A General Class of Score-Driven Smoothers*

Giuseppe Buccheri[1], Giacomo Bormetti[2], Fulvio Corsi[3,4], and Fabrizio Lillo[2,5]

[1]Scuola Normale Superiore, Italy

[2]University of Bologna, Italy

[3]University of Pisa, Italy

[4]City University of London, UK

[5]CADS, Human Technopole, Milan, Italy

Abstract

We first show that, in the steady state, Kalman filter and smoother recursions can be re-parameterized in terms of the score of the conditional density and the Fisher matrix. Since in the new representation the predictive filter has the form of score-driven models, we introduce, by analogy, a score-driven update filter (SDU) and smoother (SDS). In this new framework, we recover smoothed estimates of observation-driven models, as well as assess filtering uncertainty and construct confidence bands. We test both empirically and through simulations the advantages of SDU and SDS over standard score-driven filters and exact simulation-based methods.

Keywords: Score-driven models, Smoothing, Kalman filter, State-Space models, Filtering uncertainty

JEL codes: C22, C32, C58.

*Corresponding author: giuseppe.buccheri@sns.it, Scuola Normale Superiore, Piazza dei Cavalieri 7, 56126, Pisa (Italy). We are particularly grateful for suggestions we have received from Andrew Harvey, Marcin Zamojski and participants to the 11$^{\text{th}}$ CFE conference in London, the 11$^{\text{th}}$ SoFiE conference in Lugano and the 2018 IAAE conference in Montreal.

1 Introduction

Observation-driven models like the GARCH model of Bollerslev (1986), where time-varying parameters are driven by functions of lagged observations, are typically viewed as data generating processes. As such, all relevant information is encoded on past observations and there is no room for using actual and future observations when estimating time-varying parameters. However, they can also be viewed as predictive filters, as time-varying parameters are one-step-ahead predictable. This idea was largely exploited by Daniel B. Nelson, who explored the asymptotic properties of conditional covariances of a misspecified GARCH under the assumption that the data generating process is a diffusion[1]; see Nelson (1992), Nelson and Foster (1994), Nelson and Foster (1995) and Nelson (1996). In particular, Nelson (1996) showed how to efficiently use information in both lagged and led GARCH residuals to estimate the unobserved states of a stochastic volatility model. Despite many observation-driven models have been proposed in the econometric literature, little attention has been paid to the problem of smoothing within this class of models in case they are employed as misspecified filters rather than data generating processes.

We aim at filling this gap by introducing a smoothing method for a general class of observation-driven models, namely score-driven models of Creal et al. (2013) and Harvey (2013), also known as "Generalized Autoregressive Score" (GAS) models or "Dynamic Conditional Score" (DCS) models. We show that, in the steady state, Kalman filter and smoothing recursions for linear Gaussian models can be re-written in terms of the score of the conditional density, the Fisher information matrix and a set of static parameters. In particular, the predictive filtering recursion turns out to have the form of score-driven models. The latter can therefore be viewed as approximate filters for nonlinear non-Gaussian models. The performances of these filters have been examined by Koopman et al. (2016), who showed that misspecified score-driven models provide similar forecasting performances as correctly specified parameter-driven models. Based on the same logic, we build a new class of approximate nonlinear smoothers that have similar form to Kalman backward smoothing recursions but employ the score of the non-Gaussian density. The resulting smoothing method is very general, as it can be applied to any observation density, in a similar fashion to score-driven models. We name the newly proposed methodology Score-Driven Smoother (SDS). Similarly, we introduce a Score-Driven Update (SDU) filter, allowing to update predictive filtered estimates once new observations become available.

Smoothing with the SDS requires performing a backward recursion following the standard score-driven forward recursion to filter time-varying parameters. While going backward, the SDS updates filtered estimates by including the effect of actual and future observations and leads to a more efficient reconstruction of time-varying parameters. In our experiments, we have found that, compared to filtered estimates, the SDS provides gains up to 63% in mean square errors, for a wide class of data generating processes. Considered that the likelihood of observation-driven models can be typically written down in closed form, smoothing with the SDS is particularly advantageous from a computational point of view. In contrast, the classical theory of filtering and smoothing for nonlinear non-Gaussian models requires the use of computationally demanding simulation-based techniques (Durbin and Koopman 2012). Another relevant advantage of the SDS over traditional simulation-based methods is that extension to a setting with multiple time-varying parameters is immediate, as it maintains the same simple form as in the univariate case.

[1]The interpretation of GARCH processes as filters is well described in this statement by Nelson (1992): *"Note that our use of the term 'estimate' corresponds to its use in the filtering literature rather than the statistics literature; that is, an ARCH model with (given) fixed parameters produces 'estimates' of the true underlying conditional covariance matrix at each point in time in the same sense that a Kalman filter produces 'estimates' of unobserved state variables in a linear system".*

This general framework allows to construct confidence bands around filtered and smoothed estimates. In observation-driven models, confidence bands are typically needed because static parameters are replaced by their maximum likelihood estimates. In the language of Blasques et al. (2016), this is known as parameter uncertainty. However, if observation-driven models are employed as filters, the latent state variables are not completely revealed by past observations. Thus, also filtering uncertainty has to be considered when building confidence bands. While confidence bands reflecting parameter uncertainty can be built through the methods developed by Blasques et al. (2016), it is less clear how one can take into account filtering uncertainty in observation-driven models. Zamojski (2016) proposed a bootstrap based method to construct in-sample confidence bands for the GARCH. As acknowledged by the author, this method leads to underestimate filtering uncertainty and provides narrow confidence bands. We show that, as a byproduct of our results, one can build both in-sample and out-of-sample confidence bands accounting for filtering uncertainty in score-driven models. We examine in detail the construction of confidence bands in the case of the GARCH model. A general and systematic treatment of filtering uncertainty in score-driven models is provided by Buccheri et al. (2018).

Score-driven models have been successfully applied in the recent econometric literature. For instance, Creal et al. (2011) developed a multivariate dynamic model for volatilities and correlations using fat tailed distributions. Harvey and Luati (2014) described a new framework for filtering with heavy tails while Oh and Patton (2017) introduced high-dimensional factor copula models based on score-driven dynamics for systemic risk assessment. As shown by Blasques et al. (2015), in the class of observation-driven models, score-driven models are locally optimal from an information theoretic perspective. For any score-driven model, one can devise companion SDS and SDU recursions. In particular, the SDS is useful for off-line signal reconstruction and analysis, while the SDU can be used for on-line updating of time-varying parameters. We examine in detail the companion SDS and SDU recursions of popular observation-driven models, namely the GARCH, the MEM model of Engle (2002b) and Engle and Gallo (2006) and an AR(1) model with a time-varying autoregressive coefficient. In oder to show the effectiveness of the proposed methodology in a setting with multiple time-varying parameters, we consider the t-GAS model of Creal et al. (2011) and the Wishart-GARCH model of Gorgi et al. (2018). We show, both on simulated and empirical data, the advantages of SDS and SDU over standard filtered estimates.

A related smoothing technique for a dynamic Student's t location model was introduced by Harvey (2013), who replaced prediction errors in the Kalman smoothing recursions with a martingale difference that is proportional to the score of the t distribution. An application of this smoother can be found in Caivano et al. (2016). The main difference with our approach is that we write the Kalman recursions for the mean of time-invariant linear Gaussian models in a general form that only depends on the score and the Fisher information matrix of the observation density. The resulting smoothing recursions are different by those obtained by Harvey (2013) and are easily applicable to a generic score-driven model by replacing the Gaussian density with the observation density at hand. The SDS is also related to the "approximation via mode estimation" technique described by Durbin and Koopman (2000) and Durbin and Koopman (2012). These authors proved that one can find a sequence of approximating linear Gaussian models enabling the computation of the conditional mode of a non-Gaussian model via a Newton-Raphson algorithm. The main difference with our methodology is that the SDS requires a unique, nonlinear recursion rather than a sequence of Kalman recursions for approximating linear Gaussian models. In addition, in our methodology, the filter coincides with well-known observation-driven model (e.g. GARCH, MEM, ACD, etc) while the approximation via mode estimation technique uses a sequence of filters that are not easily interpretable as dynamic models.

By performing extensive Monte Carlo simulations of nonlinear non-Gaussian state-space models, we compare the performance of the SDS to that of correctly specified parameter-driven models. In particular, we consider two stochastic volatility models and a stochastic intensity models. Importance sampling methods allow to evaluate the full likelihood of these models. The Quasi Maximum Likelihood (QML) method of Harvey et al. (1994) is also considered as a benchmark when estimating the two stochastic volatility models. Compared to correctly specified models, the losses incurred by the SDS are very small in all the simulated scenarios and are always lower, on average, than 2.5% in mean square errors. Moreover, the SDS systematically outperforms the QML. Computational times are decisively in favour of the SDS. For the models used in the simulation study, we found that smoothing with the SDS is on average 215 times faster than smoothing with efficient importance sampling techniques. The advantages of the proposed method are also shown on empirical data. Using realized covariance as a proxy of latent covariance, we show that SDU and SDS covariance estimates obtained through the dynamic t-GAS model fitted on Russel 3000 stock returns are superior to standard filtered score-driven estimates. The analysis allows to examine the informational content of present and future log-returns from a dynamic covariance modelling perspective.

The rest of the paper is organized as follows: section 2 introduces the SDS and provides the main theoretical results; section 3 describes several examples of SDS's and discusses how to construct confidence bands; section 4 shows the results of the Monte Carlo study; in section 5 the SDS is applied on an empirical analysis involving assets of Russel 3000 index; section 6 concludes.

2 Theoretical framework

In this section, we discuss in detail the main theoretical results conveying to the formulation of our approximate, nonlinear smoothing technique. We start by showing that, in the steady state, the classical Kalman filter and smoothing recursions for linear Gaussian models can be re-written in an alternative form that only involves the score of the conditional likelihood, the Fisher information matrix and a set of static parameters. Abstracting from the linear Gaussian setting, these recursions can be viewed as the approximate filtering and smoothing recursions for a non-Gaussian model by computing scores and information based on the non-Gaussian density. We then show that filtering uncertainty in score-driven models can be evaluated as an immediate byproduct of our results.

2.1 Kalman filtering and smoothing

Let us consider a linear Gaussian state-space representation:

$$y_t = Z\alpha_t + \epsilon_t, \quad \epsilon_t \sim \text{NID}(0, H) \tag{1}$$

$$\alpha_{t+1} = c + T\alpha_t + \eta_t, \quad \eta_t \sim \text{NID}(0, Q) \tag{2}$$

where $\alpha_t \in \mathbb{R}^m$ is a column vector of state variables and $y_t \in \mathbb{R}^p$ is a column vector of observations. The parameters $Z \in \mathbb{R}^{p\times m}$, $H \in \mathbb{R}^{p\times p}$, $T \in \mathbb{R}^{m\times m}$ and $Q \in \mathbb{R}^{m\times m}$ are system matrices. Let $\mathcal{F}_t$ denote the set of observations up to time t, namely $\mathcal{F}_t = \{y_1, \dots, y_t\}$. We are interested in updating our knowledge of the underlying state variable α_t when a new observation y_t becomes available and to predict α_{t+1} based on the last observations $y_1, \dots, y_t$. We thus define:

$$a_{t|t} = \text{E}[\alpha_t|\mathcal{F}_t], \quad P_{t|t} = \text{Var}[\alpha_t|\mathcal{F}_t] \tag{3}$$

$$a_{t+1} = \text{E}[\alpha_{t+1}|\mathcal{F}_t], \quad P_{t+1} = \text{Var}[\alpha_{t+1}|\mathcal{F}_t] \tag{4}$$

The Kalman filter allows to compute recursively $a_{t|t}$, $P_{t|t}$, a_{t+1} and P_{t+1}. Assuming $\alpha_1 \sim \mathrm{N}(a_1, P_1)$, where a_1 and P_1 are known, for $t = 1, \ldots, n$, we have (Harvey 1991, Durbin and Koopman 2012):

$$v_t = y_t - Z a_t \tag{5}$$

$$a_{t|t} = a_t + P_t Z' F_t^{-1} v_t \tag{6}$$

$$a_{t+1} = c + T a_t + K_t v_t \tag{7}$$

and

$$F_t = Z P_t Z' + H \tag{8}$$

$$P_{t|t} = P_t - P_t Z' F_t^{-1} Z P_t \tag{9}$$

$$P_{t+1} = T P_t (T - K_t Z)' + Q \tag{10}$$

where $K_t = T P_t Z' F_t^{-1}$. The log-likelihood can be computed in the prediction error decomposition form, namely:

$$\log p(y_t | \mathcal{F}_{t-1}) = \text{const} - \frac{1}{2} \left(\log |F_t| + v_t' F_t^{-1} v_t \right) \tag{11}$$

Smoothed estimates $\hat{\alpha}_t = \mathrm{E}[\alpha_t | \mathcal{F}_n]$, $\hat{P}_t = \mathrm{Var}[\alpha_t | \mathcal{F}_n]$, $n > t$, can be computed through the following backward recursions:

$$r_{t-1} = Z' F_t^{-1} v_t + L_t' r_t \tag{12}$$

$$\hat{\alpha}_t = a_t + P_t r_{t-1} \tag{13}$$

and

$$N_{t-1} = Z' F_t^{-1} Z + L_t' N_t L_t \tag{14}$$

$$\hat{P}_t = P_t - P_t N_{t-1} P_t \tag{15}$$

where $L_t = T - K_t Z$, $r_n = 0$, $N_n = 0$ and $t = n, \ldots, 1$. The conditional distribution of α_t is Gaussian with mean and variance given by (a_{t+1}, P_{t+1}), $(a_{t|t}, P_{t|t})$, $(\hat{\alpha}_t, \hat{P}_t)$, depending on the conditioning set.

2.2 A more general representation

In Appendix A we prove the following:

Proposition 1 *In the steady state, Eq. (6), (7), (12), (13) can be written as:*

$$a_{t|t} = a_t + T^{-1} R \nabla_t \tag{16}$$

$$a_{t+1} = c + T a_t + R \nabla_t \tag{17}$$

and

$$r_{t-1} = \nabla_t + (T - R\mathcal{I})' r_t \tag{18}$$

$$\hat{\alpha}_t = a_t + T^{-1} R r_{t-1} \tag{19}$$

where $\nabla_t = \frac{\partial \log p(y_t | \mathcal{F}_{t-1})}{\partial a_t}$, $\mathcal{I} = Z' \bar{F}^{-1} Z$, $\bar{F} = Z' \bar{P} Z + H$, $R = T \bar{P}$ *and* $\bar{P}$ *is the steady state variance matrix which is the solution of the matrix Riccati equation:*

$$\bar{P} = T \bar{P} T' - T \bar{P} Z' \bar{F}^{-1} Z \bar{P} T' + Q \tag{20}$$

Note that a steady state solution exists whenever the system matrices are constant (Harvey 1991, Durbin and Koopman 2012). In this case, the variance matrix P_t converges to $\bar{P}$ after few time steps. The new Kalman recursions for the mean are re-parameterized in terms of the score ∇_t and the Fisher information matrix $\mathcal{I}$. This representation is equivalent to the one in equations (6), (7) and (12), (13). However, it is more general, as it only relies on the predictive density $p(y_t|\mathcal{F}_{t-1})$. In principle, the forward recursions (16), (17) and the backward recursions (18), (19) can be applied to any parameter-driven model for which a measurement density $p(y_t|\mathcal{F}_{t-1})$ is defined.

2.3 SDS recursions

Note that the predictive filter (17) has an autoregressive structure and is driven by the score of the conditional likelihood, i.e. it has the form of score-driven models of Creal et al. 2013 and Harvey 2013. Thus, if one looks at score-driven models as filters, it turns out that the score-driven filter (SDF hereafter) is optimal in case of linear Gaussian models. In case of nonlinear non-Gaussian models, the SDF can be regarded as an approximate nonlinear filter. The main difference with the Kalman filter is that the Gaussian score is replaced by the score of the true conditional density, thus providing robustness to non-Gaussianity. As shown by Koopman et al. (2016), score-driven filters have similar predictive accuracy as correctly specified nonlinear non-Gaussian models, while at the same time providing significant computational gains. Indeed, the likelihood can be written in closed form and standard quasi-Newton techniques can be employed for optimization.

Based on the same principle, we introduce an approximate nonlinear smoother allowing to estimate time-varying parameters using all available observations. In case of linear Gaussian models, the Kalman smoother is a minimum variance linear unbiased estimator (MVLUE) of the state. Thus, we define our smoother in such a way that it coincides with the latter in this specific case. In case of nonlinear non-Gaussian models, it maintains the same simple form of Kalman backward smoothing recursions but replaces the Gaussian score with the one of the non-Gaussian density.

Let us assume that observations $y_t \in \mathbb{R}^p$, $t = 1, \dots, n$, are generated by the following observation density:

$$y_t|f_t \sim p(y_t|f_t, \Theta) \tag{21}$$

where $f_t \in \mathbb{R}^k$ is a vector of time-varying parameters and Θ is a vector of static parameters. We generalize the filtering and smoothing recursions (16)-(19) for the measurement density $p(y_t|f_t, \Theta)$ as:

$$f_{t|t} = f_t + B^{-1}A\nabla_t \tag{22}$$

$$f_{t+1} = \omega + A\nabla_t + Bf_t \tag{23}$$

$t = 1, \dots, n$ and:

$$r_{t-1} = \nabla_t + (B - A\mathcal{I}_{t|t-1})'r_t \tag{24}$$

$$\hat{f}_t = f_t + B^{-1}Ar_{t-1} \tag{25}$$

where $r_n = 0$ and $t = n, \dots, 1$. The predictive filter in equation (23) has the same form of score-driven models. The term ∇_t is now the score of the measurement density $p(y_t|f_t, \Theta)$, namely:

$$\nabla_t = \frac{\partial \log p(y_t|f_t, \Theta)}{\partial f_t} \tag{26}$$

while $\mathcal{I}_{t|t-1} = \mathrm{E}_{t-1}[\nabla_t \nabla_t']$ is the information matrix, which may be time-varying. The vector $\omega \in \mathbb{R}^k$ and the two matrices $A, B \in \mathbb{R}^{k \times k}$ are static parameters included in Θ. They are estimated by maximizing the log-likelihood, namely:

$$\hat{\Theta} = \underset{\Theta}{\operatorname{argmax}} \sum_{t=1}^{n} \log p(y_t | f_t, \Theta) \tag{27}$$

Thus, one can run the backward smoothing recursions (24), (25) after computing the forward filtering recursions (22), (23), in a similar fashion to Kalman filter and smoothing recursions. Note that the above recursions are nonlinear, as the score of a non-Gaussian density is typically nonlinear in the observations. The filter $f_{t|t}$ in equation (22) allows to update the current estimate f_t once a new observation y_t becomes available. While going backward, the smoothing recursions (24), (25) update the two filters f_t and $f_{t|t}$ using all available observations. Smoothed estimates $\hat{f}_t$ are generally less noisy than filtered estimates $f_{t|t}$, f_t and provide a more accurate reconstruction of the time-varying parameters.

It is a standard practice in score-driven models replacing the score ∇_t with the scaled score $s_t = S_t \nabla_t$. The role of the scaling matrix S_t is to take into account the curvature of the log-likelihood function. Creal et al. (2013) discussed several choices of S_t based on inverse powers of the information matrix $\mathcal{I}_{t|t-1}$. For instance, given a normal density with time-varying variance, if $S_t = \mathcal{I}_{t|t-1}^{-1}$, one recovers the standard GARCH model. The filtering and smoothing recursions (22)-(25) are obtained if one sets S_t equal to the identity matrix. When using a scaled score s_t, the filtering recursions (22), (23) become:

$$f_{t|t} = f_t + B^{-1} A s_t \tag{28}$$

$$f_{t+1} = \omega + A s_t + B f_t \tag{29}$$

Since the score is now scaled by S_t, the term $A\mathcal{I}_{t|t-1}$ in equation (24) has to take into account the new normalization. We thus replace A with AS_t. As a result, we obtain the general backward smoothing recursions:

$$r_{t-1} = s_t + (B - A S_t \mathcal{I}_{t|t-1})' r_t \tag{30}$$

$$\hat{f}_t = f_t + B^{-1} A r_{t-1} \tag{31}$$

Note that the second equation is unaffected, as the term r_{t-1} already corrects for the scaling. For instance, if $S_t = \mathcal{I}_{t|t-1}^{-1}$, we obtain:

$$r_{t-1} = s_t + (B - A)' r_t \tag{32}$$

$$\hat{f}_t = f_t + B^{-1} A r_{t-1} \tag{33}$$

that is, the information matrix $\mathcal{I}_{t|t-1}$ disappears because its effect is already taken into account when scaling the score. If $S_t = \mathcal{I}_{t|t-1}^{-1/2}$, we get:

$$r_{t-1} = s_t + (B - A \mathcal{I}_{t|t-1}^{1/2})' r_t \tag{34}$$

$$\hat{f}_t = f_t + B^{-1} A r_{t-1} \tag{35}$$

From a computational point of view, the backward recursions (30), (31) are simple since s_t and $\mathcal{I}_{t|t-1}$ are typically available from the forward filtering recursion. We term the approximate smoother obtained through recursions (30), (31) as Score-Driven Smoother (SDS). Basically, for any score-driven model, one can devise a companion SDS recursion that only requires the s_t, $\mathcal{I}_{t|t-1}$ and the static parameters, as estimated through the SDF. Note that the

forward recursion (28) is the analogue of recursion (6) in the Kalman filter and allows to update SDF estimates once a new observation y_t becomes available. We denote the approximate Score-Driven Update filter (28) by SDU. The proposed methodology can thus be schematically represented through the following procedure:

1. Estimation of static parameters:
$$\tilde{\Theta} = \underset{\Theta}{\operatorname{argmax}} \sum_{t=1}^{n} \log p(y_t|f_t, \Theta)$$

2. Forward predictive and update filter:
$$f_{t+1} = \tilde{\omega} + \tilde{A}s_t + \tilde{B}f_t$$
$$f_{t|t} = f_t + \tilde{B}^{-1}\tilde{A}s_t$$

3. Backward smoother:
$$r_{t-1} = s_t + (\tilde{B} - \tilde{A}S_t\mathcal{I}_{t|t-1})'r_t$$
$$\hat{f}_t = f_t + \tilde{B}^{-1}\tilde{A}r_{t-1}$$

2.4 Filtering uncertainty

The general framework developed in section (2.3) also allows to construct in-sample and out-of-sample confidence bands around filtered and smoothed estimates. As underlined by Blasques et al. (2016), confidence bands can reflect both parameter and filtering uncertainty. Parameter uncertainty is related to the fact that static parameters are replaced by their maximum likelihood estimates. Both observation-driven and parameter-driven models are affected by parameter uncertainty. In observation-driven models, confidence bands reflecting parameter uncertainty can be constructed through the methods developed by Blasques et al. (2016). Filtering uncertainty is related to the fact that time-varying parameters are not completely revealed by observations. As such, it is absent in observation-driven models, where time-varying parameters are deterministic functions of past observations. However, if observation-driven models are regarded as filters, one is interested in constructing confidence bands around filtered and smoothed estimates reflecting the conditional distribution of the underlying state variable.

In linear Gaussian models, filtering uncertainty can be assessed through the variance matrices P_{t+1}, $P_{t|t}$, $\hat{P}_t$ introduced in section (2.1), which provide the conditional variance of the unobserved state variable. It is instead less clear how one can quantify filtering uncertainty in misspecified observation-driven models. Zamojski (2016) proposed a bootstrap based method for assessing filtering uncertainty in GARCH filters. Confidence bands constructed through this technique tend to underestimate filtering uncertainty, because they are based on bootstraps of the filter rather than the underlying state variable. In addition, the method of Zamojski (2016) does not allow to construct out-of-sample confidence bands, which are often needed in practical applications.

In our framework, in-sample and out-sample confidence bands can be constructed by exploiting the relation between Kalman filter recursions and score-driven recursions. In section 2.2, we have shown that the steady state variance matrix $\bar{P}$ can be expressed as:

$$\bar{P} = T^{-1}R \tag{36}$$

In the score-driven framework, the analogue of P_{t+1}, which we denote by J_{t+1}, is then given by:

$$J_{t+1} = B^{-1}AS_t \tag{37}$$

where the scaling matrix S_t is introduced to take into account different normalizations of the score. From eq. (9), (A.2), the analogue of $P_{t|t}$ is:

$$J_{t|t} = J_t - J_t \mathcal{I}_{t|t-1} J_t \tag{38}$$

Similarly, the analogue of $\hat{P}_t$, from eq. (14), (15), is:

$$N_{t-1} = \mathcal{I}_{t|t-1} + (B - AS_t\mathcal{I}_{t-1})' N_t (B - AS_t\mathcal{I}_{t-1}) \tag{39}$$

$$\hat{J}_t = J_t - J_t N_{t-1} J_t \tag{40}$$

with $N_n = 0$ and $t = n, \ldots, 1$.

Confidence bands can be computed as quantiles of the conditional distribution of the state variable. For a general state-space model, the latter is non-Gaussian and is not known analytically. Assuming a Gaussian density generally leads to underestimate filtering uncertainty, as the true conditional density is typically fat-tailed. In order to construct robust confidence bands, we use a more flexible density determined by matching location and scale parameters with those of the normal density. This method is described in its full generality by Buccheri et al. (2018). In section 3.2, we show an application to the GARCH and assess the performance of robust confidence bands in a simulation study.

3 Examples of SDS recursions

In this section we provide several examples of SDS estimates. As a first step, we focus on two volatility models that are quite popular in the econometric literature, namely the GARCH model of Bollerslev (1986) and the multiplicative error model (MEM) of Engle (2002b) and Engle and Gallo (2006). These are score-driven models which are susceptible of treatment within our framework. As a third example, we present an AR(1) model with a score-driven autoregressive coefficient. The time-varying autoregressive coefficient allows to capture temporal variations in persistence, as well as nonlinear dependencies (Blasques et al. 2014). Autoregressive models with time-varying coefficients have been employed by Monache and Petrella (2017) and Buccheri and Corsi (2017) for inflation and volatility forecasting, respectively.

One of the advantages of the SDS recursions (30), (31) is that they maintain the same simple form when $f_t \in \mathbb{R}^k$, $k > 1$ is a vector containing multiple time-varying parameters. In this multivariate setting, the use of simulation-based techniques would be highly computationally demanding. In order to test the SDS in a multivariate setting, we consider the t-GAS model of Creal et al. (2011) and the Wishart-GARCH model of Gorgi et al. (2018). The former is a conditional correlation model for heavy-tail returns while the latter is a joint model for the dynamics of daily returns and realized covariance matrices. In these models, the number of time-varying parameters grows as the square of the number of assets and therefore they provide an interesting multivariate framework in which to assess the performance of the SDS.

1. GARCH-SDS

Consider the model:

$$y_t = \sigma_t \epsilon_t, \quad \epsilon_t \sim \text{NID}(0, 1) \tag{41}$$

The predictive density is thus:

$$p(y_t|\sigma_t^2) = \frac{1}{\sqrt{2\pi\sigma_t^2}} e^{-\frac{y_t^2}{2\sigma_t^2}} \tag{42}$$

Setting $f_t = \sigma_t^2$ and $S_t = \mathcal{I}_{t|t-1}^{-1}$, equations (28), (29) reduce to:

$$f_{t|t} = f_t + B^{-1}A(y_t^2 - f_t) \tag{43}$$

$$f_{t+1} = \omega + A(y_t^2 - f_t) + Bf_t \tag{44}$$

In particular, the predictive filter (44) is the standard GARCH(1,1) model. The smoothing recursions (30), (31) reduce to:

$$r_{t-1} = y_t^2 - f_t + (B - A)'r_t \tag{45}$$

$$\hat{f}_t = f_t + B^{-1}Ar_{t-1} \tag{46}$$

$t = n, \ldots, 1$.

2. MEM-SDS

Consider the model:

$$y_t = \mu_t \epsilon_t \tag{47}$$

where ϵ_t has a Gamma distribution with density $p(\epsilon_t|\alpha) = \Gamma(\alpha)^{-1}\epsilon_t^{\alpha-1}\alpha^\alpha e^{-\alpha\epsilon_t}$. The predictive density is thus given by:

$$p(y_t|\mu_t, \alpha) = \Gamma(\alpha)^{-1} y_t^{\alpha-1}\alpha^\alpha \mu_t^{-\alpha} e^{-\alpha\frac{y_t}{\mu_t}} \tag{48}$$

Setting $f_t = \mu_t$ and $S_t = \mathcal{I}_{t|t-1}^{-1}$, equations (28), (29) reduce to:

$$f_{t|t} = f_t + B^{-1}A(y_t - f_t) \tag{49}$$

$$f_{t+1} = \omega + A(y_t - f_t) + Bf_t \tag{50}$$

In particular, the predictive filter (50) is the standard MEM(1,1) model. The smoothing recursions (30), (31) reduce to:

$$r_{t-1} = y_t - f_t + (B - A)'r_t \tag{51}$$

$$\hat{f}_t = f_t + B^{-1}Ar_{t-1} \tag{52}$$

$t = n, \ldots, 1$.

3. AR(1)-SDS

Consider the model:

$$y_t = c + \alpha_t y_{t-1} + \epsilon_t, \quad \epsilon_t \sim \mathrm{N}(0, q^2) \tag{53}$$

The predictive density is thus given by:

$$p(y_t|\alpha_t) = \frac{1}{\sqrt{2\pi}q}\exp\left[-\frac{1}{2}\left(\frac{y_t - c - \alpha_t y_{t-1}}{q}\right)^2\right] \tag{54}$$

Setting $f_t = \alpha_t$ and $S_t = \mathcal{I}_{t|t-1}^{-1}$, equations (28), (29) reduce to:

$$f_{t|t} = f_t + B^{-1}A\left(\frac{y_t - c - f_t y_{t-1}}{y_{t-1}}\right) \tag{55}$$

$$f_{t+1} = \omega + A\left(\frac{y_t - c - f_t y_{t-1}}{y_{t-1}}\right) + Bf_t \tag{56}$$

while the smoothing recursions (30), (31) reduce to:

$$r_{t-1} = \left(\frac{y_t - c - f_t y_{t-1}}{y_{t-1}} \right) + (B - A)' r_t \tag{57}$$

$$\hat{f}_t = f_t + B^{-1} A r_{t-1} \tag{58}$$

$t = n, \ldots, 1$.

4. t-GAS-SDS

Let $r_t \in \mathbb{R}^p$ denote a vector of demeaned daily log-returns. Consider the following observation density:

$$p(r_t|V_t, \nu) = \frac{\Gamma((\nu+p)/2)}{\Gamma(\nu/2)[(\nu-2)\pi]^{p/2}|V_t|^{1/2}} \left[1 + \frac{r_t' V_t^{-1} r_t}{(\nu-2)} \right]^{-\frac{\nu+p}{2}} \tag{59}$$

where $V_t \in \mathbb{R}^{p \times p}$ is a time-varying covariance matrix and $\nu > 2$ is the number of degrees of freedom. Note that $p(r_t|V_t, \nu)$ is a normalized Student t distribution such that $\text{Cov}(r_t|V_t, \nu) = V_t$. Applying the filtering equation (29) leads to the t-GAS model of Creal et al. (2011). Closed form formulas for the score and information matrix are reported in Creal et al. (2011). These authors also proposed two parameterizations of V_t leading to positive-definite estimates. The first is similar to the one used in the DCC model of Engle (2002a), while the second is based of hyperspherical coordinates. In the two parameterizations, the number of time-varying parameters is $k = p + p(p+1)/2$ and $k = p(p+1)/2$, respectively.

5. Wishart-GARCH-SDS

Let us assume that, in addition to daily log-returns r_t, we can compute realized measures from the intraday returns of the p assets. Let $X_t \in \mathbb{R}^{p \times p}$ denote a positive definite estimate of the realized covariance matrix. Let also $\mathcal{F}_t$ denote the σ-field generated by r_t and X_t. The observation density in the Wishart-GARCH model is:

$$r_t|\mathcal{F}_{t-1} \sim N_k(0, V_t) \tag{60}$$

$$X_t|\mathcal{F}_{t-1} \sim W_k(V_t/\nu, \nu) \tag{61}$$

where $N_k(0, V_t)$ is a multivariate zero-mean normal distribution with covariance matrix V_t and $W_k(V_t/\nu, \nu)$ is a Wishart distribution with mean V_t and degrees of freedom $\nu \geq p$. Assuming that r_t and X_t are conditionally independent given $\mathcal{F}_{t-1}$, the conditional log-likelihood can be written as:

$$\log L(r_t, X_t|V_t, \mathcal{F}_{t-1}) = \log L(r_t|V_t, \mathcal{F}_{t-1}) + \log L(X_t|V_t, \mathcal{F}_{t-1}) \tag{62}$$

where:

$$\log L(r_t|V_t, \mathcal{F}_{t-1}) = \frac{1}{2} d_r(p) - \frac{1}{2} \log|V_t| - \frac{1}{2} \text{tr}(V_t^{-1} r_t r_t') \tag{63}$$

$$\log L(X_t|V_t, \mathcal{F}_{t-1}) = \frac{1}{2} d_X(p) + \frac{\nu - p - 1}{2} \log|X_t| - \frac{\nu}{2} \log|V_t| - \frac{\nu}{2} \text{tr}(V_t^{-1} X_t) \tag{64}$$

Here $d_r(p) = -p \log(2\pi)$, $d_X(p, \nu) = \nu p \log(\nu/2) - 2 \log \Gamma_p(\nu/2)$ and Γ_p is the multivariate Gamma function of order p. We denote the vector of time-varying covariances by $f_t = \text{vech}(V_t) \in \mathbb{R}^k$, $k = \frac{p(p+1)}{2}$. The score $\nabla_t = \frac{\partial \log L(r_t, X_t|f_t, \mathcal{F}_{t-1})}{\partial f_t} \in \mathbb{R}^k$ and the information matrix $\mathcal{I}_{t|t-1} = \text{E}_{t-1}[\nabla_t \nabla_t'] \in \mathbb{R}^{k \times k}$ can be computed as reported in Gorgi et al. (2018). Opschoor et al. (2017) proposed an alternative specification with a heavy tail distribution for both returns and realized measures. Similar SDS recursions can be recovered for this fat-tail specification using our general framework.

Figures 1 - 3 show several examples of SDS estimates from the above models. The time-varying parameters follow both deterministic and stochastic patterns and are generated as described in the next paragraph.

3.1 Comparison of SDF and SDS estimates

In order to show the effectiveness of the proposed methodology, we compare SDF and SDS estimates. It is natural expecting that SDS estimates are affected by lower estimation errors, as they use more information when reconstructing time-varying parameters. However, comparing with the latter allows to provide a quantitative assessment of the benefits of using the SDS in place of standard score-driven estimates.

We first focus on the univariate models (GARCH, MEM, AR(1)) and simulate $N = 250$ time-series of $n = 4000$ observations with different dynamic patterns for the time-varying parameters. The first 2000 observations are used to estimate the models while the remaining observations are used for testing. Let β_t generically denote the time-varying parameters σ_t^2, μ_t and α_t in the three models. We consider the following data generating processes for β_t:

1. Slow sine: $\beta_1 + \frac{1}{2}\sin(\frac{2\pi t}{N})$
2. Fast sine: $\beta_1 + \frac{1}{2}\sin(\frac{20\pi t}{N})$
3. Ramp: $\beta_1 + \frac{1}{N} \mod (t, \frac{N}{\omega})$
4. Step: $\beta_1 + \frac{1}{2}(t > \frac{N}{2})$
5. Model: $\beta_{t+1} = c + \varphi\beta_t + \xi_t, \quad \xi_t \sim \mathrm{N}(0, \sigma^2)$

$t = 1, \dots, N$. We set $\omega = 4$, $k = 1$, $c = 0.01$, $\varphi = 0.98$, $\sigma^2 = 0.5$. For some of these dynamic specifications, figure 1 shows examples of filtered and smoothed estimates of time-varying parameters obtained through the GARCH. As expected, SDS estimates are less noisy than filtered estimates and provide a more accurate reconstruction of time-varying parameters.

Table 1 shows average MSE and MAE of SDF and SDS estimates, for all the patterns considered above. We also report MSE and MAE obtained through the SDU filter in equation (28). The latter updates f_t once a new observation y_t arrives. This translates into a slight improvement over filtered estimates. The SDS, using all available observations, significantly improves on SDF estimates, with relative gains larger than 35% and lower than 62% in mean square errors.

We now consider the two multivariate models, namely the t-GAS and the Wishart-GARCH. We compare SDF estimates to SDU and SDS estimates in a simulation setting where $N = 250$ time series of $n = 2000$ daily realized covariance matrices and log-returns are generated as described in Appendix B. The aim of the experiment is to estimate the true covariance matrix V_t from observations of daily returns r_t in the t-GAS model and from observations of both daily returns r_t and realized covariance matrices X_t in the Wishart-GARCH model. We consider three scenarios where the number of assets is $p = 5, 10, 20$ respectively, and thus we have $k = 15, 55, 210$ time-varying covariances[2].

For $p = 5$, figures 2 and 3 compare SDF and SDS estimates of the $V_t(1,1)$ and $V_t(1,2)$ elements of the simulated covariance matrix in the t-GAS and Wishart-GARCH models, respectively. As in the previous univariate cases, smoothed estimates provide a better reconstruction of the time-varying covariances. Note that, compared to the t-GAS model, the Wishart-GARCH provides estimates which are closer to the simulated V_t, as they are obtained by conditioning on a larger information set.

In order to quantify estimation errors, we use the root mean square error (RMSE) and the quasi-likelihood (Qlike), which are robust loss measures for covariance estimates (Patton 2011). These are defined in Appendix B. Table 2

[2]We implement the t-GAS model using hyperspherical coordinates, and thus we have $k = p(p+1)/2)$ time-varying covariances.

shows relative RMSE and Qlike gains of SDU and SDS estimates over SDF. We first note that SDU and SDS provide significantly lower RMSE. In the t-GAS model, the relative gain of the SDU is roughly equal to 3%, while the one of SDS is larger than 14% and lower than 19%. In the Wishart-GARCH model, the relative gain of SDU is larger than 7% and lower than 13%, while the one of the SDS is larger than 13% and lower than 19%. It is interesting to note that SDU gains are significantly larger in the Wishart-GARCH model. This is due to the fact that today's realized covariance X_t is a highly informative proxy of V_t, thus leading to drastic RMSE reduction when included in the information set. In contrast, daily returns are less informative and thus it is necessary to include all available observations to achieve significant RMSE reduction in the t-GAS model. If one looks at the Qlike loss, relative gains of SDU and SDS are moderate compared to RMSE but they are statistically significant. Even in this case, SDU gains are larger in the Wishart-GARCH model due to the highly informative content of realized covariance measures.

3.2 Confidence bands

In section (2.4), we have seen that an estimate of the conditional variance of the state variable is given by the variance matrices J_{t+1}, J_t and $\hat{J}_t$ defined in eq. (37), (38), (40). As in the Kalman filter, one can use these variances to construct confidence bands around filtered and smoothed estimates. However, the conditional density of the state variable is typically fat-tailed and cannot be written in closed form. Assuming normality generally provides narrow confidence bands and thus underestimates filtering uncertainty.

Robust in-sample and out-of-sample confidence bands can be constructed by computing quantiles of a more flexible distribution determined by matching location and scales parameters. We illustrate here an application of this technique in the case of the GARCH and provide a systematic treatment in Buccheri et al. (2018).

Let us consider the following stochastic volatility model:

$$y_t = e^{\frac{\theta_t}{2}}\epsilon_t, \quad \epsilon_t \sim \mathrm{N}(0,1) \tag{65}$$

$$\theta_{t+1} = \gamma \ + \phi\theta_t + \eta_t, \quad \eta_t \sim \mathrm{N}(0,\sigma_\eta^2) \tag{66}$$

We are interested in computing quantiles of the conditional density of e^{θ_t}. Filtered and smoothed estimates of the latent log-variance θ_t are recovered by computing the score-driven recursions for the following observation density:

$$p(y_t|f_t) = \frac{1}{\sqrt{2\pi e^{f_t}}} e^{-\frac{y_t^2}{2e^{f_t}}}$$

As an outcome of this procedure, we also obtain the conditional variances J_{t+1}, J_t and $\hat{J}_t$. Let $F_{\theta_t|\mathcal{F}_{t_1}}(\theta) = P(\theta_t < \theta|\mathcal{F}_{t_1})$, $1 < t_1 < n$, be the conditional distribution function of θ_t. The quantile function of e^{θ_t} is then given by:

$$F^{-1}_{e^{\theta_t}|\mathcal{F}_{t_1}} = \exp\left(F^{-1}_{\theta_t|\mathcal{F}_{t_1}}\right) \tag{67}$$

As a first approximation, we compute quantiles by assuming $\theta_t|\mathcal{F}_{t_1} \sim N(f_{t|t_1}, J_{t|t_1})$. For $t_1 = t-1$, $t_1 = t$, $t_1 = n$ we obtain $N(f_t, J_t)$, $N(f_{t|t}, J_{t|t})$, $N(\hat{f}_t, \hat{J}_t)$, respectively. These conditional densities depend on parameters which are an output of the score-driven recursions and thus confidence bands can be easily computed through eq. (67) using the Gaussian quantile function. We then assume $\theta_t|\mathcal{F}_{t_1} \sim t(f_{t|t_1}, J_{t|t_1}, \nu)$, i.e. a Student's t-distribution with location $f_{t|t_1}$, scale $J_{t|t_1}$ and ν degrees of freedom. If $\nu \to \infty$, we recover the Gaussian confidence bands. However, if ν is finite, confidence bands will be larger and provide a better approximation to the true filtering uncertainty. In this example, the parameter ν is chosen by fitting a t distribution on the residuals of an AR(1) model estimated on $\hat{f}_t$. More sophisticated techniques are developed in Buccheri et al. (2018).

In order to test the quality of confidence bands, we generate 1000 time series of $n = 4000$ observations of the stochastic volatility model (65), (66). The values of static parameters are chosen in order to be similar to those obtained when estimating the model on real financial returns: $\gamma = 0.001$, $\phi = 0.98$, $\sigma_\eta^2 = 0.02$. Figure (4) shows one of the simulated patterns, together with 95% confidence bands for filtered and smoothed estimates computed through the method described above.

We estimate the GARCH in the sub-sample comprising the first 2000 observations and construct in-sample confidence bands. In the remaining sub-sample of 2000 observations, out-of-sample bands are constructed using previous parameter estimates. Both Gaussian and robust confidence bands are built at 90%, 95%, 99% nominal confidence levels. We compare the nominal confidence level to the coverage, defined as the fraction of times the true variance path e^{θ_t} is inside the confidence bands. Table (3) shows average coverages for in-sample and out-of-sample SDF, SDU and SDS confidence bands. As expected, confidence bands constructed by assuming a Gaussian density provide an average coverage which is significantly lower than the nominal confidence level, meaning that they underestimate filtering uncertainty. In contrast, the average coverage of robust confidence bands is very close to the nominal coverage. Similar results are found when changing the variance σ_η^2 in the latent process. In particular, for larger values of σ_η^2, the quality of Gaussian confidence bands further deteriorates, while robust bands still provide a good matching to the nominal level. A systematic treatment of the technique described here and, more generally, of filtering uncertainty in observation-driven models can be found in Buccheri et al. (2018).

4 Monte Carlo analysis

In this section we perform extensive Monte Carlo simulations to test the performance of the SDS under different dynamic specifications for the time-varying parameters. Since we interpret the SDS as an approximate smoother for nonlinear non-Gaussian models, we compare its performance to that of correctly specified parameter-driven models. The main idea is to examine the extent to which the approximation leads to similar results as correctly specified parameter-driven models. In this case, the use of the SDS would be particularly advantageous from a computational point of view, as the likelihood of score-driven models can be written in closed form and smoothing can be performed through a simple backward recursion. This analysis is similar in spirit to that of Koopman et al. (2016), who compared score-driven models to correctly specified parameter-driven models and found that the two classes of models have similar predictive accuracy, with very small average losses. We find a similar result for the SDS.

4.1 Linear non-Gaussian models

We first consider an AR(1) model with a t-distributed measurement error:

$$y_t = \alpha_t + \epsilon_t, \quad \epsilon_t \sim t(0, \sigma_\epsilon^2, \nu) \tag{68}$$

$$\alpha_{t+1} = c + \phi\alpha_t + \eta_t, \quad \eta_t \sim \mathrm{N}(0, \sigma_\eta^2) \tag{69}$$

We choose $c = 0.01$ and $\phi = 0.95$. The signal-to-noise ratio is defined as $\delta = \frac{\sigma_\eta^2}{\sigma_\epsilon^2}$. The corresponding observation driven model is a t-location model (Harvey 2013) with predictive density:

$$p(y_t|f_t;\varphi,\beta) = \frac{\Gamma[(\beta+1)/2]}{\Gamma(\beta/2)\varphi\sqrt{\pi\beta}}\left[1 + \frac{(y_t - f_t)^2}{\beta\varphi^2}\right]^{-(\beta+1)/2} \tag{70}$$

Setting $S_t = \mathcal{I}_{t|t-1}^{-1}$, equation (29) reduces to:

$$f_{t+1} = \omega + A(\beta+3)\frac{y_t - f_t}{\beta + \left(\frac{y_t - f_t}{\varphi}\right)^2} + Bf_t \tag{71}$$

while the smoothing recursions (30), (31) reduce to:

$$r_{t-1} = (\beta+3)\frac{y_t - f_t}{\beta + \left(\frac{y_t - f_t}{\varphi}\right)^2} + (B-A)'r_t \tag{72}$$

$$\hat{f}_t = f_t + B^{-1}Ar_{t-1} \tag{73}$$

$t = n, \ldots, 1$. We compare standard Kalman filtered and smoothed estimates to SDF, SDU and SDS estimates. Similarly to previous simulation studies, we generate 1000 time series of 4000 observations and use the first subsample of 2000 observations for estimation and the remaining observations for testing. Table 4 shows relative MSE and MAE for different values of ν. Note that SDF, SDU and SDS provide better estimates than standard Kalman filter and smoother. In particular, we observe large differences for low values of ν, where the t-distribution strongly deviates from the Gaussian, and for low values of δ, at which accounting for the non-normality of the measurement error becomes more important. Note also that gains of SDS over Kalman smoother estimates are larger than gains of SDF over the Kalman filter for low ν and δ. These results confirm the ability of the SDS to provide robust smoothed estimates of time-varying parameters, to the same extent as the SDF provides robust filtered estimates of time-varying parameters in presence of a non-Gaussian prediction density.

4.2 Nonlinear non-Gaussian models

We now examine the behavior of the SDS in presence of nonlinear non-Gaussian parameter-driven models. In particular, we consider the following three specifications, which are quite popular in the econometric literature:

1. Stochastic volatility model with Gaussian measurement density[3]:

$$y_t = e^{\frac{\theta_t}{2}}\epsilon_t, \quad \epsilon_t \sim \mathrm{N}(0,1)$$
$$\theta_{t+1} = \gamma + \phi\theta_t + \eta_t, \quad \eta_t \sim \mathrm{N}(0, \sigma_\eta^2)$$

2. Stochastic volatility with non-Gaussian measurement density:

$$y_t = e^{\frac{\theta_t}{2}}\epsilon_t, \quad \epsilon_t \sim t(0,1,\nu)$$
$$\theta_{t+1} = \gamma \ + \phi\theta_t + \eta_t, \quad \eta_t \sim \mathrm{N}(0, \sigma_\eta^2)$$

3. Stochastic intensity model with Poisson measurement density:

$$p(y_t|\lambda_t) = \frac{\lambda_t^{y_t} e^{-\lambda_t}}{y_t!}, \quad \theta_t = \log\lambda_t$$
$$\theta_{t+1} = \gamma + \phi\theta_t + \eta_t, \quad \eta_t \sim \mathrm{N}(0, \sigma_\eta^2)$$

In order to estimate the two stochastic volatility models, Harvey et al. (1994) proposed a Quasi Maximum Likelihood method (QML) based on a Gaussian quasi-likelihood. Normality is assumed for the linearized model.

[3]Note that this is the same stochastic volatility model considered in the in section 3.2.

The latter is thus susceptible of treatment with the Kalman filter and the method can be viewed as providing approximate filtered and smoothed estimates. It is therefore interesting to compare the performance of the QML to that of the SDS.

The three models above are estimated through importance sampling (IS hereafter). As discussed by Durbin and Koopman (2000), IS methods are simple and effective for problems in time series analysis. They are based on independent samples rather than Markov chains, thus enabling to obtain accurate estimates of Monte-Carlo variances. Sandmann and Koopman (1998) devised an IS technique to evaluate the full likelihood function of stochastic volatility models. We estimate models 1 and 2 by employing the same importance sampling approach but use the recently developed "Numerically Accelerated Importance Sampling" (NAIS) technique of Koopman et al. (2015) to choose the parameters of the importance density. This method has been shown to provide several efficiency gains compared to other IS approaches. The stochastic intensity model can also be estimated through IS, as described e.g. by Durbin and Koopman (1997). Similarly to the previous cases, we choose the parameters of the importance density through the NAIS. More details on IS techniques for nonlinear non-Gaussian state-space models can be found in Durbin and Koopman (2012).

We set the measurement densities of the corresponding observation-driven models as indicated below:

1. For the two stochastic volatility models:

$$p(y_t|f_t) = \frac{\Gamma\left(\frac{\beta+1}{2}\right)}{\Gamma\left(\frac{\beta}{2}\right)\sqrt{\pi\beta e^{f_t}}}\left[1+\frac{1}{\beta}\left(\frac{y_t}{e^{\frac{f_t}{2}}}\right)^2\right]^{-\frac{\beta+1}{2}} \tag{74}$$

2. For the stochastic intensity model:

$$p(y_t|f_t) = e^{-e^{f_t}}\frac{e^{f_t y_t}}{y_t!} \tag{75}$$

The use of a t distribution for the first model is motivated by the fact that even Gaussian stochastic volatility models are able to generate a predictive density with fat-tails and over-dispersion (Carnero et al. 2004). Thus, in order for the observation-driven model to capture these dynamic features, we adopt a more flexible specification for the measurement density. This is in line with Koopman et al. (2016), who used more flexible densities for observation-driven counterparts of parameter-driven models. Note that the measurement density (74) is similar to the one leading to the Beta-t-EGARCH model of Harvey and Chakravarty (2008).

In the first case, setting $S_t = \mathcal{I}_{t|t-1}^{-1}$, the filtering recursion (29) reduces to:

$$f_{t+1} = \omega + A\frac{\beta+3}{\beta}\left[\frac{\beta+1}{\beta}\frac{\left(\frac{y_t}{e^{\frac{f_t}{2}}}\right)^2}{1+\frac{1}{\beta}\left(\frac{y_t}{e^{\frac{f_t}{2}}}\right)^2}-1\right]+Bf_t \tag{76}$$

while the smoothing recursions (30), (31) reduce to:

$$r_{t-1} = \frac{\beta+3}{\beta}\left[\frac{\beta+1}{\beta}\frac{\left(\frac{y_t}{e^{\frac{f_t}{2}}}\right)^2}{1+\frac{1}{\beta}\left(\frac{y_t}{e^{\frac{f_t}{2}}}\right)^2}-1\right]+(B-A)'r_t \tag{77}$$

$$\hat{f}_t = f_t + B^{-1}Ar_{t-1} \tag{78}$$

In the case of the measurement density in eq. (75), setting $S_t = \mathcal{I}_{t|t-1}^{-1}$, the filtering recursion (29) reduces to:

$$f_{t+1} = \omega + A(e^{-f_t}y_t - 1) + Bf_t \tag{79}$$

while the smoothing recursions (30), (31) reduce to:

$$r_{t-1} = e^{-f_t} y_t - 1 + (B - A)' r_t \tag{80}$$

$$\hat{f}_t = f_t + B^{-1} A r_{t-1} \tag{81}$$

Estimation of parameter-driven models is performed through the NAIS with $S = 200$ simulations. We also introduce control variables as described by Koopman et al. (2015). Smoothed estimates are computed with $G = 200$ simulations. Larger values of S and G do not lead to significant improvements of parameter-driven model estimates compared to SDS estimates. The simulation setting is the same as in the previous experiment, with 1000 simulated time series of 4000 observations. Figure 5 compares smoothed estimates obtained through both IS and the SDS for the three different models at hand. A simple visual inspection shows that the estimates provided by the two methods are very close.

In order to examine in more detail differences between SDS and IS estimates, Tables 5, 6, 7 show the results of Monte Carlo experiments for the three models. In the case of stochastic volatility models, we consider several scenarios by varying the autoregressive coefficient ϕ and the coefficient of variation CV. The latter is defined as in Sandmann and Koopman (1998), namely:

$$CV = \exp\left(\frac{\sigma_\eta^2}{1 - \phi^2}\right) - 1 \tag{82}$$

Note that CV is related to the variance of the signal innovation. The values of both ϕ, CV and of the remaining parameters are chosen to be close to those estimated on real financial time series, as discussed in Sandmann and Koopman (1998). For the stochastic intensity model, we consider scenarios characterized by different autoregressive coefficients ϕ and different values of the variance σ_η^2 of the signal. In the case of the two stochastic volatility models, the SDS largely outperforms the QML in all scenarios. The performance of the latter tends to worsen as CV decreases, according to the fact that the non-normality of the measurement equation becomes more relevant for low CV. Compared to IS, the relative MSE loss of the SDS is very small. In particular, it is always less than 2% in the Gaussian case, while it is always lower than 2.5% in the non-Gaussian case. Larger losses are observed for large values of CV, where σ_η is large and accounting for the non-normality of observations is less relevant.

In the case of the stochastic intensity model, we observe a similar behavior but the relative MSE loss is slightly larger for $\phi = 0.98$ and $\sigma_\eta^2 = 0.01$, where it is found to be around 5%. Overall, average MSE losses are less than 2.5% if one averages across all scenarios. This result is in agreement with what Koopman et al. (2016) found by comparing the prediction performance of score-driven models to that of correctly specified parameter-driven models.

Finally, it is interesting to look at computational times. Table 8 shows average computational times of IS relative to those of SDS. We report both the time required to estimate the model parameters and that required for smoothing. Estimation of static parameters parameters is much faster, as the likelihood can be computed in closed forms. In contrast, estimating correctly specified parameter-driven models with $S = 200$ simulations is on average 80 times slower. Note that decreasing S would lead to faster estimates at the expense of reducing efficiency. Smoothing with IS is on average 215 times slower compared to smoothing with the SDS. Note also that, while it is generally difficult to extend the NAIS and other IS methods to a setting with multiple time-varying parameters, the SDS maintains the same form in case f_t is a large vector of time-varying parameters. Thus, using the SDS allows to obtain smoothed estimates which are very close to those of correctly specified parameter-driven models but reducing considerably the computational burden.

5 Empirical illustration

It is interesting to investigate whether the results found in the simulation study in section 3 also hold on empirical data. In particular, we aim to provide a quantitative assessment of the improvement of smoothed SDS estimates over standard score-driven estimates in a problem of empirical relevance. Unlike the simulation study, it is generally difficult to perform such analysis empirically, given that time-varying parameters do not belong to the econometrician's information set and cannot be employed as a benchmark in the loss function. However, when dealing with conditional covariance estimates computed from *daily* log-returns, one can use realized covariance computed from *intraday* log-returns as an accurate proxy of the true latent covariance (Andersen and Bollerslev 1997). Loss functions can therefore be built as if covariances were observed. This empirical analysis will also show the advantages of the SDS methodology in an highly multivariate framework, where the use of simulation-based methods is computationally problematic or even unfeasible.

Our dataset consists of unbalanced 1-minute transaction data of Russel 3000 constituents over the period from 18-11-1999 to 27-09-2013. The total number of assets is 4166. The analysis is performed on the subsample comprising the last $T = 2000$ days, in order to avoid discontinuities due to changes on index composition. We consider trades from 9:30 to 16:00, leading to 390 timestamps per day. Assets having less than 10 trades per day are excluded in order to avoid poor and ill-conditioned realized covariance estimates. As a final outcome of our filtering procedure, we obtain $N = 1682$ assets.

In order to estimate conditional covariances from daily log-returns, we use the t-GAS model of Creal et al. (2011) described in section (3). Compared to standard conditional covariance models, the main advantage of the t-GAS is that it updates covariances by taking into account the full shape of the Student t observation density and thus providing robustness against outliers (see discussions on Creal et al. 2011). We implement the parameterization based on hyperspherical coordinates, as it generally leads to better estimates. The number of time-varying parameters grows as p^2, where p is the number of assets.

Among the universe of $N = 1682$ assets, we select random groups of $p = 5, 10, 20$ assets. In particular, for each cross-section dimension p, we randomly choose four groups. The analysis is thus performed on 12 different groups of assets. As done in the simulation study, we use the RMSE and Qlike as loss functions. The benchmark used in the loss function is the realized covariance estimator of Barndorff-Nielsen and Shephard (2004) computed at the 5-minutes sampling frequency. The use of other estimators does not alter the outcome of the experiment. For each group of assets, the t-GAS is estimated on the time-series of $T = 2000$ open-to-close log-returns. We thus compute (i) the predictive filter f_t, (ii) the update filter $f_{t|t}$ and (iii) the smoother $\hat{f}_t$. The statistical significance of loss differences is tested through the model confidence set of Hansen et al. (2011) at the 90% confidence level.

Tables 9, 10, 11 show the results of the analysis, for $p = 5$, 10, 20, respectively. We first note that covariance estimates constructed through the predictive filter feature larger RMSE and Qlike and are always excluded from the model confidence set. They are in fact less informative, since only past log-returns are used when reconstructing time-varying parameters. Covariance estimates built through $f_{t|t}$ and $\hat{f}_t$ are both included in the model confidence set constructed through the RMSE. If the Qlike is used, only smoothed estimates are included. The latter provide a better reconstruction of realized covariance, suggesting that future observations contain relevant information on today's covariance. Note also that relative gains are similar across different dimensions, meaning that the SDS is not affected by the proliferation of time-varying parameters when the number of assets increases.

The above results suggest that, when extracting latent covariance, the smoothing provided by the t-GAS model is

effective in aggregating all available information. Compared to standard score-driven filtered estimates, the update filter $f_{t|t}$ and the smoother $\hat{f}_t$ can thus be regarded as providing a more accurate estimate of latent covariance. As seen in the simulation study, this is true for a large class of dynamic models. We thus use these results to solicit the use of the SDS in place of standard filtered estimates in signal reconstruction analysis.

6 Conclusions

In this paper we have introduced the SDS, a general smoothing methodology for observation-driven models. The SDS is based on viewing observation-driven models, and in particular score-driven models, as filters rather than data generating processes. As such, actual and future observations can be used to refine estimates of time-varying parameters. The SDS has similar form to Kalman backward smoothing recursions for linear Gaussian state-space models but employs the score of the observation density at hand. Thus, it is particularly advantageous from a computational point of view and can be easily applied to any observation density, with a potentially large vector of time-varying parameters. Similarly, we have introduced an update filter, the SDU, which updates standard score-driven filtered estimates once new observations become available.

We have examined several examples of SDS and SDU estimates corresponding to popular univariate observation-driven models (GARCH, MEM), an AR(1) model with a time-varying autoregressive coefficient and two multivariate covariance models (t-GAS, Wishart-GARCH). The SDS updates SDF estimates based on all available observations and thus leads to a more efficient reconstruction of time-varying parameters. In particular, we have found relative gains up to 63% in mean square errors when comparing filtered and smoothed estimates of time-varying parameters.

Our general framework allows to straightforwardly build in-sample and out-of-sample confidence bands accounting for filtering uncertainty in score-driven models. We have examined in detail the construction of robust confidence bands in the GARCH model and showed in a simulation study that they are able to capture filtering uncertainty.

Extensive Monte Carlo simulations of nonlinear non-Gaussian state-space models showed that SDS estimates are very similar to estimates of correctly specified parameter-driven models. Indeed, losses are always smaller on, average, than 2.5% in mean square errors. At the same time, the SDS is more appealing from a computational point of view, being hundreds of times faster than simulation-based techniques.

In our empirical study, we used realized covariance computed from high-frequency data as a proxy of the latent covariance and showed that reconstructing covariances using all available log-returns leads to better estimates compared to only using past log-returns, as typically done in observation-driven models. This analysis strengthens the results found in the simulation study and suggests that one can benefit from employing the SDS in place of standard filtered estimates when reconstructing time-varying parameters from the observed data.

References

Andersen, T., Bollerslev, T., 1997. Intraday periodicity and volatility persistence in financial markets. Journal of Empirical Finance 4 (2-3), 115–158.

Barndorff-Nielsen, O. E., Shephard, N., 2004. Econometric analysis of realized covariation: High frequency based covariance, regression, and correlation in financial economics. Econometrica 72 (3), 885–925.

Blasques, F., Koopman, S. J., Lucas, A., Aug. 2014. Optimal formulations for nonlinear autoregressive processes. Working paper Available at https://ssrn.com/abstract=2478575.

Blasques, F., Koopman, S. J., Lucas, A., 2015. Information-theoretic optimality of observation-driven time series models for continuous responses. Biometrika 102 (2), 325.

Blasques, F., Koopman, S. J., Łasak, K., Lucas, A., 2016. In-sample confidence bands and out-of-sample forecast bands for time-varying parameters in observation-driven models. International Journal of Forecasting 32 (3), 875 – 887.

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

Buccheri, G., Bormetti, G., Corsi, F., Lillo, F., 2018. Assessing filtering uncertainty in observation-driven models. Mimeo.

Buccheri, G., Corsi, F., 2017. Hark the shark: Realized volatility modelling with measurement errors and nonlinear dependencies. Working Paper, Available at https://ssrn.com/abstract=3089929.

Caivano, M., Harvey, A., Luati, A., Mar 2016. Robust time series models with trend and seasonal components. SERIEs 7 (1), 99–120.

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

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

Creal, D., Koopman, S. J., Lucas, A., 2013. Generalized autoregressive score models with applications. Journal of Applied Econometrics 28 (5), 777–795.

Durbin, J., Koopman, S., 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 (Statistical Methodology) 62 (1), 3–56.

Durbin, J., Koopman, S., 2012. Time Series Analysis by State Space Methods: Second Edition. Oxford Statistical Science Series. OUP Oxford.

Durbin, J., Koopman, S. J., 1997. Monte carlo maximum likelihood estimation for non-gaussian state space models. Biometrika 84 (3), 669–684.

Engle, R., 2002a. Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. Journal of Business & Economic Statistics 20 (3), 339–50.

Engle, R., 2002b. New frontiers for arch models. Journal of Applied Econometrics 17 (5), 425–446.

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

Gorgi, P., Hansen, P. R., Janus, P., Koopman, S. J., 2018. Realized wishart-garch: A score-driven multi-asset volatility model*. Journal of Financial Econometrics, nby007.

Hansen, P. R., Lunde, A., Nason, J. M., 2011. The model confidence set. Econometrica 79 (2), 453–497.

Harvey, A., 1991. Forecasting, Structural Time Series Models and the Kalman Filter. Cambridge University Press.

Harvey, A., Chakravarty, T., Sep. 2008. Beta-t-(E)GARCH. Cambridge Working Papers in Economics 0840, Faculty of Economics, University of Cambridge.

Harvey, A., Luati, A., 2014. Filtering with heavy tails. Journal of the American Statistical Association 109 (507), 1112–1122.

Harvey, A., Ruiz, E., Shephard, N., 1994. Multivariate stochastic variance models. The Review of Economic Studies 61 (2), 247–264.

Harvey, A. C., 2013. Dynamic Models for Volatility and Heavy Tails: With Applications to Financial and Economic Time Series. Econometric Society Monographs. Cambridge University Press.

Koopman, S. J., Lucas, A., Scharth, M., 2015. Numerically accelerated importance sampling for nonlinear non-gaussian state-space models. Journal of Business & Economic Statistics 33 (1), 114–127.

Koopman, S. J., Lucas, A., Scharth, M., March 2016. Predicting Time-Varying Parameters with Parameter-Driven and Observation-Driven Models. The Review of Economics and Statistics 98 (1), 97–110.

Monache, D. D., Petrella, I., 2017. Adaptive models and heavy tails with an application to inflation forecasting. International Journal of Forecasting 33 (2), 482 – 501.

Nelson, D. B., 1992. Filtering and forecasting with misspecified arch models i: Getting the right variance with the wrong model. Journal of Econometrics 52 (1), 61 – 90.

Nelson, D. B., 1996. Asymptotically optimal smoothing with arch models. Econometrica 64 (3), 561–573.

Nelson, D. B., Foster, D. P., 1994. Asymptotic filtering theory for univariate arch models. Econometrica 62 (1), 1–41.

Nelson, D. B., Foster, D. P., 1995. Filtering and forecasting with misspecified arch models ii: Making the right forecast with the wrong model. Journal of Econometrics 67 (2), 303 – 335.

Oh, D. H., Patton, A. J., 2017. Time-varying systemic risk: Evidence from a dynamic copula model of cds spreads. Journal of Business & Economic Statistics 0 (0), 1–15.

Opschoor, A., Janus, P., Lucas, A., Dijk, D. V., 2017. New heavy models for fat-tailed realized covariances and returns. Journal of Business & Economic Statistics 0 (0), 1–15.

Patton, A. J., 2011. Volatility forecast comparison using imperfect volatility proxies. Journal of Econometrics 160 (1), 246 – 256, realized Volatility.

Sandmann, G., Koopman, S. J., 1998. Estimation of stochastic volatility models via monte carlo maximum likelihood. Journal of Econometrics 87 (2), 271–301.

Zamojski, M., 2016. Filtering with confidence: In-sample confidence bands for garch filters. Working paper.

	Slow sine	Fast sine	Ramp	Step	Model
			MSE		
			GARCH		
SDF	1.0000	1.0000	1.0000	1.0000	1.0000
SDU	0.9958	0.9879	0.9888	0.9889	0.9912
SDS	0.4231	0.5398	0.5494	0.5029	0.5289
			MEM		
SDF	1.0000	1.0000	1.0000	1.0000	1.0000
SDU	0.9928	0.9718	0.9769	0.9794	0.9791
SDS	0.3874	0.4300	0.5239	0.5140	0.5132
			AR(1)		
SDF	1.0000	1.0000	1.0000	1.0000	1.0000
SDU	0.9954	0.9849	0.9935	0.9966	0.9991
SDS	0.5657	0.5202	0.6257	0.6303	0.6441
			MAE		
			GARCH		
SDF	1.0000	1.0000	1.0000	1.0000	1.0000
SDU	0.9974	0.9902	0.9975	0.9966	0.9948
SDS	0.6279	0.7017	0.7479	0.7314	0.7351
			MEM		
SDF	1.0000	1.0000	1.0000	1.0000	1.0000
SDU	0.9957	0.9813	0.9968	0.9962	0.9885
SDS	0.5990	0.6318	0.7370	0.7497	0.7178
			AR(1)		
SDF	1.0000	1.0000	1.0000	1.0000	1.0000
SDU	0.9991	0.9891	0.9978	0.9988	0.9995
SDS	0.7074	0.7010	0.7940	0.7934	0.7705

Table 1: Average MSE and MAE of filtered SDU and smoothed SDS estimates relative to standard filtered SDF estimates.

p	5	10	20
		RMSE	
		t-GAS	
SDF	1.0000	1.0000	1.0000
SDU	0.9723	0.9654	0.9666
SDS	0.8165	0.8202	0.8596
		Wishart-GARCH	
SDF	1.0000	1.0000	1.0000
SDU	0.8733	0.8821	0.9330
SDS	0.8164	0.8070	0.8742
		Qlike	
		t-GAS	
SDF	1.0000	1.0000	1.0000
SDU	0.9989	0.9965	0.9922
SDS	0.9821	0.9888	0.9898
		Wishart-GARCH	
SDF	1.0000	1.0000	1.0000
SDU	0.9858	0.9909	0.9901
SDS	0.9658	0.9700	0.9712

Table 2: Average root mean square error (RMSE) and quasi-likelihood (Qlike) of SDU and SDS estimates relative to SDF estimates of t-GAS and Wishart-GARCH model

Nominal c.l.	90%	95%	99%
	In-sample		
	Normal		
SDF	0.8356	0.9026	0.9703
SDU	0.8346	0.9022	0.9701
SDS	0.8303	0.8980	0.9680
	Robust		
SDF	0.8968	0.9523	0.9930
SDU	0.8963	0.9520	0.9929
SDS	0.8921	0.9489	0.9918
	Out-of-sample		
	Normal		
SDF	0.8304	0.8975	0.9669
SDU	0.8296	0.8971	0.9667
SDS	0.8268	0.8951	0.9661
	Robust		
SDF	0.8917	0.9483	0.9913
SDU	0.8912	0.9481	0.9912
SDS	0.8890	0.9467	0.9909

Table 3: Average coverage of in-sample and out-of-sample confidence bands obtained by assuming a normal conditional density and the fat-tailed density described in section 3.2. We report results for three different nominal confidence levels, namely 90%, 95%, 99%.

.

	SDF − KF(p)			SDU − KF(u)			SDS − KS		
δ	0.1	1	10	0.1	1	10	0.1	1	10
	$\nu = 3$								
MSE	0.8610	0.9522	0.9991	0.8074	0.8357	1.0223	0.8093	0.8876	0.9618
MAE	0.9389	0.9859	1.0036	0.9110	0.9439	1.0150	0.9128	0.9634	1.0169
	$\nu = 5$								
MSE	0.9552	0.9912	1.0032	0.9407	0.9681	1.0032	0.9376	0.9880	1.0058
MAE	0.9792	0.9973	0.9999	0.9720	0.9886	1.0032	0.9698	0.9949	1.0112
	$\nu = 8$								
MSE	0.9877	0.9981	1.0029	0.9895	0.9928	1.0107	0.9844	0.9954	1.0117
MAE	0.9939	0.9992	1.0039	0.9927	0.9973	1.0070	0.9917	0.9982	1.0136

Table 4: Average MSE and MAE of SDF, SDU, SDS relative to Kalman filtered (predicted and updated) and smoothed estimates of AR(1) model plus non-Gaussian noise.

CV	0.1	1	5	10	0.1	1	5	10
	MSE				MAE			
	$\phi = 0.98$							
IS	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
SDS	0.9988	1.0050	1.0001	1.0162	1.0004	1.0043	1.0017	1.0097
QML	1.4153	1.3880	1.3333	1.3138	1.1797	1.1739	1.1564	1.1475
	$\phi = 0.95$							
IS	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
SDS	1.0057	0.9983	0.9988	1.0059	1.0034	1.0023	1.0024	1.0059
QML	1.3131	1.3737	1.3246	1.3168	1.1450	1.1758	1.1567	1.1524
	$\phi = 0.90$							
IS	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
SDS	1.0076	0.9956	0.9974	1.0086	1.0044	1.0010	1.0033	1.0093
QML	1.2371	1.3157	1.2893	1.2750	1.1109	1.1508	1.1422	1.1370

Table 5: Average MSE and MAE of IS, SDS and QML smoothed estimates normalized by IS loss in case of stochastic volatility model with Gaussian measurement density.

CV	0.1	1	5	10	0.1	1	5	10
	MSE				MAE			
	$\phi = 0.98$, $\nu = 3$							
IS	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
SDS	1.0026	0.9950	1.0140	1.0169	1.0015	0.9997	1.0077	1.0098
QML	1.3962	1.2553	1.2125	1.1998	1.1735	1.1184	1.1013	1.0939
	$\phi = 0.95$, $\nu = 3$							
IS	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
SDS	1.0014	1.0049	1.0121	1.0200	1.0008	1.0031	1.0064	1.0105
QML	1.3058	1.2639	1.2447	1.2246	1.1354	1.1230	1.1158	1.1056
	$\phi = 0.90$, $\nu = 3$							
IS	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
SDS	1.0020	1.0033	1.0149	1.0221	1.0016	1.0023	1.0081	1.0117
QML	1.2306	1.2325	1.2262	1.2200	1.1026	1.1075	1.1062	1.1034

Table 6: Average MSE and MAE of IS, SDS and QML smoothed estimates normalized by IS loss in case of stochastic volatility model with non-Gaussian measurement density.

$\sigma_\eta^2 \times 100$	0.1	0.5	1	0.1	0.5	1
	MSE			MAE		
	$\phi = 0.98$					
IS	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
SDS	1.0149	1.0281	1.0521	1.0067	1.0132	1.0244
	$\phi = 0.95$					
IS	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
SDS	1.0203	1.0176	1.0254	1.0097	1.0083	1.0120
	$\phi = 0.90$					
IS	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
SDS	1.0310	1.0160	1.0205	1.0142	1.0079	1.0099

Table 7: Average MSE and MAE of IS and SDS smoothed estimates normalized by IS loss in case of stochastic intensity model with Poisson measurement density.

	SV Gaussian	SV Fat tail	Intensity
Estimation	45.35	107.90	82.33
Smoothing	179.74	229.79	236.34

Table 8: Average computational times of IS relative to SDS.

	Portfolio 1	Portfolio 2	Portfolio 3	Portfolio 4
	RMSE$\times 10^7$			
SDF	0.7967*	0.2904	0.2181	0.9222
	1.0000	1.0000	1.0000	1.0000
SDU	0.7967*	0.2806*	0.1997	0.8976*
	0.9733	0.9665	0.9157	0.9732
SDS	0.7445*	0.2791*	0.1732*	0.8101*
	0.9345	0.9611	0.7941	0.8784
	Qlike			
SDF	-34.1291	-37.3352	-32.4399	-37.7819
	1.0000	1.0000	1.0000	1.0000
SDU	-34.2181	-37.4292	-32.6133	-37.9202*
	0.9974	0.9975	0.9947	0.9964
SDS	-34.3482*	-37.5424*	-32.6987*	-37.9751*
	0.9936	0.9945	0.9921	0.9949

Table 9: Absolute and relative RMSE and Qlike of SDF, SDU, SDS estimates of the t-GAS model for the four randomly selected portfolios with $n = 5$. The asterisk implies that the estimator is included in the model confidence set.

	Portfolio 1	Portfolio 2	Portfolio 3	Portfolio 4
	RMSE$\times 10^5$			
SDF	0.1271	0.2080	0.3105	0.1398
	1.0000	1.0000	1.0000	1.0000
SDU	0.1229*	0.2021*	0.3008	0.1356*
	0.9662	0.9716	0.9688	0.9704
SDS	0.1189*	0.1946*	0.2868*	0.1345*
	0.9354	0.9355	0.9237	0.9623
	Qlike			
SDF	-65.5001	-68.3171	-70.4809	-70.9020
	1.0000	1.0000	1.0000	1.0000
SDU	-65.7213	-68.4716	-70.6554	-71.0882
	0.9974	0.9977	0.9975	0.9974
SDS	-65.8784*	-68.6694*	-70.9327*	-71.2838*
	0.9936	0.9949	0.9936	0.9946

Table 10: Absolute and relative RMSE and Qlike of SDF, SDU, SDS estimates of the t-GAS model for the four randomly selected portfolios with $n = 10$. The asterisk implies that the estimator is included in the model confidence set.

	Portfolio 1	Portfolio 2	Portfolio 3	Portfolio 4
	RMSE$\times 10^5$			
SDF	0.7061	0.3705	0.2146	0.3255
	1.0000	1.0000	1.0000	1.0000
SDU	0.6950*	0.3634*	0.2099	0.3170*
	0.9843	0.9808	0.9782	0.9739
SDS	0.6709*	0.3565*	0.1998*	0.3046*
	0.9501	0.9622	0.9311	0.9358
	Qlike			
SDF	-134.7951	-128.6277	-140.1625	-137.9562
	1.0000	1.0000	1.0000	1.0000
SDU	-135.1702	-129.0375	-140.5694	-138.2863
	0.9972	0.9968	0.9971	0.9976
SDS	-135.7584*	-129.3180*	-141.2055*	-138.8856*
	0.9929	0.9947	0.9926	0.9933

Table 11: Absolute and relative RMSE and Qlike of SDF, SDU, SDS estimates of the t-GAS model for the four randomly selected portfolios with $n = 20$. The asterisk implies that the estimator is included in the model confidence set.

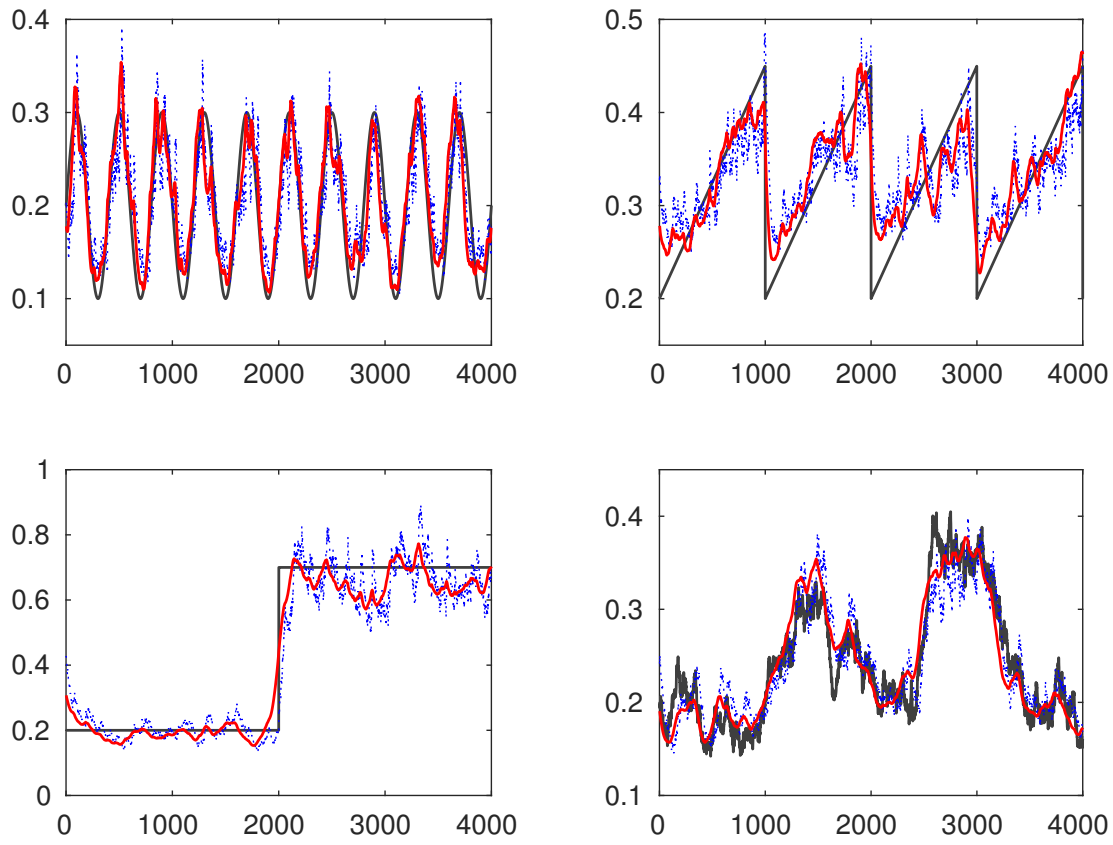

Figure 1: Comparison among simulated (black lines), filtered (blue dotted lines) and smoothed (red lines) variance σ_t^2 of GARCH(1,1) model.

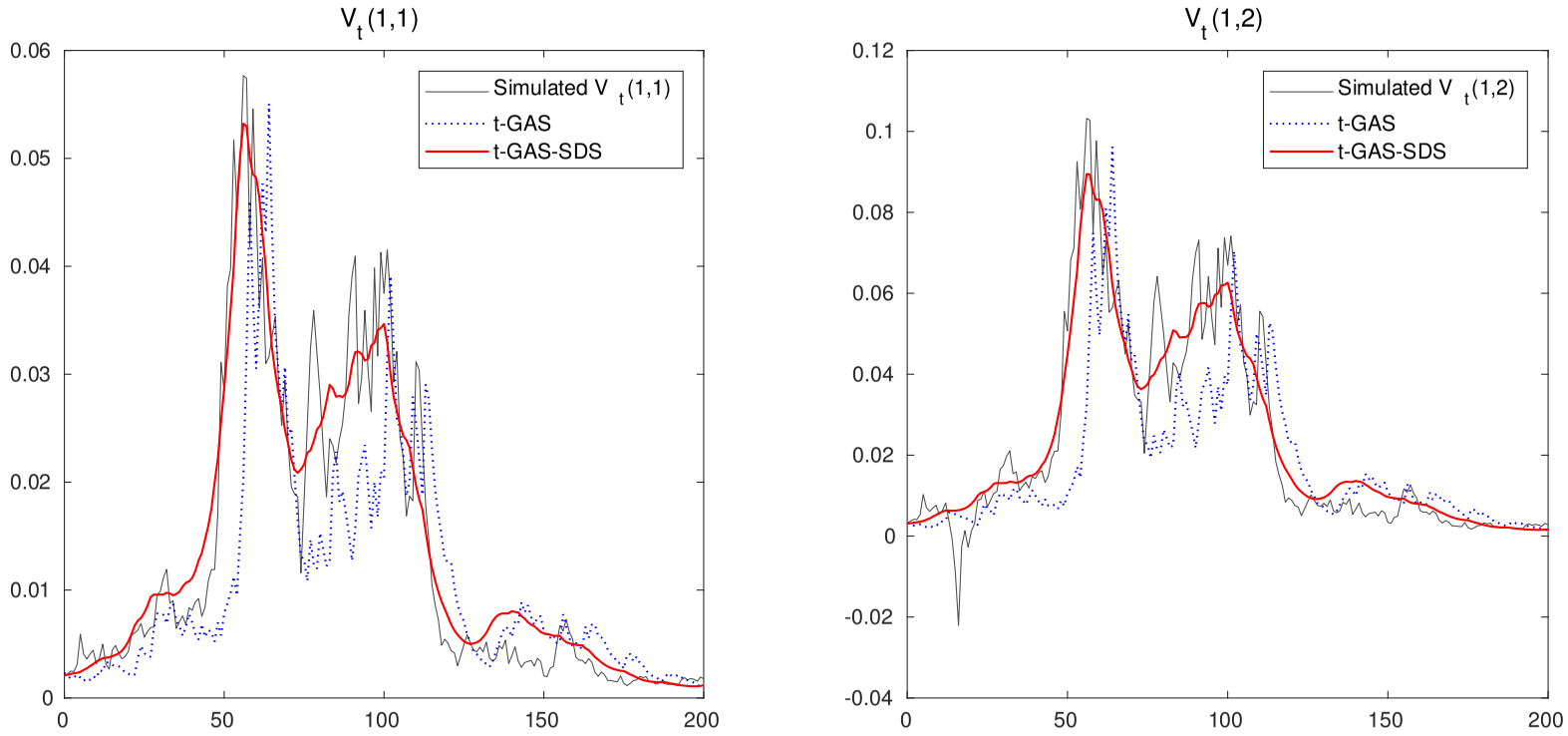

Figure 2: Comparison among simulated true covariances V_t (black lines), filtered (blue dotted lines) and smoothed (red lines) (co)variances of t-GAS model in the case $k = 5$. We show the variance corresponding to the first asset on the left and the covariance between the first and the second asset on the right.

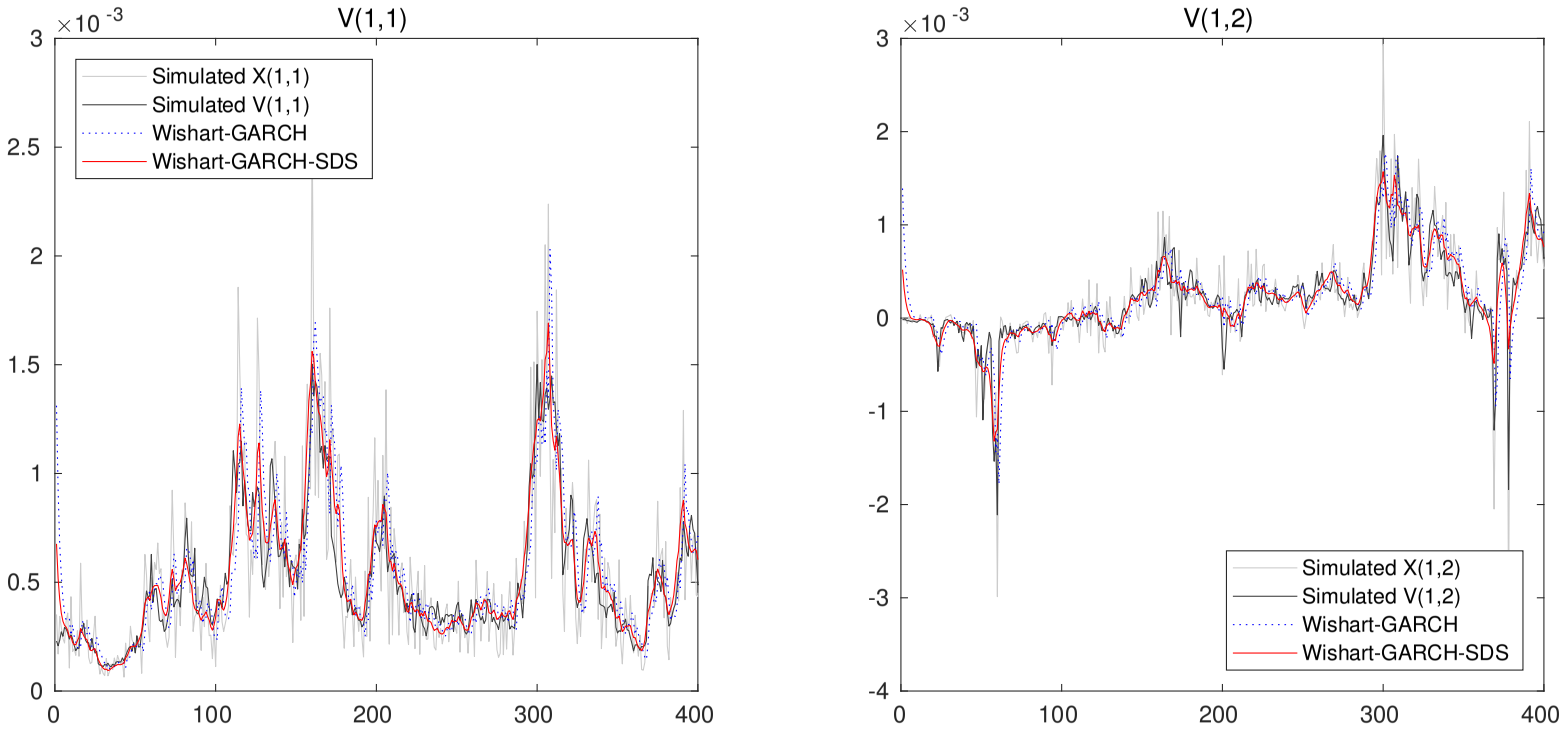

Figure 3: Comparison among simulated observations of X_t (grey lines), simulated true covariances V_t (black lines), filtered (blue dotted lines) and smoothed (red lines) (co)variances of realized Wishart-GARCH model in the case $k = 5$. We show the variance corresponding to the first asset on the left and the covariance between the first and the second asset on the right.

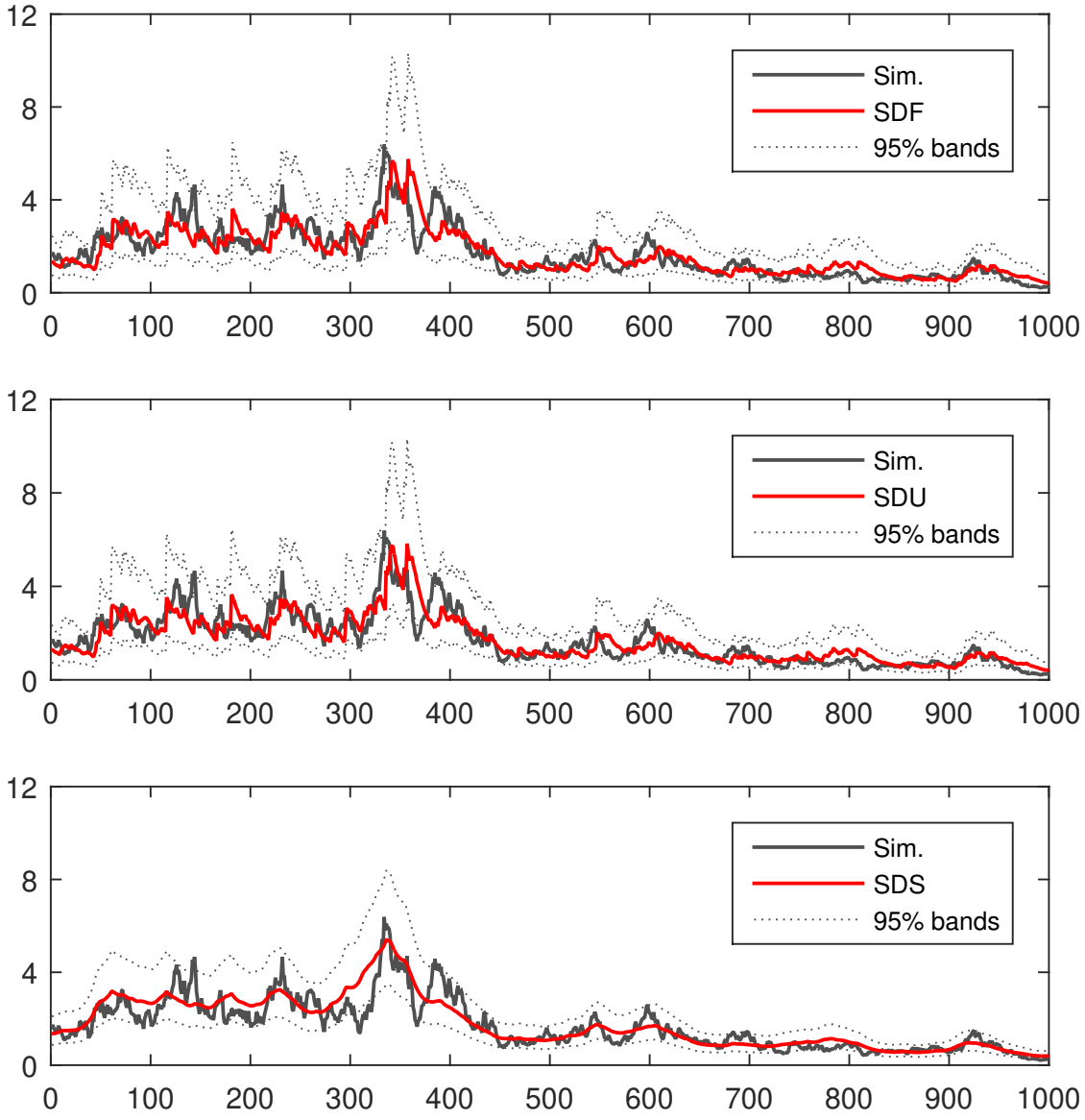

Figure 4: Simulated volatility and SDF, SDU, SDS estimates together with 95% in-sample confidence bands computed through the technique described in section 3.2.

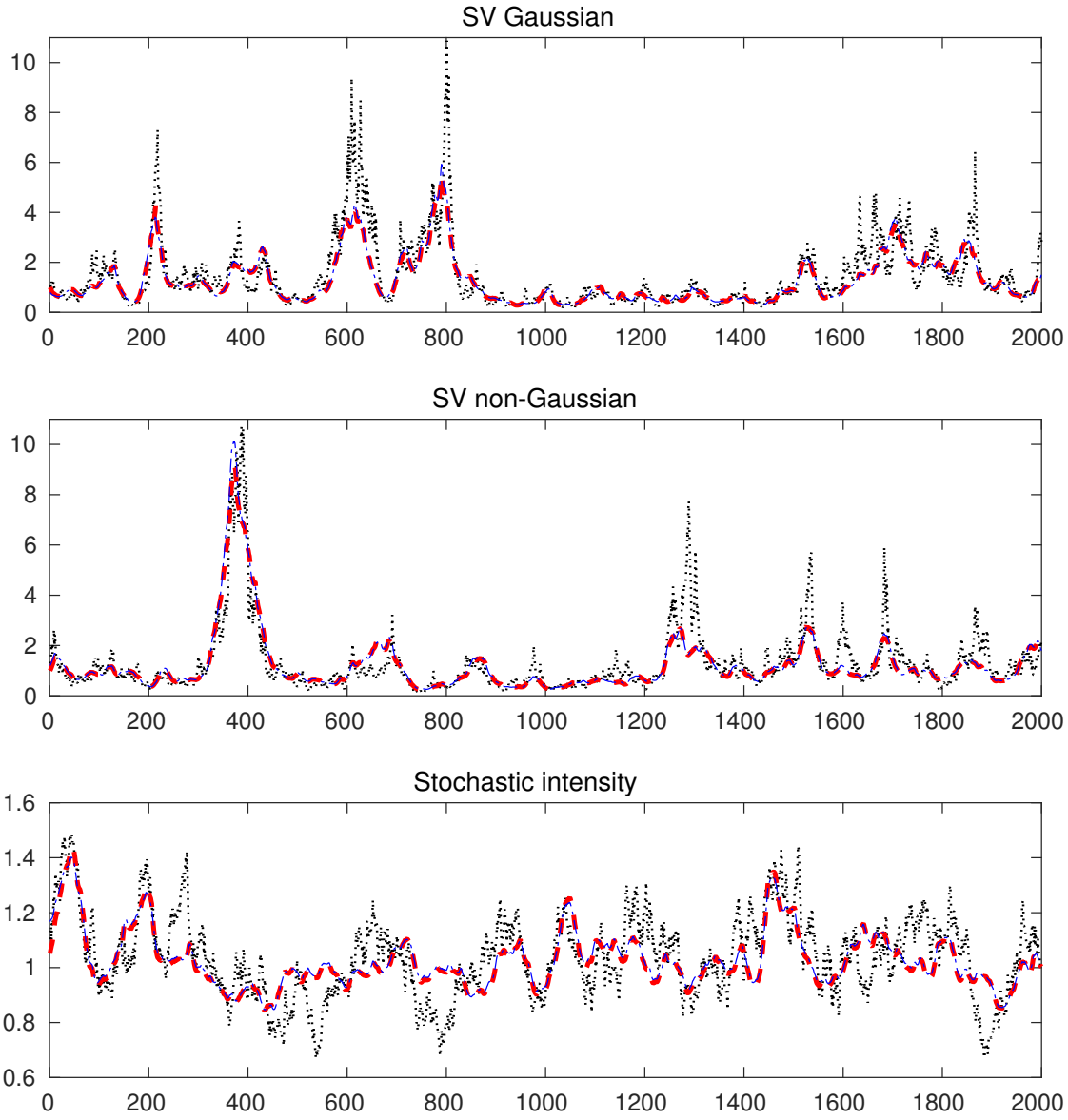

Figure 5: Comparison among simulated unobserved components (black dotted lines), IS smoothed estimates (red dashed lines) and SDS smoothed estimates (blue dotted and dashed).

Appendix

A Proof of Proposition 1

We first compute the score:

$$\left[\frac{\partial \log p(y_t|\mathcal{F}_{t-1})}{\partial a_t'}\right]' = \left[\frac{\partial \log p(y_t|\mathcal{F}_{t-1})}{\partial v_t'}\frac{\partial v_t}{\partial a_t'}\right]' = [v_t' F_t^{-1} Z]' = Z' F_t^{-1} v_t \tag{A.1}$$

The information matrix is then computed as:

$$\mathcal{I}_{t|t-1} = \mathrm{E}_{t-1}[\nabla_t \nabla_t'] = Z' F_t^{-1} Z \tag{A.2}$$

Thus, we can re-write recursions for $a_{t|t}$ and a_{t+1} as:

$$a_{t|t} = a_t + P_t \nabla_t \tag{A.3}$$

$$a_{t+1} = c + T a_t + T P_t \nabla_t \tag{A.4}$$

and the backward recursion for $\hat{\alpha}_t$ as

$$r_{t-1} = \nabla_t + L_t' r_t \tag{A.5}$$

$$\hat{\alpha}_t = a_t + P_t r_{t-1} \tag{A.6}$$

where $L_t = T - T P_t \mathcal{I}_{t|t-1}$. In the steady state, P_t converges to the solution $\bar{P}$ of the matrix Riccati equation (20) (Harvey 1991, Durbin and Koopman 2012). By defining $R = T\bar{P}$, $\bar{F} = Z'\bar{P}Z + H$ and $\mathcal{I} = Z'\bar{F}^{-1}Z$, we can re-write the Kalman filtering and smoothing recursions for the mean in the steady state as:

$$a_{t|t} = a_t + T^{-1} R \nabla_t \tag{A.7}$$

$$a_{t+1} = c + T a_t + R \nabla_t \tag{A.8}$$

and

$$r_{t-1} = \nabla_t + (T - R\mathcal{I})' r_t \tag{A.9}$$

$$\hat{\alpha}_t = a_t + T^{-1} R r_{t-1} \tag{A.10}$$

Q.E.D.

B DGP for time-varying covariances

We generate daily time series of realized covariance matrices and log-returns through the following method: (a) using daily log-returns of $k = 5, 10, 15$ randomly selected NYSE stocks, we compute the sample covariance matrix C over the period from 03-01-2006 to 31-12-2014, (b) consider the decomposition $C = UDU'$, where $D = \mathrm{diag}[\lambda_1, \ldots, \lambda_k]$ is a diagonal matrix of decreasing eigenvalues of C and U is the corresponding matrix of normalized eigenvectors, (c) let the first r eigenvalues of C evolve through an AR(1) process:

$$\log \lambda_{t,j} = c_j + \phi_j \log \lambda_{t-1,j} + \beta_{t,j}, \quad \beta_{t,j} \sim \mathrm{N}(0, q_j) \tag{B.1}$$

for $t = 2, \ldots, T$ and $j = 1, \ldots, r$. We therefore consider the sequence of matrices $V_t = U D_t U'$, where:

$$D_t = \mathrm{diag}[\lambda_{t,1}, \ldots, \lambda_{t,r}, \lambda_{r+1}, \ldots, \lambda_k] \tag{B.2}$$

for $t = 1, \dots T$. We set $r = 3$ and choose the parameters of the AR(1) model as: $\phi_1 = 0.99$, $q_1 = 0.05$, $\phi_2 = 0.95$, $q_2 = 0.05$, $\phi_1 = 0.90$, $q_1 = 0.5$ while constants c_j are chosen by setting the unconditional mean equal to $\log \lambda_j$, for $j = 1, 2, 3$. We then use the simulated covariance matrices V_t to generate observations of daily log-returns r_t through the t-GAS density in eq. (59) and of both daily log-returns r_r and realized covariance matrices X_t through the Wishart-GARCH density in eq. (60), (61). In the former case we set $\nu = 5$ while in the latter we set $\nu = p + 1$.

As loss measures, we use the root mean square error (RMSE) and the quasi-likelihood (Qlike). The former is defined as:

$$L_t^{\text{RMSE}} = \sqrt{\text{Tr}[(\widehat{V}_t - V_t)'(\hat{V}_t - V_t)]} \tag{B.3}$$

while the Qlike loss is given by:

$$L_t^{\text{Qlike}} = \log |\widehat{V}_t| + \text{Tr}[\widehat{V}_t^{-1} V_t] \tag{B.4}$$

where $\widehat{V}_t$ generically denotes the estimate obtained through the SDF or the SDS. See Patton (2011) for a more detailed discussion on the use of these loss measures in the field of (co)variance estimation.