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

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

[(a)] VU University Amsterdam, The Netherlands

[(b)] Tinbergen Institute, The Netherlands

[(c)] CREATES, Aarhus University, Denmark

[(d)] Australian School of Business, University of New South Wales

This version: March 2014

*We would like to thank Andrew Harvey for his helpful comments. André Lucas acknowledges the financial support of the Dutch Science Foundation (NWO). Marcel Scharth wishes to acknowledge partial support from the Australian Research Council (ARC) grants DP0988579 and LP0774950. Contact author: S.J. Koopman (s.j.koopman@vu.nl).

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. In most cases, the differences in mean squared errors are smaller than 1% and model confidence sets have low power when comparing the two different model classes. Within the class of observation-driven models, dynamic models relying on the predictive score outperform specifications based on moments. Our main findings are supported by the results from an extensive empirical study in volatility forecasting. We conclude that dynamic models driven by the score function lead to accurate forecasts without the large computational costs associated with parameter-driven models.

1 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 generalised autoregressive score model. Creal, Koopman, and Lucas (2012) 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 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 the latter set of models under the general heading of autoregressive conditional moment (ACM) models.

Cox (1981) classifies time-varying parameter models into 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 their popularity 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 large amount of effort 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 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 a similar degree of generality as obtained for nonlinear non-Gaussian state space models. The GAS model 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, Koopman, and Lucas (2012) 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, Schwaab, Koopman, and Lucas (2011). In the same way as parameter-driven models, GAS models can accommodate different parameterisations 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 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, 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 (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 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 square 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 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 moment (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. 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 analyse 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 and more complex two-factor specifications for the volatility dynamics.

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 introduces several new GAS models for continuous mixtures. Section 4 discusses the estimation of parameters for the different model classes. Section 5 presents the Monte Carlo results. Section 6 presents the empirical application. Section 7 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 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 $p(y_t|\theta_t;\psi)$.

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 $\sigma_\eta^2/(1-\phi^2)$.

The equations (1) and (2) characterise a class of nonlinear non-Gaussian state space models; see Durbin and 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, Lucas, and Scharth (2011) report recent developments on Monte Carlo methods for the analysis of general nonlinear non-Gaussian state space models.

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 consider the autoregressive update

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

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 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 (2011).

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 main 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

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

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

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

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 non-degenerate 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, Koopman, and Lucas (2012). 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 $\mathcal{I}_t$ which does not depend on α_t, such that the choice of scaling matrix S_t becomes irrelevant. In this paper 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. Results for a selection of alternative specifications and parametrisations are available upon request.

2.1.3 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

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

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 realised 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 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 autoregressive conditional moment 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), some autoregressive copulas in Patton (2006), and realised 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.

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 with ν degrees of freedom.

Model type	Distribution	Density	Link function
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	$\frac{\frac{1}{2\pi\sqrt{1-\rho_t^2}}\exp\left[-\frac{z_{1t}^2+z_{2t}^2-2\rho_t z_{1t}z_{2t}}{2(1-\rho_t^2)}\right]}{\prod_{i=1}^2\frac{1}{\sqrt{2\pi}}e^{-z_{it}^2/2}}$	$\rho_t = \frac{1-\exp(-\alpha_t)}{1+\exp(-\alpha_t)}$
Copula	Student's t	$\frac{\Gamma(\frac{\nu+2}{2})\Gamma(\frac{\nu}{2})}{\Gamma(\frac{\nu+1}{2})}\frac{\frac{1}{\sqrt{1-\rho_t^2}}\left[1+\frac{z_{1t}^2+z_{2t}^2-2\rho_t z_{1t}z_{2t}}{\nu(1-\rho_t^2)}\right]^{-\frac{\nu+2}{2}}}{\prod_{i=1}^2(1+z_{it}/\nu)^{-\frac{\nu+1}{2}}}$	$\rho_t = \frac{1-\exp(-\alpha_t)}{1+\exp(-\alpha_t)}$

2.2 Observation densities

In Table 1 we present the observation densities $p(y_t|\theta_t;\psi)$ from (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 (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 (2012).

Table 2 completes the specifications of the considered observation-driven models.

It presents the generalised 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 appear in the original paper by Creal, Koopman, and Lucas (2012). The Student's t volatility model is also discussed in Harvey and Chakravarty (2008). The GAS model for the Student's t copula is obtained in Creal, Koopman, and Lucas (2011).

2.3 Parameter-driven versus observation-driven models

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

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

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 (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 (2) is conditionally leptokurtic. Similarly, the stochastic count and duration models we study below display conditional over-dispersion.

On the 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 Durbin and Koopman (1997). Simulation methods are also necessary for the estimation and forecasting of

Table 2: Observation-driven model updates.

The table displays the score and information matrix for the models given in Table 1. The GAS update $s_t = \nabla_t(\alpha_t)\mathcal{I}_t(\alpha_t)^{-1/2}$ we consider in this paper is invariant under non-degenerate 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 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 $\nabla_t(\theta_t)$	GAS $\mathcal{I}_t(\theta_t)$	ACM 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	$\frac{(1+\rho^2)(\omega_t\hat{z}_{1,t}-\rho_t)-\rho_t(\omega_t\hat{z}_{2,t}-2)}{(1-\rho^2)^2}$ $\omega_t = \frac{\nu+2}{\nu+\frac{\hat{z}_{2,t}-2\rho_t\hat{z}_{1,t}}{1-\rho^2}}$	$\frac{\left(\nu+2+\nu\rho_t^2\right)}{(\nu+4)(1-\rho_t^2)^2}$	$z_{1,t}z_{2,t}$

the time-varying parameter θ_t. We discuss parameter estimation further in Section 4.

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. 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, 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, 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, ACM models rely exclusively on a specific moment of $p(y_t|\theta_t;\psi)$, such as the mean or the 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

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

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 ν 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

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

If $\nu^{-1}\to 0$, the GAS update step recovers the GARCH(1,1) model as it should. However, if ν 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 (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.

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 step 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 or Burr density is given by

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

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. They advocate the Burr model for the empirical analysis of price durations on the basis of its ability to account for non-monotonic hazard functions; see also the discussions in Andres and Harvey (2012).

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 with respect to μ_t are given by

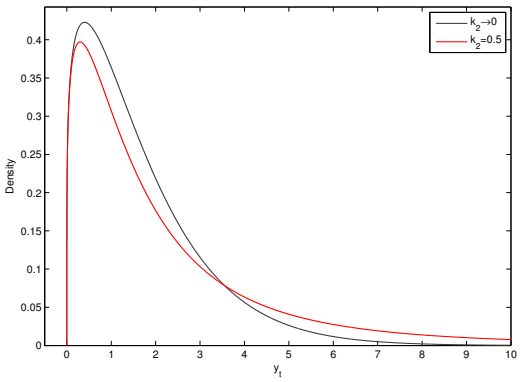
$$\nabla_t = \frac{1}{\mu_t} - (1 + k_2)\frac{y_t^{k_1}}{1 + k_2\mu_t y_t^{k_1}}, \qquad \mathcal{I}_t^{-1} = \mu_t^2(1 + 2k_2), \tag{14}$$

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

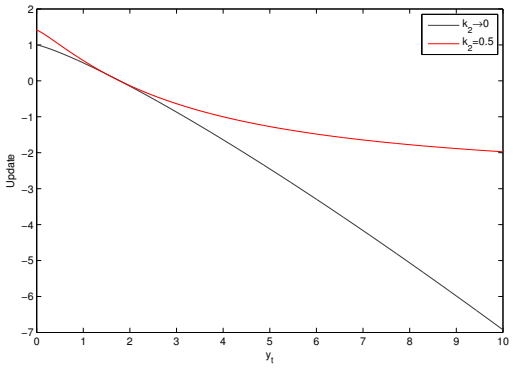
$$s_t = \mathcal{I}_t^{-1/2}\nabla_t = \sqrt{1 + 2k_2}\left(1 - (1 + k_2)\frac{\mu_t y_t^{k_1}}{1 + k_2\,\mu_t\,y_t^{k_1}}\right). \tag{15}$$

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 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.



(a) Weibull-gamma mixture PDF



(b) GAS updates

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

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 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

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

Hence the intuition for the GAS updates of the gamma-gamma mixture distribution is 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 ACM) 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) 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 $\hat{\psi}$.

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

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

where ω_s is the realised importance weight function in (23) for $\alpha = \alpha^{(s)}$. We base our estimations in 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 moment (ACM) 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 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 $\mathbb{E}_j(\theta_{t+1}|y_1, \ldots, y_t; \widehat{\psi}_j)$, 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 realisations of θ_t which would otherwise be unobserved in an empirical study. Hence we compare the predictions with the true θ_t and are able to accurately 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 listed in Table 1. The autoregressive state equation (2) completes the specifications of all parameter-driven models. We draw $1,000$ realisations of time series with length $n = 4,000$ for each DGP, where the parameter values for the different DPGs are in Table 3. The parameter values are chosen to be typical for 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 models are listed in Table 2; (iii) the ACM model for the corresponding specification; (iv) in the case of the exponential, gamma, Weibull, and Gaussian models, a robust variant of the GAS and ACM specification. We base the robust specifications on the exponential-gamma, Weibull-gamma, double gamma (see Section 3) and Student's t distributions, respectively. These alternative specifications allow for fatter tails and more over-dispersion and are therefore more comparable to the parameter driven specifications; see the discussion in Section 3.

We compute one-step ahead predictions for the next 2,000 values of θ_t given the parameter values estimated from the first 2,000 observations y_t. We therefore consider two million $(2,000 \times 1,000)$ forecasts for each specification. For reference purposes, we also compute the predictions for the true model specification but with estimated parameters. In the case of the state space model, we estimate the parameters and predict θ_t one-step ahead using the NAIS method; see Appendix B. For the gamma and Weibull models, we predict the means for $\theta_t k$ and $\theta_t \Gamma(1 + 1/k)$, respectively, and compare it to the true simulated mean. 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 two 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 (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 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	$\nu = 10$

5.2 Results

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 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 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.

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 smaller margin. For the exponential and

Student's t volatility models, the result is due to the fact that the ACM 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

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

see Creal, Koopman, and Lucas (2012). 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, Koopman, and Lucas (2012) argue that the ACM driver can be sensitive to large realisations 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, even though the second scenario intuitively 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 $-\rho_t(\Phi^{-1}(y_{1t})^2 + \Phi^{-1}(y_{1t})^2 - 2)$, which discounts the large realisation of $\Phi^{-1}(y_{2t})$.

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 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 do not report the results for other loss functions such as mean-absolute 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 properties of the time-varying processes underlying the results in Tables 4 and 5 are therefore comparable.

The direct comparison of 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. However, by inspecting the results of Tables 4 and 5, we observe that the state space models seem to be more sensitive to misspecification

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 moment (ACM) specification (column 1, only for some DGPs), and an ACM model based on the same observation density as the DGP (column 2). The GAS and ACM 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		ACM	
		True	Estimated	(1)	(2)	(1)	(2)
		Relative mean-squared 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-squared 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 moment (ACM) specification. The ACM 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-squared error			Mean-squared error		
		State Space	GAS	ACM	State Space	GAS	ACM
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

under the GAS DGPs than the GAS models under the state space DGPs. We further observe that 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 for the mixture models. Table 5 also shows that the forecasting performances of the ACM models are comparable with those of the parameter-driven models. This result further stresses the distinction between GAS and ACM models.

6 Empirical example

To illustrate the relevance of our Monte Carlo findings for empirical applications, we analyse the performance of different parameter-driven and observation-driven models

for predicting the volatility of stocks and indices. 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.

6.1 Descriptions 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) have 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 which are able to account for the slow decays observed in the empirical autocorrelations of squared and absolute returns and realised volatilities. The results for one factor models are available upon request.

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.

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

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

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 $T_\nu(\cdot)$ denotes the CDF of a standardised Student's t random variable with ν degrees of freedom and $\Phi^{-1}(\cdot)$ the inverse standard normal CDF. We assume that

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

This stochastic volatility specification has been considered by Koopman, Lucas, and Scharth (2012); we adopt their importance method for simulated maximum likelihood estimation and signal forecasting.

6.2 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 indices between January 1996 and October 2012. Parameter estimation for all eight models is based on daily close-to-close returns. We compute one-step ahead forecasts starting in 2001 and 2004 for the stocks and indices, respectively. This choice implies that we use approximately 2000 observations for initial estimation and 2900 (2200 for the indices) observations for predictive evaluation. For each model the parameters are re-estimated every three months, in an expanding window including all previous daily returns. The precision of the forecasts from a model is evaluated by comparing the volatility forecasts with the daily realised volatilities as measured from high-frequency data. We adopt the realised kernel of Barndorff-Nielsen, Hansen, Lunde, and Shephard (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 realised 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.

6.3 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 indices, whether or not we consider leverage effects. The relative difference in variance between these models is less than 2% for 19 out of 25 series and never reaches more than 6% for the remaining series. In contrast, the GARCH and GJR-GARCH models appear to be less robust, generating similar forecasts to the previous two models for some stocks but leading to markedly higher forecasting variances for others. The EGARCH model performs better than the other GARCH models, achieving a similar performance to the SV and GAS in a majority of cases and even outperforming these models by a small margin in four series. 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 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 best performing, it appears in the 90% model confidence set and is the only alternative in the model confidence set. The model confidence sets are proposed by Hansen, Lunde, and Nason (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 indices. The SV model achieves the lowest forecasting errors for 48% of the series, closely followed by the GAS model (36%) and to a lesser extent the EGARCH model (16%). The 90% model confidence set includes the SV, GAS and EGARCH models for 88%, 68% 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. The EGARCH model is often competitive but is rejected for half of the series in favour

of the SV and/or GAS models. Finally, the model confidence sets provide substantial evidence that the GJR-GARCH model is outperformed statistically by the the other specifications.

7 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 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 compared to that of GAS models. When considering model confidence sets, GAS and state space models are hard to distinguish statistically. Even for large samples, the power of model confidence sets to single out the correct specifications is low because the models produce forecasts of similar accuracy. 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.

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 for the GAS model stems from the fact that the likelihood function for the GAS model is available in closed-form. The analysis of GAS models 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.

Table 6: Results for the empirical example.

The table shows the relative the performance of different two-factor models with Student's t return innovations. We estimate the models using daily returns and use them to forecast daily volatilities one-step ahead. The comparison criterion is the variance of the residuals of Mincer-Zarnowitz regressions of the realised volatilities on the daily volatility forecasts by each specification. In the bottom lines, we report the proportion of series in which a given specification is the best performing model, appears in a 90% model confidence set and is the only alternative in the model confidence set (MCS).

Stock/index	No leverage				Leverage			
	SV	GAS	GARCH	EGARCH	SV	GAS	GJR	EGARCH
Alcoa	1.11	1.11	1.11	1.10	1.00	0.99	1.02	1.00
American Express	1.08	1.08	1.09	1.08	1.00	0.99	1.02	0.99
Boeing	1.07	1.06	1.13	1.06	1.00	0.99	1.04	1.00
Caterpillar	1.18	1.16	1.26	1.17	1.00	1.00	1.24	1.05
Chevron	1.12	1.13	1.21	1.13	1.00	1.00	1.20	1.00
Walt Disney	1.13	1.19	1.18	1.19	1.00	1.05	1.09	1.10
General Electric	1.06	1.04	1.06	1.05	1.00	0.99	1.01	1.01
IBM	1.12	1.11	1.23	1.13	1.00	0.98	1.11	1.00
Intel	1.02	1.02	1.07	1.04	1.00	0.99	1.15	1.02
Johnson & Johnson	1.05	1.08	1.17	1.06	1.00	1.01	1.20	1.03
JPMorgan	1.07	1.09	1.07	1.08	1.00	1.02	1.09	1.02
Coca-Cola	1.07	1.06	1.13	1.08	1.00	0.99	1.09	1.02
McDonald's	1.06	1.10	1.17	1.12	1.00	1.04	1.14	1.15
Merck	1.05	1.04	1.36	1.31	1.00	1.02	1.53	1.30
Microsoft	1.08	1.07	1.21	1.11	1.00	1.01	1.20	1.06
Pfizer	1.03	1.00	1.12	1.04	1.00	1.01	1.20	1.04
Procter & Gamble	1.06	1.07	1.06	1.06	1.00	0.99	1.04	0.99
AT&T	1.06	1.06	1.11	1.08	1.00	1.03	1.08	1.04
Wal-Mart	1.04	1.04	1.07	1.03	1.00	0.99	1.05	0.98
Exxon	1.10	1.13	1.18	1.12	1.00	1.01	1.19	1.00
DAX 30	1.27	1.26	1.27	1.23	1.00	1.01	1.14	0.99
FTSE 100	1.20	1.16	1.22	1.17	1.00	1.06	1.16	1.08
NASDAQ	1.20	1.20	1.21	1.22	1.00	0.99	1.01	1.00
Nikkei 225	1.07	1.08	1.14	1.10	1.00	0.97	1.10	1.00
S&P 500	1.28	1.30	1.35	1.30	1.00	1.04	1.22	1.05
Best model	0.00	0.00	0.00	0.00	0.48	0.36	0.00	0.16
In 90% MCS	0.08	0.12	0.00	0.04	0.88	0.68	0.12	0.48
Only model in MCS	0.00	0.00	0.00	0.00	0.24	0.08	0.00	0.04

ACM 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.

The findings rely mostly on our large-scale Monte Carlo study. We have verified these findings with an empirical study for 20 individual stocks and 5 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

Andres, P. and A. C. Harvey (2012). The dynamic location/scale model: with applications to intra-day financial data.

Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard (2008). Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise. *Econometrica 76*, 1481–1536.

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

Durbin, J. and S. J. Koopman (2012). *Time Series Analysis by State Space Methods* (2 ed.). Oxford University Press.

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. and G. Lee (1999). A long-run and short-run component model of stock return volatility. In R. Engle and H. White (Eds.), *Cointegration, Causality, and Forecasting: A Festschrift in Honour of Clive W. J. Granger*, pp. 475–497. Oxford University Press.

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.

Glosten, L. R., R. Jagannathan, and D. E. Runkle (1993). On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks. *Journal of Finance 48*(5), 1779–1801.

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 (2012). Dynamic stochastic copula models: Estimation, inference and applications. *Journal of Applied Econometrics* (27), 269–295.

Hansen, P., Z. Huang, and H. H. Shek (2012). Realized GARCH: a joint model for returns and realized measures of volatility. *Journal of Applied Econometrics 27*, 877–906.

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.

Harvey, A. C. and T. Chakravarty (2008). Beta-t-(E)GARCH.

Harvey, A. C. and A. Luati (2014). Filtering with heavy tails. *Journal of the American Statistical Association*, forthcoming.

Janus, P., S. J. Koopman, and A. Lucas (2011). Long Memory Dynamics for Multivariate Dependence Under Heavy Tails. Working paper 11-175/5/DSF28, Tinbergen Institute and Duisenberg school of finance.

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.

Koopman, S. J., A. Lucas, and M. Scharth (2012). Efficient likelihood evaluation for non-Gaussian measurement state space models with multiple time-varying parameters. Technical report, VU University Amsterdam.

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.

Nelson, D. B. (1991). Conditional Heteroskedasticity in Asset Returns: A New Approach. *Econometrica 59*(2), 347–370.

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 computa-

tionally 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 (2012), 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 impor-

tance 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

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