A general framework for score-driven filtering and smoothing

Giuseppe Buccheri

Scuola Normale Superiore

Giacomo Bormetti

University of Bologna

Fulvio Corsi

University of Pisa, City University of London

Fabrizio Lillo

University of Bologna

Econometrics in the Arena

12-13 September 2019, Verona

Key facts

Following Cox (1981), time-varying parameter models can be divided in two classes:

1. **Parameter-driven models:** Parameters evolve based on idiosyncratic innovations (e.g. Local-Level, Stochastic Volatility (SV), Stochastic Intensity)
2. **Observation-driven models:** Parameters evolve based on past observations (e.g. GARCH, DCC, Score-Driven models)

There is a trade-off between:

1. **Flexibility**
   - Here parameter-driven models are superior
2. **Estimation complexity** and computational speed
   - Here observation-driven models are superior

Why a difference in flexibility?

- Parameter-driven: $\text{Var}[f_{t+1}] > 0$ and $\textbf{Var}[f_{t+1}|\mathcal{F}_t] > 0$
- Observation-driven: $\text{Var}[f_{t+1}] > 0$ but $\textbf{Var}[f_{t+1}|\mathcal{F}_t] = 0$

Score-Driven (SD) models

- SD models (Creal et al. 2013, Harvey 2013) general class of obs-driven models
- Provide general framework to update parameters based on past observations

Assume obs $y_t \in \mathbb{R}^p$ are generated by a conditional obs-density $p(y_t|f_t, \Theta)$:

$$y_t|\mathcal{F}_{t-1} \sim p(y_t|f_t, \Theta)$$

where $f_t \in \mathbb{R}^k$ is a vector of t.v.p. for which the following updating rule is proposed:

$$f_{t+1} = \omega + Bf_t + AS_t\nabla_t$$

∇_t is the score of the conditional likelihood $p(y_t|f_t, \Theta)$:

$$\nabla_t \equiv \frac{\partial \log p(y_t|f_t)}{\partial f_t}, \qquad S_t = g(\mathcal{H}_{t|t-1}), \qquad \mathcal{H}_{t|t-1} \equiv \mathbb{E}[\nabla_t \nabla_t' | \mathcal{F}_{t-1}]$$

Main advantages:

- Exploit the full shape of the observation density
- The likelihood can be written in closed-form
- Encompass many well-known models (GARCH, MEM, EGARCH, ACD, etc.)
- Information theoretic optimality (Blasques et al. 2015)

Observation-driven models as misspecified filters

Consider a standard GARCH(1,1) model:

$$y_t|\mathcal{F}_{t-1} \sim \mathrm{N}(0, \sigma_t^2)$$
$$\sigma_{t+1}^2 = c + a y_t^2 + b\sigma_t^2$$

There are two possible interpretations for the recursion $\sigma_{t+1}^2 = g(y_t, \sigma_t^2)$:

1. It is the true **D.G.P.** of volatility
2. It is a **predictive filter**, since σ_{t+1}^2 is $\mathcal{F}_t$-measurable

$$\sigma_{t+1}^2 \approx \mathrm{E}[\zeta_{t+1}^2|\mathcal{F}_t]$$

   where ζ_{t+1} is the volatility of the true, parameter-driven DGP (e.g. a SV model)

Assumption 1 is more common in the financial econometrics literature, while 2 is closer to the filtering literature.

Motivations and Objectives

Observation-driven models as **D.G.P.**

Parameters are completely revealed by past observations ($\underline{\text{Var}[f_{t+1}|\mathcal{F}_t] = 0}$)

$\downarrow$

No room for smoothing

Observation-driven models as **predictive filters**

Conditionally on the past, parameters are stochastic ($\underline{\text{Var}[f_{t+1}|\mathcal{F}_t] > 0}$)

$\downarrow$

Smoothing is useful

Despite the large amount of observation-driven models, little attention has been paid to the problem of smoothing with *misspecified* observation-driven models (Harvey 2013).

Goal of this paper:

- Filling the gap by proposing a general methodology to update and smooth filtered estimates of **Score-Driven models**

Filtering and smoothing in linear Gaussian models

$$y_t = Z\alpha_t + \epsilon_t, \quad \epsilon_t \sim \mathrm{N}(0, H)$$
$$\alpha_{t+1} = c + B\alpha_t + \eta_t, \quad \eta_t \sim \mathrm{N}(0, Q)$$

$y_t \in \mathbb{R}^p$, $\alpha_t \in \mathbb{R}^k$.

Kalman predictive filter: $a_{t+1} = \mathrm{E}[\alpha_{t+1}|\mathcal{F}_t],\ P_{t+1} = \mathrm{Var}[\alpha_{t+1}|\mathcal{F}_t]$

Kalman update filter: $a_{t|t} = \mathrm{E}[\alpha_t|\mathcal{F}_t],\ P_{t|t} = \mathrm{Var}[\alpha_t|\mathcal{F}_t]$

$$v_t = y_t - Za_t, \qquad F_t = ZP_tZ' + H$$
$$a_{t+1} = c + Ba_t + K_tv_t, \qquad P_{t+1} = BP_t(B - K_tZ)' + Q$$
$$a_{t|t} = a_t + P_tZ'F_t^{-1}v_t, \qquad P_{t|t} = P_t - P_tZ'F_t^{-1}ZP_t$$

with $K_t = BP_tZ'F_t^{-1}$ and $t = 1, \ldots, n$

Kalman backward smoother: $\hat{\alpha}_t = \mathrm{E}[\alpha_t|\mathcal{F}_n],\ \hat{P}_t = \mathrm{Var}[\alpha_t|\mathcal{F}_n],\ t \leq n.$

$$r_{t-1} = Z'F_t^{-1}v_t + L_t'r_t, \qquad N_{t-1} = Z'F_t^{-1}Z + L_t'N_tL_t$$
$$\hat{\alpha}_t = a_t + P_tr_{t-1}, \qquad \hat{P}_t = P_t - P_tN_{t-1}P_t$$

$L_t = B - K_tZ,\ r_n = 0,\ N_n = 0,\ t = n, \ldots, 1.$

If Z, H, B, Q are constant, the variance recursions have a **steady-state** solution

Score-driven representation of Kalman Filter: Steady-state

Being

$$\log p(y_t|\mathcal{F}_{t-1}) = \text{const} - \frac{1}{2}(log|F_t| + v_t' F_t^{-1} v_t), \quad \text{with} \quad v_t = y_t - Z a_t$$

$$\nabla_t \equiv \left[\frac{\partial \log p(y_t|\mathcal{F}_{t-1})}{\partial a_t'}\right]' = Z' F_t^{-1} v_t, \quad \mathcal{H}_{t|t-1} = \mathcal{I}_{t|t-1} = \mathrm{E}_{t-1}[\nabla_t \nabla_t']' = Z' F_t^{-1} Z$$

In the steady-state: $\mathcal{H} = Z'\bar{F}^{-1}Z$, $\bar{F} = Z'\bar{P}Z + H$ with $\bar{P}$ the steady-state variance matrix. Let's define $A \equiv B\bar{P}$, then:

Proposition

In the steady-state, KFS recursions for the mean can be represented as:

$$\begin{aligned} a_{t+1} &= c + Ba_t + B\bar{P}Z'F_t^{-1}v_t \\ &= c + Ba_t + A\nabla_t \end{aligned} \qquad \begin{aligned} a_{t|t} &= a_t + \bar{P}Z'F_t^{-1}v_t \\ &= a_t + B^{-1}A\nabla_t \end{aligned}$$

$$\begin{aligned} r_{t-1} &= Z'F_t^{-1}v_t + (B - B\bar{P}Z'\bar{F}^{-1}Z)'r_t \\ &= \nabla_t + (B - A\mathcal{H})'r_t \end{aligned} \qquad \begin{aligned} \hat{\alpha}_t &= a_t + \bar{P}r_{t-1} \\ &= a_t + B^{-1}Ar_{t-1} \end{aligned}$$

- Kalman recursions for the mean re-parametrized in terms of ∇_t and $\mathcal{H}$
- The new representation is more general, as it only relies on the conditional density $p(y_t|\mathcal{F}_{t-1})$, which is defined for any observation-driven model.

Score-driven representation of Kalman Filter: General case

It is a standard practice in SD models to use the **scaled-score** $s_t \equiv S_t \nabla_t$ with scaling matrix S_t typically chosen as $S_t = \mathcal{I}_{t|t-1}^{-\varphi}$ with $\varphi = 0, 1/2, 1$.

Proposition

Outside the steady-state, KF recursions for the mean can be represented as:

$$a_{t+1} = c + \boldsymbol{B}a_t + \boldsymbol{B}P_t Z' F_t^{-1} v_t = c + \boldsymbol{B}a_t + \boldsymbol{B}P_t \nabla_t$$

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

$$P_{t+1} = \boldsymbol{Q} + \boldsymbol{B}P_t\boldsymbol{B} - \boldsymbol{B}P_t Z' F_t^{-1} Z P_t \boldsymbol{B}' = \boldsymbol{Q} + \boldsymbol{B}P_t\boldsymbol{B} - \boldsymbol{B}P_t \mathcal{H}_{t|t-1} P_t \boldsymbol{B}'$$

$$P_{t|t} = P_t - P_t Z' F_t^{-1} Z P_t = P_t - P_t \mathcal{H}_{t|t-1} P_t$$

- The **static parameters** to be estimated are **B** and **Q**
- The prediction error matrix of the state P_t in KF is the scaling matrix S_t in SD
- KF (outside steady-state) is then equivalent to a SD model with scaling matrix P_t driven by an addional obs-driven recursion inversely related to $\mathcal{H}_{t|t-1}$
- **Remark:** The variance matrix P_t in the KF is related to **filtering uncertainty**

Score-driven representation of the Kalman Filter: Implications

The formal equivalence of the Kalman filter predictive equation in the steady-state with score-driven models allows:

- provide a further justification for the updating rule adopted in score-driven models
- extend tools available for the Kalman filter to the score-driven approach. Namely:
  1. the update filter,
  2. the smoother
  3. the computation of confidence bands accounting for filtering uncertainty.

The Score-Driven Update (SDU) and Score-Driven Smoother (SDS)

- Score-driven models can be viewed as **misspecified predictive filters** for nonlinear non-Gaussian state-space models
- We can adopt the same logic for the **update filter** $a_{t|t}$ and for the **smoother** $\hat{\alpha}_t$

Recalling that $y_t|\mathcal{F}_{t-1} \sim p(y_t|f_t, \Theta)$ and **scaled-score** $s_t \equiv S_t \nabla_t$, we generalize the KF recursions for the mean as:

$$f_{t+1} = \omega + Bf_t + As_t \quad (1)$$

$$f_{t|t} = f_t + B^{-1}As_t \quad (2)$$

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

$$r_{t-1} = s_t + (B - AS_t\mathcal{H}_{t|t-1})' r_t \quad (3)$$

$$\hat{f}_t = f_t + B^{-1}Ar_{t-1} \quad (4)$$

where $r_n = 0$ and $t = n, \ldots, 1$.

We name (2) **Score-Driven Update (SDU)** and (3)-(4) **Score-Driven Smoother (SDS)**.

SDU and SDS methodology

$$y_t|\mathcal{F}_{t-1} \sim p(y_t|f_t, \Theta)$$

1. Estimation of static parameters

$$\{\hat{\omega}, \hat{A}, \hat{B}\} = \underset{\omega, A, B}{\text{argmax}} \sum_{t=1}^{n} \log p(y_t|f_t, \Theta)$$

2. Predictive filter

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

3. Update filter

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

4. Backward smoother

$$r_{t-1} = s_t + (\hat{B} - \hat{A}S_t\mathcal{H}_{t|t-1})' r_t$$
$$\hat{f}_t = f_t + \hat{B}^{-1}\hat{A}r_{t-1}$$

- Maximization of closed-form likelihood + forward/backward recursions
- Can handle any obs density $p(y_t|f_t, \Theta)$, with a potentially large number of t.v.p.

Example: GARCH-SDS

Consider the observation density:

$$y_t | \mathcal{F}_{t-1} \sim \mathrm{N}(0, \sigma_t^2)$$

Setting $f_t = \sigma_t^2$ and $S_t = \mathcal{H}_{t|t-1}^{-1}$, the previous filtering recursions become:

$$\begin{aligned} f_{t+1} &= \omega + b f_t + a(y_t^2 - f_t) \\ f_{t|t} &= f_t + b^{-1} a(y_t^2 - f_t) \end{aligned}$$

$t = 1, \ldots, n$. The predictive filter is the standard GARCH(1,1) model. The smoothing recursions reduce to:

$$\begin{aligned} r_{t-1} &= y_t^2 - f_t + (b - a) r_t \\ \hat{f}_t &= f_t + b^{-1} a r_{t-1} \end{aligned}$$

$t = n, \ldots, 1, \; r_n = 0.$

Example: GARCH-SDS

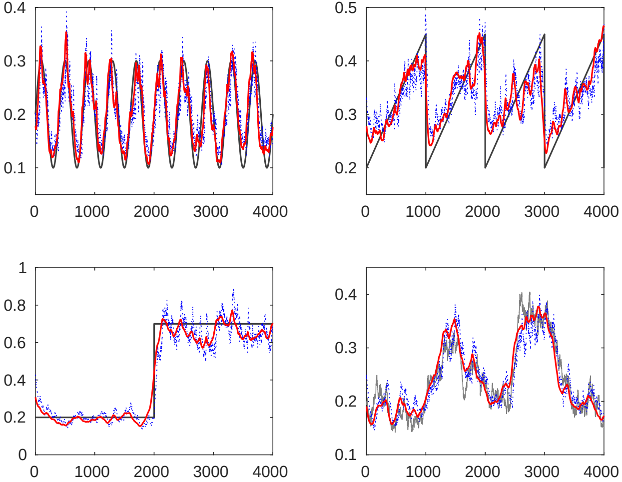


Figure: Filtered (blue dotted), and smoothed (red) estimates of GARCH(1,1) model

Other examples

- Multiplicative Error Model (Engle 2002)

$$y_t|\mathcal{F}_{t-1} \sim \text{Gamma}(\mu_t, \alpha)$$

- AR(1) with a time-varying coefficient

$$y_t|\mathcal{F}_{t-1} \sim \text{N}(c + \alpha_t y_{t-1}, \sigma^2)$$

- Beta-*t*-GARCH / *t*-GAS model (Harvey 2013, Creal et al. 2011)

$$y_t|\mathcal{F}_{t-1} \sim t_p(0, V_t, \nu)$$

- Realized Wishart-GARCH (Gorgi et al. 2018)

$$r_t|\mathcal{F}_{t-1} \sim \text{N}_p(0, V_t)$$
$$X_t|\mathcal{F}_{t-1} \sim \text{Wishart}_k(V_t/\nu, \nu)$$

Other examples: t-GAS

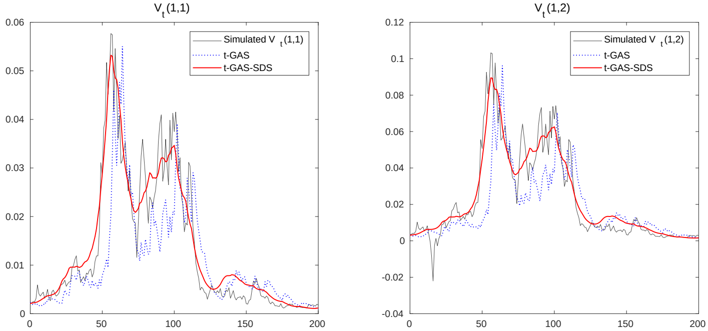


Figure: Comparison among simulated true covariances V_t (black lines), filtered (blue dotted lines) and smoothed (red lines) (co)variances of t-GAS model in the case $k = 5$. We show the variance corresponding to the first asset on the left and the covariance between the first and the second asset on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

Other examples: Realized Wishart-GARCH-SDS

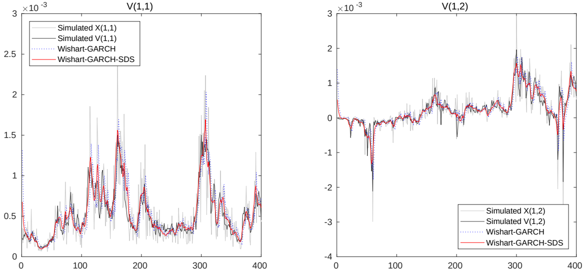


Figure: Comparison among simulated observations of X_t (grey lines), simulated true covariances V_t (black lines), filtered (blue dotted lines) and smoothed (red lines) (co)variances of realized Wishart-GARCH model in the case $k = 5$. We show the variance corresponding to the first asset on the left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

Kalman smoother vs SDS in linear non-Gaussian models

AR(1) + non-Gaussian noise:

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

Signal-to-noise ratio: $\delta = \sigma_\eta^2/\sigma_\epsilon^2$, $c = 0.01$, $\phi = 0.95$

- Counterpart observation-driven density: $y_t|\mathcal{F}_{t-1} \sim t(\mu_t, \varphi, \beta)$

	SDF-KF			SDS-KS		
δ	0.1	1	10	0.1	1	10
		$\nu = 3$			$\nu = 3$	
MSE	0.8610	0.9522	0.9991	0.8093	0.8876	0.9618
MAE	0.9389	0.9859	1.0036	0.9128	0.9634	1.0169
		$\nu = 5$			$\nu = 5$	
MSE	0.9552	0.9912	1.0032	0.9376	0.9880	1.0058
MAE	0.9792	0.9973	0.9999	0.9698	0.9949	1.0112
		$\nu = 8$			$\nu = 8$	
MSE	0.9877	0.9981	1.0029	0.9844	0.9954	1.0117
MAE	0.9939	0.9992	1.0039	0.9917	0.9982	1.0136

Parameter-driven Smoother vs SDS in non-linear models

1. SV with Gaussian measurement density

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

2. SV with non-Gaussian measurement density

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

3. Stochastic intensity

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

Two alternative methods:

- **NAIS** method of Koopman et al. (2015) for estimation + **importance sampling** for smoothing (for 1, 2, 3)
- **QML** approximate method of Harvey et al. (1994) for 1, 2

Loss measures computed on $N = 1000$ simulations of $n = 2000$ observations.

Nonlinear models - SV with Gaussian disturbances

- Counterpart observation-driven density: $y_t|\mathcal{F}_{t-1} \sim t(0, \varphi_t, \beta)$
- Reduces to the Beta-t-EGARCH model of Harvey and Chakravarty (2008)
- Coefficient of variation: $CV = \exp\left(\frac{\sigma_\eta^2}{1-\phi^2}\right) - 1$

	MSE				MAE			
CV	0.1	1	5	10	0.1	1	5	10
	$\phi = 0.98$				$\phi = 0.98$			
NAIS	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
	$\phi = 0.95$				$\phi = 0.95$			
NAIS	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
	$\phi = 0.90$				$\phi = 0.90$			
NAIS	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

Nonlinear models - SV with non-Gaussian disturbances

- Counterpart observation-driven density: $y_t|\mathcal{F}_{t-1} \sim t(0, \varphi_t, \beta)$
- Reduces to the Beta-t-EGARCH model of Harvey and Chakravarty (2008)
- Coefficient of variation: $CV = \exp\left(\frac{\sigma_\eta^2}{1-\phi^2}\right) - 1$

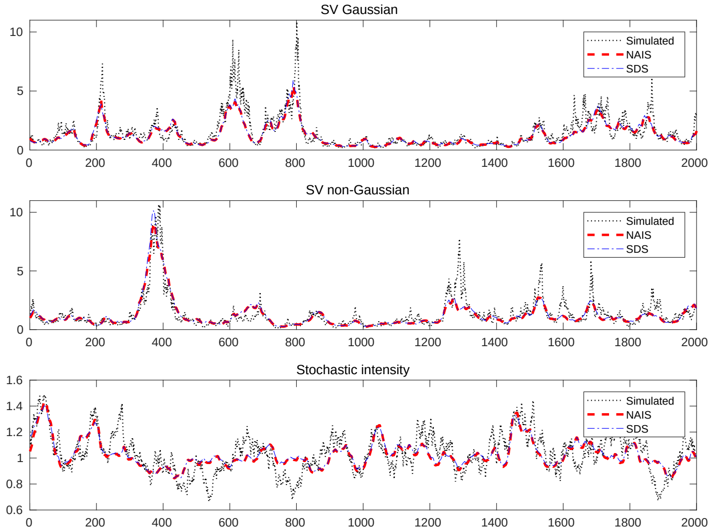
	MSE				MAE			
CV	0.1	1	5	10	0.1	1	5	10
	$\phi = 0.98,\ \nu = 3$				$\phi = 0.98,\ \nu = 3$			
NAIS	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
SDS	1.0026	0.9950	1.0140	1.0169	1.0015	0.9997	1.0077	1.0098
QML	1.3962	1.2553	1.2125	1.1998	1.1735	1.1184	1.1013	1.0939
	$\phi = 0.95,\ \nu = 3$				$\phi = 0.95,\ \nu = 3$			
NAIS	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
	$\phi = 0.90,\ \nu = 3$				$\phi = 0.90,\ \nu = 3$			
NAIS	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

Nonlinear models - Stochastic intensity

- Counterpart observation-driven density: $y_t|\mathcal{F}_{t-1} \sim \text{Poisson}(\lambda_t)$

	MSE			MAE		
$\sigma_\eta^2 \times 100$	0.1	0.5	1	0.1	0.5	1
		$\phi = 0.98$			$\phi = 0.98$	
NAIS	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
		$\phi = 0.95$			$\phi = 0.95$	
NAIS	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
		$\phi = 0.90$			$\phi = 0.90$	
NAIS	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

Comparison with exact smoothers



Empirical evidence

Does future (daily) returns help in estimating today's (co)volatilities?

- Compare SDF, SDU, SDS estimates of t-GAS model
- Proxy of today's covariance: Realized Covariance

	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: Absolute and relative RMSE and Qlike of SDF, SDU, SDS estimates of the t-GAS model for the four randomly selected portfolios with $n = 20$ (* implies inclusion in the Model Confidence Set).

Filtering uncertainty

What are the sources of statistical uncertainty in observation-driven models?

Observation-driven models as **D.G.P.**

Parameters are completely revealed by past observations ($P_{t+1} = \text{Var}[f_{t+1}|\mathcal{F}_t] = 0$)

$\downarrow$

Parameter uncertainty

Pascual et al. (2001), Blasques et al. (2016)

Observation-driven models as **filters**:

Conditionally on the past, parameters are stochastic ($P_{t+1} = \text{Var}[f_{t+1}|\mathcal{F}_t] > 0$)

$\downarrow$

Parameter + Filtering uncertainty

Filtering uncertainty, cont'd

Assume the MLE $\hat{\Theta}$ of static parameters are asymptotically normally distributed:

$$\sqrt{T}(\hat{\Theta} - \Theta_0) \overset{d}{\Rightarrow} \mathrm{N}(0, W)$$

We adopt the Bayesian perspective that Θ is a r.v. with posterior $\Theta \sim \mathrm{N}(\hat{\Theta}, \frac{1}{T}\hat{W})$

In a linear Gaussian model, the predictive filter $a_t^{\hat{\Theta}} = \mathrm{E}[\alpha_t | \mathcal{F}_{t-1}, \hat{\Theta}]$ depends on $\hat{\Theta}$.

It is possible to show (Hamilton 1986) that:

$$\begin{aligned}
&\mathrm{E}[(\alpha_t - a_t^{\hat{\Theta}})(\alpha_t - a_t^{\hat{\Theta}})' | \mathcal{F}_{t-1}] = \\
&\mathrm{E}_\Theta[(\alpha_t - a_t^{\Theta})(\alpha_t - a_t^{\Theta})' | \mathcal{F}_{t-1}] + \mathrm{E}_\Theta[(a_t^{\Theta} - a_t^{\hat{\Theta}})(a_t^{\Theta} - a_t^{\hat{\Theta}})'] = \\
&\underbrace{\mathrm{E}_\Theta[P_t(\Theta)]}_{\text{Filtering uncertainty}} + \underbrace{\mathrm{E}_\Theta[(a_t^{\Theta} - a_t^{\hat{\Theta}})(a_t^{\Theta} - a_t^{\hat{\Theta}})']}_{\text{Parameter uncertainty}}
\end{aligned}$$

where $\mathrm{E}_\Theta[\cdot]$ is the expectation w.r.t. the posterior density of Θ.

$\Rightarrow$ construct confidence bands quantifying both filtering and parameter uncertainty.

Confidence bands for the SV model

Nominal confidence level	90%	95%	99%
	Average coverage		
	Parameter uncertainty		
SDF	0.8421	0.9133	0.9783
SDU	0.8458	0.9078	0.9755
SDS	0.8501	0.9190	0.9890
	Parameter + filtering uncertainty		
SDF	0.8921	0.9459	0.9912
SDU	0.8952	0.9506	0.9931
SDS	0.8987	0.9482	0.9930

Table: Average coverage of in-sample confidence bands of SV model. We report results for three different nominal confidence levels, namely 90%, 95%, 99%.

.

Conclusions

We introduced a general methodology allowing to recover smoothed estimates of observation-driven models, as well as evaluating filtering uncertainty.

Main features of the methodology

- General, as it maintains the same form for any observation density
- Computationally simple, being forward/backward recursions
- Can easily handle models with high-dimensional t.v.p. vector
- Very close to exact methods based on IS (MSE loss lower, on average, than 2.5%)
- Average coverage of confidence bands in agreement with nominal confidence levels

Appendix

Other examples: MEM-SDS

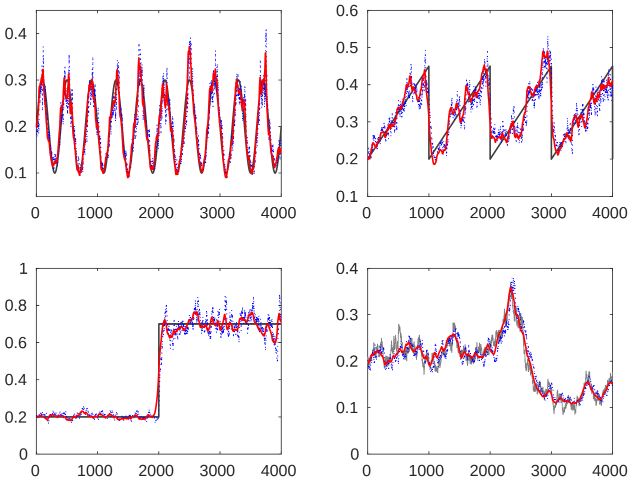


Figure: Filtered (blue dotted), and smoothed (red) estimates of MEM(1,1) model

Other examples: RV-GB2 (Burr)

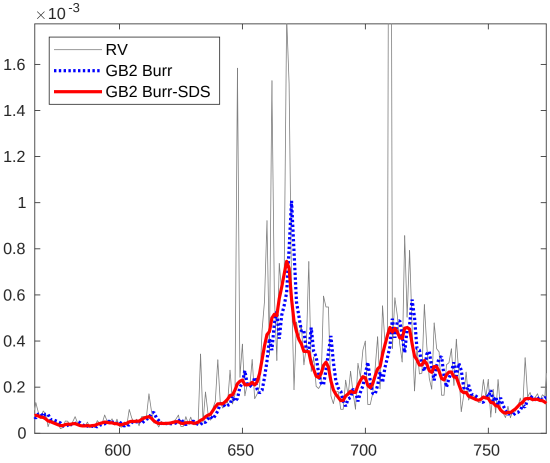


Figure: Filtered (blue dotted), and smoothed (red) estimates of RV-GB2 (Burr) model on S&P500 future data