5}$$

therefore linking the scaling matrix to the variance of the score. In this paper we focus on the scaling matrix $S_t = \mathcal{I}_t^{-1/2}$. For this choice of scaling the step s_t has constant unit variance and is invariant under non-degenerate parameter transformations $\Lambda(\cdot)$. The constant unit variance property is an 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, Koopman, and Lucas (2012). We can also specify the observation-driven dynamics directly for θ_t.

An useful feature of the GAS approach is the automatic treatment of parameter transformations $\Lambda(\cdot)$. 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 $\mathcal{I}_t$ which does not depend on α_t, such that the choice of scaling matrix S_t becomes irrelevant. In this paper we only report the results for GAS models with appropriate parameter transformations and scaling matrix $S_t = \mathcal{I}_t^{-1/2}$. We have found that these specifications lead to greater predictive accuracy across different DGPs. The evidence for a selection of alternative specifications and parametrisations is available upon request.

2.1.3 Autoregressive conditional parameter models

Many observation-driven models directly relate the time-varying parameter to a natural transformation of the data. A common approach is to define s_t such that

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

We refer to this class as autoregressive conditional parameter (ACP) models.

ACP models adhere to the intuitive notion that the observation-driven parameter should increase (decrease) if the realised s_t is higher (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 ACP update is $s_t = y_t$. Similarly, if θ_t is the conditional variance of y_t, then $s_t = (y_t - \mu_{y,t-1})^2$ with $\mu_{y,t-1} = \mathbb{E}(y_t|\mathcal{F}_{t-1})$.

Examples of autoregressive conditional parameter models include the generalised autoregressive conditional heteroscedasticity (GARCH) model of Bollerslev (1986), the autoregressive conditional duration (ACD) and intensity (ACI) models of Engle and Russell (1998), the autoregressive conditional Poisson model of Rydberg and Shephard (2000), the dynamic conditional correlation model of Engle (2002), specific autoregressive copulas in Patton (2006), and the HEAVY model of Shephard and Sheppard (2010). Due to their widespread use, the class of ACP models provides a useful benchmark to our analysis.

2.2 Observation densities

We present the observation densities for our study below in Table 1. These densities represent $p(y_t|\theta_t; \psi)$ in (1). We consider a wide range of specifications, including densities for count, intensity, duration, volatility and copula models. The combination

Table 1: Observation densities.

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 θ_t is the time-varying parameter of interest, and $\Lambda(.)$ 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, $z_{i,t} = \Phi^{-1}(y_{i,t})$, where the observations $y_{i,t}$ have uniform $(0,1)$ marginal distributions and $\Phi^{-1}(\cdot)$ denotes the inverse normal CDF. For the Student t copula, $z_{i,t} = T_\nu^{-1}(y_{i,t})$, where the observations $y_{i,t}$ have uniform $(0,1)$ marginal distributions and $T_\nu^{-1}(y_{i,t})$ is the inverse CDF of a Student's t distribution with ν degrees of freedom.

Model type	Distribution	Density	Parameterisation
Count	Poisson	$\frac{\lambda_t^{y_t}}{y_t!}e^{-\lambda_t}$	$\lambda_t = \exp(\alpha_t)$
Count	Neg. 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 t	$\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{(\nu-2)\pi}\Gamma\left(\frac{\nu}{2}\right)\sigma_t}\left(1+\frac{y_t^2}{(\nu-2)\sigma_t^2}\right)^{-\frac{\nu+1}{2}}$	$\sigma_t^2 = \exp(\alpha_t)$
Copula	Gaussian	$\frac{\frac{1}{2\pi\sqrt{1-\rho_t^2}}\exp\left[\frac{z_{1t}^2+z_{2t}^2-2\rho_t z_{1t}z_{2t}}{2(1-\rho_t^2)}\right]^\dagger}{\prod_{i=1}^2\frac{1}{\sqrt{2\pi}}e^{-z_{it}^2/2}}$	$\rho_t = \frac{1-\exp(-\alpha_t)}{1+\exp(-\alpha_t)}$
Copula	Student's t	$\frac{\Gamma(\frac{\nu+2}{2})\Gamma(\frac{\nu}{2})}{\Gamma(\frac{\nu+1}{2})}\frac{\frac{1}{\sqrt{1-\rho_t^2}}\left[1+\frac{z_{1t}^2+z_{2t}^2-2\rho_t z_{1t}z_{2t}}{\nu(1-\rho_t^2)}\right]^{-\frac{\nu+2}{2}\ddagger}}{\prod_{i=1}^2(1+z_{it}/\nu)^{-\frac{\nu+1}{2}}}$	$\rho_t = \frac{1-\exp(-\alpha_t)}{1+\exp(-\alpha_t)}$

of the dynamic gamma, Weibull, normal, Student's t and copula densities in Table 1 with (2) directly lead 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 (2011).

Table 2 completes the specification of the observation-driven models by listing the generalised autoregressive score and autoregressive conditional parameter updates s_t for the densities in Table 1. The ACP updates lead to the familiar autoregressive conditional Poisson, autoregressive conditional duration, autoregressive conditional intensity, GARCH and autoregressive copula specifications. The ACP model for the

Table 2: Observation-driven model updates.

The table displays the score and information matrix for the models given in Table 1. The GAS update $s_t = \nabla_t \mathcal{I}_t^{-1/2}$ we consider in this paper is invariant under non-degenerate parameter transformations. The equivalent ACP models are without parameter transformation, see Table 1. The data is 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 $z_{i,t} = T_\nu^{-1}(y_{i,t})$ for the Student t copula, with $\Phi^{-1}(\cdot)$ and $T_\nu^{-1}(y_{i,t})$ denoting the is the inverse normal CDF and the inverse CDF of a Student's t distribution with ν degrees of freedom, respectively.

Model type	Distribution	GAS ∇_t	GAS $\mathcal{I}_t$	ACP s_t
Count	Poisson	$\frac{y_t}{\lambda_t} - 1$	$\frac{1}{\lambda_t}$	y_t
Count	Neg. 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	$\frac{1}{2\sigma_t^2}\left(\frac{\omega_t y_t^2}{\sigma_t^2} - 1\right)$ $\omega_t = \frac{\nu+1}{(\nu-2)+y_t^2/\sigma_t^2}$	$\frac{\nu}{2(\nu+3)\sigma_t^4}$	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	$\frac{(1+\rho^2)(\omega_t\hat{z}_{1,t}-\rho_t)-\rho_t(\omega_t\hat{z}_{2,t}-2)}{(1-\rho^2)^2}$ $\omega_t = \frac{\nu+2}{\nu+\frac{\hat{z}_{1,t}-2\rho_t\hat{z}_{2,t}}{1-\rho^2}}$	$\frac{\left(\nu+2+\nu\rho_t^2\right)}{(\nu+4)(1-\rho_t^2)^2}$	$z_{1,t} z_{2,t}$

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 appear in the original paper by Creal, Koopman, and Lucas (2012). Creal, Koopman, and Lucas (2011) obtain the GAS model for the Student's t copula.

2.3 Parameter-driven versus observation-driven models

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

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

In a parameter-driven framework the estimation of θ_t by construction takes into account the full density structure of past observations. As we discuss below, the same is not necessarily true for observation-driven models. The mixture distribution (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 and (2) is conditionally leptokurtic. Similarly, the stochastic count and duration models we study below display conditional over-dispersion.

On other hand, a major obstacle for the application of parameter-driven models is that $p(\theta_t;\psi|\mathcal{F}_{t-1})$ is typically not available in closed-form. This is the case for all models in Table 1. The 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 characterises the likelihood function of the model; see, for instance, Shephard and Pitt (1997) and Section 4. Monte Carlo methods are also necessary to estimate and forecast the time-varying vector θ_t.

In observation-driven models the time-varying parameter θ_t is perfectly predictable one step ahead given past information. This implies that the likelihood functions for ACP and GAS models are available in closed-form. Parameter estimation is therefore straightforward for observation-driven models, contributing to their widespread use in applied econometrics and statistics. However, the self-referential structure of observation-driven models complicates their theoretical analysis. For example, the stability properties of the sequence of observations, such as stationarity and ergodicity, are typically difficult to derive.

Within the class of observation-driven models, there are important differences between GAS and ACP specifications. GAS models can handle parameter transformations and are applicable in cases where ACP updates are not readily available; see Creal, Koopman, and Lucas (2012). By making use of the observation density score,

GAS model updates also take the full density information into account. By contrast, ACP models rely exclusively on the moments of $p(y_t|\theta_t;\psi)$, such as the mean or the variance.

We follow Creal, Koopman, and Lucas (2012) and illustrate the difference using time-varying volatility models for the Gaussian and Student's t distributions based on the GAS and ACP 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

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

which is equivalent to the standard GARCH$(1,1)$ model (the corresponding ACP model). If we replace the normal distribution by the Student's t distribution with ν degrees of freedom, the ACP model updating equation remains the same such that the ACP model reduces to the GARCH(1,1) model with Student's t distributed errors. However, the GAS update equation becomes

$$s_t = \mathcal{I}_t^{-1}\nabla_t = \left(1+3\nu^{-1}\right)\cdot\left(\frac{(1+\nu^{-1})}{(1-2\nu^{-1})(1+\nu^{-1}y_t^2\ /((1-2\nu^{-1})\ \alpha_t))}y_t^2 - \alpha_t\right), \tag{9}$$

see also Table 2. If $\nu^{-1}\to 0$, the GAS model recovers the GARCH(1,1) specification. However, if ν is finite and observations are fat-tailed, large values of y_t^2 receive less weight due to the presence of the denominator in (9). This feature is intuitively appealing. The GARCH update 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 updating equation for the Student's t distribution discounts large values of y_t^2 in comparison to the Gaussian case.

3 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. These specifications also relate to the simpler negative binomial (Poisson-gamma) GAS model of Table 2. Apart from our current motivation, the new model specifications are of intrinsic interest as new duration and multiplicative error models which combine the flexibility of mixture distributions, robust score based updates and the log parametrisation.

3.1 Weibull-gamma and exponential-gamma mixture models

We consider the following parametrisation of the Weibull distribution

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

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 $\gamma_t = \mu_t\,\nu_t$ where $\alpha_t = \log(\mu_t)$ follows a GAS updating equation and ν_t is an identically and independently $\Gamma(k_2^{-1}, k_2)$ distributed random error with density function

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

The multiplicative error ν_t has mean one and variance $k_2 < \infty$.

The Weibull-gamma mixture density is

$$p(y_t|\mu_t; k) = \int_0^\infty p(y_t|\mu_t, \nu_t; k_1) p(\nu_t)\,\mathrm{d}\nu_t = \mu_t k_1 y_t^{k_1-1} \left(1 + k_2 \mu_t y_t^{k_1}\right)^{-(1+k_2^{-1})}. \tag{12}$$

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, which they refer to as the Burr distribution. They advocate the Weibull-gamma model for the empirical analysis of price durations on the basis of its ability to account for non-monotonic hazard functions.

We notice that

$$\mathbb{E}(y_t|\mu_t;k_1,k_2)=\mu_t^{-k_1^{-1}}\Gamma(k_1^{-1}+1)\mathbb{E}\left(\nu_t^{-k_1^{-1}}\right)=(\mu_t k_2)^{-k_1^{-1}}\frac{\Gamma(k_2^{-1}-k_1^{-1})}{\Gamma(k_2^{-1})}, \tag{13}$$

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 are

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

respectively. This update recovers the Weibull model when $k_2\to 0$. We base the GAS update equation on the scaled score

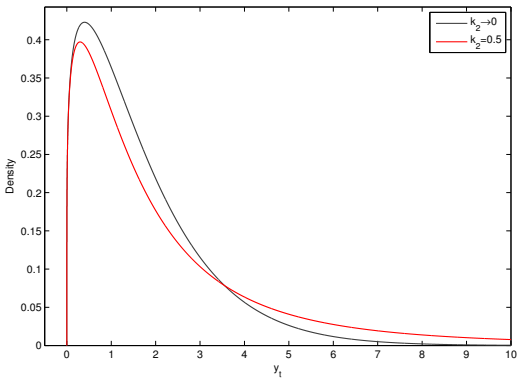
$$s_t=\mathcal{I}_t^{-1/2}\nabla_t=\sqrt{1+2k_2}\left(1-(1+k_2)\frac{\mu_t y_t^{k_1}}{1+k_2\mu_t y_t^{k_1}}\right). \tag{15}$$

By setting $k_1=1$ above, the specification specialises to the exponential-gamma GAS model.

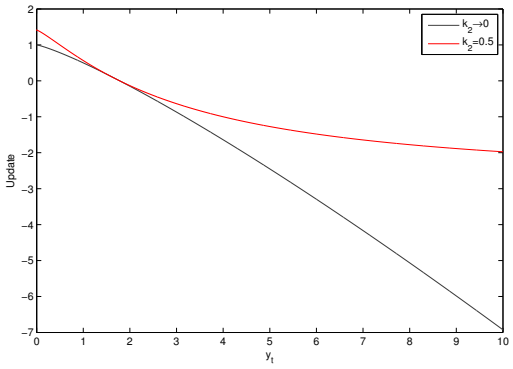
Figure 1 illustrates the probability density function and the GAS updates for the Weibull ($k_2=0$) and the Weibull-gamma mixture model ($k_2=0.5$) for $k_1=1.2$ and $\mu_t=0.5$. For $k_2=0$, the scaled score for the update step simply collapses to $s_t=1-\mu_t y_t^k$. Panel (a) in Figure 1 shows that the mixture density function significantly stretches the right tail of the distribution. Hence, large values of y_t typically signal a large realisation of ν_t, containing little information about μ_t. Accordingly, panel (b) shows that realisations 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 to the corresponding ACP model, in which the update equation for the conditional mean is linear in y_t irrespective of the value of the mixture variance k_2.

3.2 Double gamma mixture models

We can follow a similar approach to obtain a gamma-gamma mixture model. Let y_t be identically and independently $\Gamma(k_1,\gamma_t^{-1})$ distributed random variables with shape



(a) Weibull-gamma mixture PDF



(b) GAS updates

Figure 1: The Weibull-gamma mixture GAS model with $k = 1.2$ and $\mu_t = 0.5$.

coefficient k_1, time-varying scale variable γ_t^{-1} and density function

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

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

The mixture density is

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

We have

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

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

$$\nabla_t = \frac{k_1}{\mu_t} - (1+k_1k_2)\frac{y_t}{1+k_2\mu_t y_t}, \qquad \mathcal{I}_t = \frac{k_1}{\mu_t^2(1+k_2(k_1+1))}. \tag{19}$$

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

4 Maximum Likelihood Estimation

We estimate the parameter vectors in the three classes of models (state space, GAS and ACP) by the method of maximum likelihood. We maximise the log-likelihood function numerically with respect to the parameters using numerical gradient-based optimisation 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.

4.1 Observation-driven models: maximum likelihood

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

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

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

4.2 Parameter-driven models: simulated maximum likelihood

The numerically accelerated importance sampling (NAIS) method of Koopman, Lucas, and Scharth (2011) 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 only requires the specification of (1) and a linear state equation such as (2). The computation times for parameter estimation range from a few seconds to slightly less than a minute for the sample size of two thousand observations we consider in Section 5.

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

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

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

$$L(y;\psi) = \int \frac{p(\alpha, y;\psi)}{g(\alpha, y;\psi)} g(\alpha, y;\psi)\, \mathrm{d}\alpha = g(y;\psi) \int \omega(\alpha, y;\psi) g(\alpha|y;\psi)\, \mathrm{d}\alpha, \qquad (22)$$

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

$$\omega(\alpha, y;\psi) = p(y, \alpha;\psi) \,/\, g(y, \alpha;\psi) = p(y|\alpha;\psi) \,/\, g(y|\alpha;\psi). \qquad (23)$$

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

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

where ω_s is the realised importance weight function in (23) for $\alpha = \alpha^{(s)}$. We base our

estimations of Section 5 on $S = 100$ simulated trajectories.

The choice of the importance sampling density partly determines the accuracy of the likelihood estimate (24). We follow the approach of Richard and Zhang (2007) and choose an importance sampling density that (approximately) minimises the variance of (24). Koopman, Lucas, and Scharth (2011) 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 Appendix A. To further improve the numerical efficiency of the likelihood estimate (24), we also use the control variables proposed by Koopman, Lucas, and Scharth (2011). The NAIS method also facilitates the computation of the smoothed estimates of the state vector α_t; see Appendix B.

5 Predictive analysis: a Monte Carlo study

We conduct a large scale Monte Carlo study to investigate the predictive performances of state space, generalised autoregressive score (GAS) and autoregressive conditional parameter (ACP) models. We simulate series of observations $y_1, \ldots, y_n$ from both parameter-driven and observation-driven data generation processes (DGPs), estimate the parameters for the parameter-driven and observation-driven models, and analyse the forecasts generated by these different specifications. In this setting, we can observe the true values of the time-varying parameter that would be otherwise unknown in empirical studies. We are therefore able to measure which models perform best across a range of empirically relevant DGPs.

5.1 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 of Table 1. The autoregressive state equation (2) completes the specifications of the parameter-driven models. We draw $1,000$ realisations of time series with length $n = 4,000$ for each DGP, where the parameters for the different DPGs are in Table 3. The parameter values reflect typical of what is found in empirical work for these and related models.

In each simulation, we use the first $2,000$ observations to estimate the parameters

for the following model specifications: (i) the correctly specified state space model; (ii) the GAS model based on the same conditional observation density as the DGP, with the appropriate parameter transformation and scaling $S_t = \mathcal{I}_t^{-1/2}$. The functional forms of the parameter updates for the different models are in Table 2; (iii) the corresponding ACP model specification; (iv) in the case of the exponential, gamma, Weibull, and Gaussian models, a robust variant of the GAS and ACP specification. We base the robust specifications on the exponential-gamma, Weibull-gamma, double gamma (see Section 3) and Student's t distributions respectively.

We compute one-step-ahead predictions for the next 2,000 values of θ_t given the estimated parameters, therefore considering two million forecasts in total for each specification. For reference, we also compute the predictions for the true model specification. We compute the forecasts from the state space model using the NAIS method; see Appendix B. For the gamma and Weibull models, we predict their means $\theta_t k$ and $\theta_t \Gamma(1 + 1/k)$, respectively, rather than θ_t.

We measure the accuracy by means of the mean-squared error (MSE), in levels and relative to the MSE of the state space model with estimated parameters. We compute the MSE across the two million forecasts of θ_t. We can calculate the MSEs for θ_t since we know the simulated "true" values of the parameter in the Monte Carlo study.

For the second experiment we adopt the GAS model as the DGP. We consider the nine observation densities in Table 1 and the updating equation (3). Table 2 provides the scaled scores s_t. We choose the GAS specification with the appropriate parameter transformation and scaling by the square root of the information matrix. The remaining 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.

5.2 Results

Table 4 presents the results for the state space DGPs. We focus on two key findings. First, the differences in forecasting accuracy for the estimated state space models and for the corresponding estimated GAS models are small. We find that the MSE increase for the GAS specifications is less than 1% percent for seven of the models and less than 2.5% for the exponential and Gaussian copula specifications. The intuition for this result follows from our discussions in Sections 2.3 and 3. GAS models construct an

Table 3: State Space and GAS DGPs.

We specify the state space models as $y_t|\theta_t \sim p(y_t|\Lambda(\alpha_t);\psi)$, $t = 1,\ldots,n$, $\alpha_{t+1} = \delta + \phi\alpha_t + \eta_t$, $\eta_t \sim N(0,\sigma_\eta^2)$, $\alpha_1 \sim N(\delta/(1-\phi), \sigma_\eta^2/(1-\phi^2))$. We parameterise the generalised autoregressive score models as $y_t|\theta_t \sim p(y_t|\Lambda(\alpha_t);\psi)$, $t = 1,\ldots,n$, $\alpha_{t+1} = d + a\, s_t + b\,\alpha_t$, 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 parameterisations.

Model Type	Distribution	State Space, GAS			
		δ, d	ϕ, b	σ_η, a	other
Count	Poisson	0.00	0.98	0.15	
Count	Neg. 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	$\nu = 10$

updating equation 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 well approximate the conditional densities implied by the state space models.

The second finding is that the GAS specifications lead to large gains in forecasting performance over ACP models for the exponential, Student's t volatility, Gaussian copula and Student's t copula models. The GAS models also outperform the ACP specifications for the other models, but by a smaller margin. For the exponential and Student's t volatility models, the result is due to the fact that the ACP updates are sensitive to realisations 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.

We can also provide further insight for the copula models. We learn from Table 2

that the GAS copula update is given by

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

see Creal, Koopman, and Lucas (2012). The ACP update is $\Phi^{-1}(y_{1t})\Phi^{-1}(y_{2t})$, which is as an unbiased estimator of ρ_t; see Patton (2006). Whereas the ACP driver is sensitive to large realisations of $\Phi^{-1}(y_{1t})$ or $\Phi^{-1}(y_{2t})$, the distinguishing feature of the GAS copula update is the presence of the adjustment term $-\rho_t(\Phi^{-1}(y_{1t})^2+\Phi^{-1}(y_{1t})^2-2)$. Creal, Koopman, and Lucas (2012) 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 $\Phi^{-1}(y_{2t}) = 4$. While the ACP update is the same for the two scenarios, the GAS update is able to separate the two possibilities via the presence of the $\Phi^{-1}(y_{1t})^2$ and $\Phi^{-1}(y_{2t})^2$ terms in s_t.

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 only due to estimation error. Hence, the correct specification strongly outperforms the other two models in relative terms. We focus on the actual differences in mean-square errors. 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 results in Tables 4 and 5 are therefore comparable.

By inspecting the results of Tables 4 and 5, we observe that the state space models seem to be more sensitive to misspecification under the GAS DGPs than the GAS models under the state space DGPs. However, the results vary substantially for different DGPs. While the state space models perform poorly for the gamma duration and Gaussian volatility models, the differences are small and sometimes favour the state space models for the remaining densities. The results for the negative binomial and Student's t models further support our discussion in Section 2.3: the state space models generate better predictions if the GAS observation density is fat-tailed such as those for the mixture models. Table 5 also shows that the forecasting performances of the ACP models are comparable with those of the parameter-driven models. This result further stresses the distinction between GAS and ACP models.

Table 4: Results for the State Space DGPs.

We draw $1,000$ realisations of time series length $n = 4,000$ for the state space DGPs of Table 3. We use the first $2,000$ observations to estimate the correct specification, a robust GAS model (column 1, only for some DGPs), a GAS model based on the same observation density as the DGP (column 2), a robust autoregressive conditional parameter (ACP) specification (column 1, only for some DGPs), and an ACP model based on the same observation density as the DGP (column 2). The GAS and ACP updates are in Table 2. The robust GAS models are the mixture models of Section 3 for the exponential, gamma and Weibull densities and the Student's t GAS model for the volatility model. We compute one-step ahead out-of-sample predictions for the next two thousand values of the time-varying parameter θ_t (or $\theta_t k$ and $\theta_t \Gamma(1+1/k)$ for the gamma and Weibull models respectively, as we are interested in the mean of these distributions) using the true specification and the estimated models.

Model Type	Distribution	State Space		GAS		ACP	
		True	Estimated	(1)	(2)	(1)	(2)
		Relative mean-square error					
Count	Poisson	0.987	1.000	—	1.005	—	1.059
Count	Neg. Binomial	0.982	1.000	—	1.008	—	1.030
Intensity	Exponential	0.979	1.000	1.022	1.200	1.117	1.260
Duration	Gamma	0.985	1.000	1.004	1.050	1.033	1.032
Duration	Weibull	0.981	1.000	1.005	1.057	1.040	1.023
Volatility	Gaussian	0.973	1.000	1.009	1.203	1.041	1.038
Volatility	Student's t	0.968	1.000	—	1.004	—	1.145
Copula	Gaussian	0.957	1.000	—	1.014	—	1.312
Copula	Student's t	0.946	1.000	—	1.006	—	1.430
		Mean-square error					
Count	Poisson	0.280	0.283	—	0.285	—	0.300
Count	Neg. Binomial	0.336	0.342	—	0.345	—	0.352
Intensity	Exponential	0.433	0.442	0.452	0.531	0.494	0.557
Duration	Gamma	0.771	0.783	0.786	0.822	0.809	0.808
Duration	Weibull	0.317	0.324	0.325	0.342	0.337	0.331
Volatility	Gaussian	0.542	0.558	0.563	0.671	0.580	0.579
Volatility	Student's t	0.570	0.589	—	0.591	—	0.674
Copula	Gaussian	0.018	0.019	—	0.021	—	0.027
Copula	Student's t	0.021	0.023	—	0.023	—	0.032

Table 5: Results for the GAS DGPs.

We draw 1,000 realisations of time series length $n = 4,000$ for the GAS DGPs from Table 3. We use the first 2,000 observations to estimate three statistical models: the correct specification, the state space specification with the same observation density as the DGP, and the autoregressive conditional parameter (ACP) specification. The ACP updates are in Table 2. We compute one-step-ahead out-of-sample predictions for the next two thousand values of θ_t, or $\theta_t k_1$ and $\theta_t \Gamma(1 + k_1^{-1})$ for the gamma and Weibull models as we are interested in the mean of these distributions.

Model type	Distribution	Relative mean-square error			Mean-square error		
		State Space	GAS	ACP	State Space	GAS	ACP
Count	Poisson	2.888	1.000	9.187	0.012	0.004	0.038
Count	Neg. 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 t	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 t	1.175	1.000	5.490	0.002	0.002	0.010

5.3 Analysis based on model confidence sets

To verify whether we are able to statistically distinguish parameter-driven models from observation-driven models, we continue with an analysis based on model confidence sets (MCS) as recently proposed by Hansen, Lunde, and Nason (2011). The design of a MCS is such that it contains the best model in terms of a chosen loss function with a certain level of confidence.

We have constructed the model confidence sets as follows. We consider forecast samples of length 100, 250, 500, 1,000 and 2,000. For each sample, we construct a 90% model confidence set based on the MSE criterion. We compute the MSE loss function based on the true parameters as an infeasible benchmark. We then evaluate the loss function using the generated time series as in any empirical application. For the count, intensity and duration models, we simply use the observations y_t. For the volatility and copula models, we assume the presence of a realised measure. The realised measure is $\Lambda_t(\alpha_t + \epsilon_t) - \xi_t$, where $\epsilon_t \sim \mathrm{N}(0, \sigma_\epsilon^2)$ and ξ_t is a bias correction. We choose σ_ϵ^2 to be approximately half of the prediction variance for α_t under the state space model

DGP. We report the proportion of samples in which the state space, GAS and ACP models appear in the model confidence set. We base the comparison on the robust GAS models when applicable.

Tables 6, 7, 8, and 9 display the results. For the state space DGPs in Tables 6 and 8, We find that it is hard to identify the correct specification in practice, even when we assume that the true parameters are known. For all models, the number of times that GAS models are present in the MCS is almost as high as the number of times that the correctly specified state space models are present. The ability of MCS to distinguish between the two alternatives is better for the two copula models. However, even in these cases the GAS models are in the MCS at least 60% of time for sample sizes of 2,000 observations. Table 6 also shows that even though the ACP specifications perform less well, the MCS also has difficulty in excluding the ACP models for the count, intensity, duration and volatility DGPs. In the cases of the copula densities, the ACP forecasts frequently drop out from the MCS with a sufficient number of observations.

When the GAS model is the DGP, the distinction between the feasible and infeasible model confidence sets becomes important. In Table 7 the MCS is based on the true parameters and we observe that MCS is able to single out the correct specification when the sample size increases. On the other hand, the feasible MCS in Table 9 has satisfactory power only for the copula models. We conclude that the observed differences in MSE for forecasting parameters and observations will typically be statistically insignificant. Nevertheless, we do find that the GAS forecasts are in all situations satisfactory and robust to model misspecification.

5.4 Multi-step forecasting

We extend our analysis by considering whether our findings of Section 5.2 also hold for multi-step forecasting. We directly compare the estimates for the unconditional mean of θ_t with the true values implied by Table 3. The unconditional mean is 1.329 for the count, intensity, duration and volatility models and 0.441 for the copula models, while the unconditional variances are 1.35 and 0.038, respectively. We base the analysis on the estimation results in Section 5.2, so that our sample of unconditional predictions contains 1,000 observations for each model. We calculate the unconditional mean estimates for the copula and GAS models by simulation. GAS and ACP mod-

Table 6: Model Confidence Set Results for the State Space DGPs (evaluation based on the true parameters).

The table shows the proportions of replications in which a 90% model confidence set includes the state space, GAS and ACP specifications. We base the comparison on the robust GAS models when applicable. We report the results for different out of sample time lengths (n).

	Count (Poisson)			Count (Negative Binomial)			Intensity (Exponential)		
	State Space	GAS	ACP	State Space	GAS	ACP	State Space	GAS	ACP
$n = 100$	0.94	0.94	0.87	0.89	0.90	0.87	0.81	0.79	0.87
$n = 250$	0.95	0.95	0.90	0.92	0.91	0.89	0.86	0.84	0.84
$n = 500$	0.96	0.96	0.91	0.93	0.92	0.90	0.90	0.86	0.78
$n = 1000$	0.97	0.96	0.91	0.94	0.93	0.89	0.91	0.85	0.68
$n = 2000$	0.98	0.97	0.85	0.96	0.93	0.85	0.95	0.84	0.56
	Duration (Gamma)			Duration (Weibull)			Volatility (Gaussian)		
	State Space	GAS	ACP	State Space	GAS	ACP	State Space	GAS	ACP
$n = 100$	0.88	0.83	0.88	0.89	0.81	0.88	0.85	0.78	0.84
$n = 250$	0.93	0.86	0.88	0.93	0.84	0.88	0.90	0.81	0.84
$n = 500$	0.95	0.87	0.85	0.95	0.84	0.86	0.94	0.80	0.80
$n = 1000$	0.97	0.88	0.81	0.96	0.84	0.83	0.95	0.80	0.76
$n = 2000$	0.97	0.86	0.74	0.97	0.83	0.79	0.97	0.78	0.70
	Volatility (Student's t)			Copula (Gaussian)			Copula (Student's t)		
	State Space	GAS	ACP	State Space	GAS	ACP	State Space	GAS	ACP
$n = 100$	0.85	0.83	0.81	0.84	0.81	0.64	0.85	0.78	0.61
$n = 250$	0.91	0.87	0.79	0.89	0.84	0.54	0.89	0.80	0.50
$n = 500$	0.94	0.89	0.74	0.91	0.82	0.41	0.92	0.77	0.36
$n = 1000$	0.96	0.87	0.63	0.91	0.77	0.25	0.92	0.71	0.20
$n = 2000$	0.96	0.85	0.53	0.92	0.68	0.10	0.92	0.61	0.07

Table 7: Model Confidence Set Results for the GAS DGPs (evaluation based on the true parameters).

The table shows the proportions of replications in which a 90% model confidence contains the state space, GAS and ACP specifications. We report the results for different out of sample time lengths (n).

	Count (Poisson)			Count (Negative Binomial)			Intensity (Exponential)		
	State Space	GAS	ACP	State Space	GAS	ACP	State Space	GAS	ACP
$n = 100$	0.40	0.96	0.28	0.66	0.91	0.34	0.13	0.94	0.40
$n = 250$	0.26	0.96	0.12	0.55	0.92	0.20	0.14	0.94	0.31
$n = 500$	0.18	0.95	0.04	0.48	0.92	0.11	0.16	0.94	0.24
$n = 1000$	0.14	0.95	0.01	0.41	0.90	0.06	0.17	0.93	0.19
$n = 2000$	0.12	0.95	0.00	0.36	0.90	0.04	0.15	0.92	0.12
	Duration (Gamma)			Duration (Weibull)			Volatility (Gaussian)		
	State Space	GAS	ACP	State Space	GAS	ACP	State Space	GAS	ACP
$n = 100$	0.39	0.96	0.56	0.47	0.99	0.48	0.43	0.95	0.65
$n = 250$	0.33	0.97	0.49	0.40	0.99	0.43	0.49	0.97	0.65
$n = 500$	0.27	0.96	0.43	0.34	0.99	0.38	0.50	0.98	0.61
$n = 1000$	0.23	0.95	0.40	0.27	0.99	0.33	0.47	0.99	0.55
$n = 2000$	0.19	0.94	0.33	0.22	0.98	0.27	0.41	0.99	0.46
	Volatility (Student's t)			Copula (Gaussian)			Copula (Student's t)		
	State Space	GAS	ACP	State Space	GAS	ACP	State Space	GAS	ACP
$n = 100$	0.68	0.92	0.61	0.56	0.93	0.23	0.67	0.89	0.14
$n = 250$	0.70	0.94	0.65	0.44	0.94	0.11	0.61	0.91	0.05
$n = 500$	0.67	0.94	0.66	0.34	0.94	0.05	0.53	0.90	0.02
$n = 1000$	0.60	0.92	0.60	0.26	0.94	0.03	0.45	0.89	0.01
$n = 2000$	0.49	0.89	0.49	0.21	0.93	0.01	0.38	0.88	0.00

Table 8: Model Confidence Set Results for the State Space DGPs (evaluation based on the observations).

The table shows the proportions of replications in which a 90% model confidence contains the state space, GAS and ACP specifications. We base the comparison on the robust GAS models when applicable. We report the results for different out of sample time lengths (n).

	Count (Poisson)			Count (Negative Binomial)			Intensity (Exponential)		
	State Space	GAS	ACP	State Space	GAS	ACP	State Space	GAS	ACP
$n = 100$	0.94	0.94	0.87	0.92	0.95	0.91	0.93	0.82	0.93
$n = 250$	0.95	0.95	0.90	0.93	0.95	0.94	0.97	0.86	0.91
$n = 500$	0.96	0.96	0.91	0.94	0.95	0.95	0.98	0.87	0.86
$n = 1000$	0.97	0.96	0.91	0.94	0.95	0.95	0.99	0.88	0.81
$n = 2000$	0.98	0.97	0.85	0.96	0.94	0.93	1.00	0.86	0.74
	Duration (Gamma)			Duration (Weibull)			Volatility (Gaussian)		
	State Space	GAS	ACP	State Space	GAS	ACP	State Space	GAS	ACP
$n = 100$	0.93	0.84	0.93	0.94	0.83	0.93	0.86	0.78	0.84
$n = 250$	0.97	0.88	0.91	0.97	0.86	0.91	0.91	0.82	0.84
$n = 500$	0.98	0.89	0.88	0.99	0.88	0.89	0.94	0.82	0.81
$n = 1000$	0.99	0.90	0.83	0.99	0.88	0.84	0.96	0.81	0.77
$n = 2000$	1.00	0.91	0.76	1.00	0.87	0.79	0.97	0.78	0.71
	Volatility (Student's t)			Copula (Gaussian)			Copula (Student's t)		
	State Space	GAS	ACP	State Space	GAS	ACP	State Space	GAS	ACP
	State Space	GAS	GAM	State Space	GAS	GAM	State Space	GAS	GAM
$n = 100$	0.85	0.83	0.82	0.85	0.82	0.65	0.85	0.78	0.62
$n = 250$	0.91	0.88	0.79	0.89	0.84	0.55	0.90	0.80	0.51
$n = 500$	0.94	0.89	0.74	0.92	0.82	0.41	0.92	0.78	0.36
$n = 1000$	0.96	0.88	0.64	0.92	0.77	0.25	0.93	0.71	0.20
$n = 2000$	0.97	0.85	0.54	0.91	0.68	0.10	0.91	0.60	0.07

Table 9: Model Confidence Set Results for the GAS DGPs (evaluation based on the observations).
The table shows the proportions of replications in which a 90% model confidence contains the state space, GAS and ACP specifications. We report the results for different out of sample time lengths (n).

	Count (Poisson)			Count (Negative Binomial)			Intensity (Exponential)		
	State Space	GAS	ACP	State Space	GAS	ACP	State Space	GAS	ACP
$n = 100$	0.97	0.92	0.86	0.93	0.94	0.90	0.96	0.96	0.93
$n = 250$	0.97	0.95	0.88	0.94	0.95	0.93	0.97	0.98	0.95
$n = 500$	0.95	0.97	0.87	0.94	0.95	0.95	0.98	0.99	0.95
$n = 1000$	0.93	0.98	0.84	0.93	0.95	0.95	0.99	0.99	0.94
$n = 2000$	0.86	1.00	0.73	0.95	0.96	0.92	0.98	0.99	0.94
	Duration (Gamma)			Duration (Weibull)			Volatility (Gaussian)		
	State Space	GAS	ACP	State Space	GAS	ACP	State Space	GAS	ACP
$n = 100$	0.95	0.96	0.93	0.96	0.98	0.92	0.72	0.97	0.90
$n = 250$	0.98	0.97	0.94	0.98	0.99	0.93	0.73	0.99	0.90
$n = 500$	0.99	0.98	0.93	0.99	0.99	0.91	0.69	0.99	0.86
$n = 1000$	0.99	0.98	0.91	0.99	0.99	0.88	0.63	0.99	0.76
$n = 2000$	0.99	0.99	0.91	0.99	0.99	0.86	0.53	0.99	0.63
	Volatility (Student's t)			Copula (Gaussian)			Copula (Student's t)		
	State Space	GAS	ACP	State Space	GAS	ACP	State Space	GAS	ACP
$n = 100$	0.89	0.96	0.79	0.77	0.95	0.42	0.82	0.92	0.31
$n = 250$	0.90	0.97	0.81	0.64	0.96	0.23	0.76	0.93	0.12
$n = 500$	0.87	0.96	0.80	0.50	0.96	0.10	0.69	0.91	0.04
$n = 1000$	0.81	0.95	0.77	0.37	0.95	0.04	0.58	0.91	0.01
$n = 2000$	0.73	0.94	0.68	0.28	0.94	0.02	0.48	0.89	0.00

els sometimes lead to outliers because of their misspecification. We therefore exclude cases in which the estimated persistence for either of these models is higher than 0.997. We observe this problem in up to 0.8% and 5% of estimations for the GAS and ACP models, respectively.

Table 10 reports the mean squared errors for the unconditional mean estimates and the unconditional forecasts. We define the latter as the sum of the MSE of the unconditional mean estimates and the unconditional variance of θ_t. We report the unconditional forecasting MSE as a ratio, with the correct specification as the benchmark. The results show that the GAS models are able estimate the unconditional means of the time-varying parameters almost as accurately as the correct specifications. The ACP models, by contrast, seem to be less precise. This difference in performance is primarily due to fact that the ACP models do not allow for a parameter transformation, making the estimates more sensitive to extreme realisations.

6 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, generalised autoregressive score (GAS) models as flexible representatives of observation-driven models, and autoregressive conditional parameter (ACP) models such as the well-known GARCH and 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 (DGP) 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 decrease compared to those of GAS models. We extend our analysis by considering model confidence sets. Even when considering large samples, the ability of model confidence sets to single out the correct specifications is low. For example, when the state space model is the DGP, we observe that the GAS specification is part of the 90% model confidence set in at least 60% of samples of size 2,000.

Table 10: Unconditional Mean results for the State Space DGPs.

We draw $1,000$ realisations of time series length $n = 2,000$ for the state space DGPs of Table 3. We estimate the correct specification, a robust GAS model (column 1, only for some DGPs), a GAS model based on the same observation density as the DGP (column 2), a robust autoregressive conditional parameter (ACP) specification (column 1, only for some DGPs), and an ACP model based on the same observation density as the DGP (column 2). The GAS and ACP updates are in Table 2. The robust GAS models are the mixture models of Section 3 for the exponential, gamma and Weibull densities and the Student's t GAS model for the volatility model. We use estimated parameters to compute the unconditional mean estimates for θ_t (or $\theta_t k$ and $\theta_t \Gamma(1+1/k)$ for the gamma and Weibull models respectively, as we are interested in the mean of these distributions). The table reports the mean squared errors for the unconditional mean estimates and the unconditional forecasts. We define the latter as the sum of the MSE of the unconditional mean estimates and the unconditional variance of θ_t. We report the unconditional forecasting MSE as a ratio.

Model Type	Distribution		GAS		ACP	
		State Space	(1)	(2)	(1)	(2)
		Relative unconditional mean-square error				
Count	Poisson	1.000	—	0.997	—	1.002
Count	Neg. Binomial	1.000	—	1.004	—	1.044
Intensity	Exponential	1.000	1.004	1.023	1.129	1.210
Duration	Gamma	1.000	1.018	1.009	1.403	1.512
Duration	Weibull	1.000	1.054	1.004	1.567	1.185
Volatility	Gaussian	1.000	1.004	1.000	1.245	1.194
Volatility	Student's t	1.000	—	1.002	—	1.369
Copula	Gaussian	1.000	—	1.001	—	1.010
Copula	Student's t	1.000	—	1.001	—	1.189
		Mean-square error for the unconditional mean				
Count	Poisson	0.058	—	0.055	—	0.061
Count	Neg. Binomial	0.056	—	0.061	—	0.118
Intensity	Exponential	0.055	0.062	0.088	0.236	0.350
Duration	Gamma	0.127	0.154	0.140	0.722	0.883
Duration	Weibull	0.048	0.123	0.053	0.841	0.307
Volatility	Gaussian	0.059	0.065	0.059	0.405	0.333
Volatility	Student's t	0.058	—	0.061	—	0.578
Copula	Gaussian	0.002	—	0.002	—	0.002
Copula	Student's t	0.002	—	0.002	—	0.010

We conclude that GAS models provide a competitive alternative to state space models from a forecasting perspective. Even though the GAS models perform slightly worse if the true DGP is based on a state space model, they seem to be more robust to model misspecification. The practical advantage of this finding stems from the fact that the likelihood function for the GAS model is available in closed-form, such that the analysis of GAS models does not require the use of simulation methods.

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

References

Bauwens, L. and N. Hautsch (2006). Stochastic conditional intensity processes. *Journal of Financial Econometrics 4*(3), 450–493.

Bauwens, L. and D. Veredas (2004). The stochastic conditional duration model: A latent factor model for the analysis of financial durations. *Journal of Econometrics 119*(2), 381–412.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics 21*, 307–328.

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

Carnero, M. A., D. Peña, and E. Ruiz (2004). Persistence and kurtosis in GARCH and stochastic volatility models. *Journal of Financial Econometrics 2*, 319–342.

Cox, D. R. (1981). Statistical Analysis of Time Series: Some Recent Developments. *Scandinavian Journal of Statistics 8*(2), 93–115.

Creal, D., S. J. Koopman, and A. Lucas (2011). A dynamic multivariate heavy-tailed model for time-varying volatilities and correlations. *Journal of Business and Economic Statistics 29*(4), 552–563.

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

Creal, D., B. Schwaab, S. J. Koopman, and A. Lucas (2011). Observation Driven Mixed-Measurement Dynamic Factor Models with an Application to Credit Risk. Discussion paper 11-042/2/DSF16, Tinbergen Institute and Duisenberg school of finance.

Das, S. and K. Srinivasan (1997, June). Duration of firms in an infant industry: the case of Indian computer hardware. *Journal of Development Economics 53*(1), 157–167.

de Jong, P. and N. Shephard (1995). The simulation smoother for time series models. *Biometrika 82*, 339–350.

Durbin, J. and S. J. Koopman (1997). Monte carlo maximum likelihood estimation for non-gaussian state space models. *Biometrika* (84), 669–684.

Durbin, J. and S. J. Koopman (2000). 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), 3–56.

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

Durbin, J. and S. J. Koopman (2002). A simple and efficient simulation smoother for state space time series analysis. *Biometrika* (89), 603–616.

Engle, R. (2002). Dynamic Conditional Correlation. *Journal of Business and Economic Statistics 20*(3), 339–350.

Engle, R. and G. Gallo (2006). A multiple indicators model for volatility using intra-daily data. *Journal of Econometrics 131*(1-2), 3–27.

Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. *Econometrica 50*, 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.

Ghysels, E., A. Harvey, and E. Renault (1996). Stochastic volatility. In G. Maddala and C. Rao (Eds.), *Handbook of Statistics, Vol 14*. Elsevier, Amsterdam.

Grammig, J. and K.-O. Maurer (2000, June). Non-monotonic hazard functions and the autoregressive conditional duration model. *The Econometrics Journal 3*(1), 16–38.

Hafner, C. and H. Manner (2011). Dynamic stochastic copula models: Estimation, inference and applications. *Journal of Applied Econometrics*. forthcoming.

Hansen, P. and A. Lunde (2005). A forecast comparison of volatility models: Does anything beat a GARCH(1,1) model? *Journal of Applied Econometrics 20*, 873–889.

Hansen, P. R., A. Lunde, and J. M. Nason (2011). The Model Confidence Set. *Econometrica 79*(2), 453–497.

Jungbacker, B. and S. J. Koopman (2007). Monte carlo estimation for nonlinear non-gaussian state space models. *Biometrika* (94), 827–839.

Koopman, S. J., A. Lucas, and M. Scharth (2011). Numerically accelerated importance sampling for nonlinear non-Gaussian state space models. Working paper 2011-057/4, Tinbergen Institute. update 2012.

Lancaster, T. (1979). Econometric Methods for the Duration of Unemployment. *Econometrica 47*(4), 939–956.

Liesenfeld, R. and J. Richard (2003). Univariate and multivariate stochastic volatility models: Estimation and diagnostics. *Journal of Empirical Finance 10*(4), 505–531.

Melino, A. and S. Turnbull (1990). Pricing foreign currency options with stochastic volatility. *Journal of Econometrics 45*, 239–265.

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

Richard, J. and W. Zhang (2007). Efficient high-dimensional importance sampling. *Journal of Econometrics 141*, 1385–1411.

Rydberg, T. H. and N. Shephard (2000). A modelling framework for the prices and times of trades made on the NYSE. In W. J. Fitzgerald, R. L. Smith, A. T. Walden, and P. C. Young (Eds.), *Nonlinear and nonstationary signal processing*, pp. 217– 246. Cambridge University Press.

Shephard, N. and M. Pitt (1997). Likelihood analysis of non-gaussian measurement time series. *Biometrika 84*, 653–667.

Shephard, N. and K. Sheppard (2010, March). Realising the future: forecasting with high-frequency-based volatility (HEAVY) models. *Journal of Applied Econometrics 25*(2), 197–231.

Tauchen, G. and M. Pitts (1983). The price variability-volume relationship in speculative markets. *Econometrica 51*, 485–505.

Taylor, S. J. (1986). *Modelling Financial Time Series*. Chichester, UK: John Wiley.

A Numerically Accelerated Importance Sampling

We represent the Gaussian importance density as

$$g(\alpha, y; \psi) = \prod_{t=1}^{n} g(y_t|\alpha_t; \psi) g(\alpha_t|\alpha_{t-1}; \psi), \tag{A.1}$$

where $g(\alpha_t|\alpha_{t-1}; \psi)$ is the Gaussian density for α_t as implied by (1) and

$$g(y_t|\alpha_t; \psi) = \exp\left\{a_t + b_t' \alpha_t - \frac{1}{2}\alpha_t' C_t \alpha_t\right\}, \tag{A.2}$$

with a_t, b_t and C_t defined as functions of the data vector y and the parameter vector ψ, for $t = 1, \ldots, n$. The constants $a_1, \ldots, a_n$ ensure that $g(\alpha, y; \psi)$ integrates to one. The set of importance sampling parameters is

$$\chi = \{b_1, \ldots, b_n, C_1, \ldots, C_n\}. \tag{A.3}$$

Following Shephard and Pitt (1997), the importance density (A.2) is equivalent to the density function associated with observation $y_t^* = C_t^{-1} b_t$ and the linear Gaussian

observation equation

$$y_t^* = \alpha_t + \varepsilon_t, \qquad \varepsilon_t \sim \mathrm{N}(0, C_t^{-1}), \qquad t = 1, \ldots, n. \tag{A.4}$$

The importance sampling algorithm is then based on standard linear state space methods. de Jong and Shephard (1995) and Durbin and Koopman (2002) have developed simulation smoothing methods for sampling α from $g(\alpha|y^*;\psi)$ in a computationally efficient way. The Kalman filter calculates $g(y^*;\psi)$ via its evaluation of the likelihood function for the linear state space model (A.4).

The choice of importance parameters in χ determines the variance of the likelihood estimate (24). Following Richard and Zhang (2007) and Koopman, Lucas, and Scharth (2011), we obtain an efficient set of importance parameters $\chi_t = \{b_t, C_t\}$ via the (approximate) variance minimisation problem

$$\min_{\chi_t} \int \lambda^2(\alpha_t, y_t; \psi)\omega(\alpha_t, y_t; \psi)g(\alpha_t|y;\psi)\,\mathrm{d}\alpha_t \tag{A.5}$$

where $\omega(\alpha_t, y_t; \psi) = p(y_t|\alpha_t;\psi) \,/\, g(y_t|\alpha_t;\psi)$ and $\lambda(\alpha_t, y_t;\psi) = \log p(y_t|\alpha_t;\psi) - \log g(y_t|\alpha_t;\psi) - \lambda_{0t}$, for $t = 1, \ldots, n$, where λ_{0t} is the normalising constant.

For a given set of values in $\chi = \chi^+ = \left\{b_1^+, \ldots, b_n^+, C_1^+, \ldots, C_n^+\right\}$ of (A.3), we have that the smoothed importance density $g(\alpha_t|y;\psi) = g(\alpha_t|y^*;\psi)$ based on the linear Gaussian model (A.4) is given by

$$g(\alpha_t|y^*;\psi) = \mathrm{N}(\widehat{\alpha}_t\,,\, V_t) = \exp\left\{-\frac{1}{2}V_t^{-1}(\alpha_t - \widehat{\alpha}_t)^2\right\} \,/\, \sqrt{2\pi\, V_t}, \tag{A.6}$$

where we compute $\widehat{\alpha}_t$ and V_t by KFS methods applied to the importance model (A.4) for $y_t^* = (C_t^+)^{-1}b_t^+$.

For $\chi = \chi^+$, we evaluate the integral in (A.5) by means of a Gauss-Hermite quadrature with $M = 30$ abscissae z_j and associated weights $h(z_j)$ with $j = 1, \ldots, M$. The required inputs are available in standard computational packages. We express the minimisation in (A.5) as

$$\min_{\chi_t} \sum_{j=1}^{M} \lambda^2(\widetilde{\alpha}_{tj}, y_t; \psi)w_{tj}, \qquad w_{tj} = g(\widetilde{\alpha}_{tj}|y^*;\psi)\,\omega^*(\widetilde{\alpha}_{tj}, y_t;\psi)\,h(z_j)e^{z_j^2}, \tag{A.7}$$

where $\widetilde{\alpha}_{tj} = \widehat{\alpha}_t + V_t^{1/2} z_j$, for $j = 1, \ldots, M$. It follows from (A.6) that

$$g(\widetilde{\alpha}_{tj}|y^*; \psi) = \exp\left\{-\frac{1}{2} z_j^2\right\} / \sqrt{2\pi}, \qquad t = 1, \ldots, n.$$

The minimisation (A.7) takes place via an iterative method. For a given $\chi = \chi^+$, we obtain $\widehat{\alpha}_t$ and V_t from the KFS applied to (A.4), for $t = 1, \ldots, n$. Minimisation (A.7) for a scalar $\widetilde{\alpha}_{tj}$ reduces to weighted least squares computations, for each t, with dependent variable $p(y_t|\widetilde{\alpha}_{tj}; \psi)$, explanatory variables $\widetilde{\alpha}_{tj}$, $\widetilde{\alpha}_{tj}^2$ (including a constant) and weights w_{tj}. We obtain the minimum in (A.7) by setting $\chi_t = \{b_t, C_t\}$ equal to the least squares estimates associated with explanatory variables $\widetilde{\alpha}_{tj}$ and $\widetilde{\alpha}_{tj}^2$, respectively. The new value for χ_t becomes χ_t^+ in the next iteration. The iterative procedure terminates after convergence.

B Forecasting for the parameter-driven models

We now provide the details on how we calculate the prediction

$$\mathbb{E}(\theta_{t+1}|y; \psi) = \int \Lambda(\alpha_{t+1}) p(\alpha_{t+1}|y_1, \ldots, y_t) d\alpha_{t+1}, \tag{B.1}$$

for the state space model specified by (1) and (2) and where the conditional density $p(\alpha_{t+1}|y_1, \ldots, y_n)$ is not available in closed-form.

We follow a Monte Carlo approach based on the importance sampling techniques we have discussed in Section 4. We rewrite (B.1) as

$$\mathbb{E}(\theta_{t+1}|y; \psi) = \int \mathbb{E}(\Lambda(\alpha_{t+1})|\alpha_t) p(\alpha_t|y_1, \ldots, y_t) d\alpha_t, \tag{B.2}$$

where we can typically calculate $\mathbb{E}(\Lambda(\alpha_{t+1})|\alpha_t)$ analytically. To simplify the notation we define $f(\alpha_t) = \mathbb{E}(\Lambda(\alpha_{t+1})|\alpha_t)$ and $\bar{f} = \mathbb{E}(\theta_{t+1}|y; \psi)$. By focusing on $f(\alpha_t)$, we are able to obtain an efficient Rao-Blackwellised estimate of $\bar{f}$ without the need to simulate α_{t+1} under the importance density.

Let $\alpha' = (\alpha_1', \ldots, \alpha_t')$ and $y' = (y_1', \ldots, y_t')$. Durbin and Koopman (2001), among others, show that by considering an importance density $g(\alpha|y; \psi)$, we can estimate $\bar{f}$

by exploiting the fact that

$$\bar{f} = \int f(\alpha_t) \frac{p(\alpha|y;\psi)}{g(\alpha|y;\psi)} g(\alpha|y;\psi) d\alpha = \frac{\mathbb{E}_g \left[f(\alpha_t) \frac{p(\alpha,y;\psi)}{g(\alpha|y;\Psi)} \right]}{\mathbb{E}_g \left[\frac{p(\alpha,y;\psi)}{g(\alpha|y;\psi)} \right]}. \tag{B.3}$$

We estimate $\bar{f}$ by drawing S trajectories $\alpha^{(1)}, \ldots, \alpha^{(S)}$ from the efficient importance density $g(\alpha|y;\psi)$ of Appendix A and computing

$$\widehat{f} = \sum_{i=1}^{S} f\left(\alpha_t^{(s)}\right) \omega_s \,/\, \sum_{i=1}^{S} \omega_s, \tag{B.4}$$

where he have defined ω_s in (24).

Two strategies allow us to improve the efficiency of $\widehat{f}$. First, we use antithetic variables for variance reduction; see for example Durbin and Koopman (2002). Second, we note that observations far in the past add little or no information about the current state α_t, but contribute to the variance of the importance weights ω_s. We therefore implement the steps above for a shorter sample of recent observations (we use the most recent 250 observations in Section 5).

For the Gaussian and Student's t copula models, no analytical expression for the expectation $\mathbb{E}(\Lambda(\alpha_{t+1})|\alpha_t)$ is available for our choice of transformation $\Lambda(\cdot)$. Hence, we use a second order Taylor approximation of $\Lambda(\alpha_{t+1})$ around $\widehat{\alpha}_{t+1} = \mathbb{E}(\alpha_{t+1}|\alpha_t)$. We have that

$$\begin{aligned}
\mathbb{E}(\Lambda(\alpha_{t+1})|\alpha_t) &\approx \mathbb{E}\left(\Lambda(\widehat{\alpha}_{t+1}) + \Lambda'(\widehat{\alpha}_{t+1})(\alpha_{t+1} - \widehat{\alpha}_{t+1}) + \frac{\Lambda''(\widehat{\alpha}_{t+1})}{2}(\alpha_{t+1} - \widehat{\alpha}_{t+1})^2 \;\middle|\; \alpha_t \right) \\
&= \Lambda(\widehat{\alpha}_{t+1}) + \frac{\Lambda''(\widehat{\alpha}_{t+1})}{2} \mathrm{Var}(\alpha_{n+1} \mid \alpha_t) \\
&= \Lambda(\widehat{\alpha}_{t+1}) + \frac{\Lambda''(\widehat{\alpha}_{t+1})}{2} \sigma^2_{\eta,t}.
\end{aligned} \tag{B.5}$$