Smooth particle filters for likelihood evaluation and maximisation

Michael Pitt

Department of Economics, University of Warwick, Coventry CV4 7AL

M.Pitt@warwick.ac.uk

September 20, 2001

Models:

General Form:

$$\begin{array}{lcl} \text{meas eqn: } y_t & \sim & f(y_t|\alpha_t) \\ \text{trans eqn: } \alpha_{t+1} & \sim & f(\alpha_{t+1}|\alpha_t), \;\; t = 1, ..., n \end{array}$$

The task we are concerned is the estimation of the log-likelihood,

$$\begin{aligned} \log L(\theta) &= \log f(y_1, ..., y_n|\theta) \\ &= \log f(y_1|\theta) + \sum_{t=0}^{n-1} \log f(y_{t+1}|\theta; \mathcal{F}_t), \end{aligned}$$

via the *prediction decomposition*. $\mathcal{F}_t = (y_1', ..., y_t')'$.

In order to estimate the log-likelihood we exploit the relationship

$$f(y_{t+1}|\theta; \mathcal{F}_t) = \int f(y_{t+1}|\alpha_{t+1}; \theta) f(\alpha_{t+1}|\mathcal{F}_t; \theta) d\alpha_{t+1}. \tag{0·1}$$

for the *prediction density*.

- We have a filtering device, the PF, which delivers samples from $\alpha_t^k \sim f(\alpha_t|\mathcal{F}_t;\theta)$, $k = 1, ..., M$.

- We can sample from the transition density $f(\alpha_{t+1}|\alpha_t;\theta)$.

- So can estimate estimate (0·1).

$$\widehat{f}(y_{t+1}|\mathcal{F}_t,\theta) = \int f(y_{t+1}|\alpha_{t+1},\theta)\left\{\frac{1}{M}\sum_{k=1}^{M} f(\alpha_{t+1}|\alpha_t^k)\right\} d\alpha_{t+1}.$$

Lots of alternative procedures for such latent models:

- MCMC: Bayesian inference, Carlin et al. (1992), Shephard (1996),Jacquier et al. (1994)..allows us to get a strictly stationary sample form $f(\alpha, \theta|y)$.
- Indirect Inference. Simulation based, can be inefficient, Gourieroux & Renault (1993).
- Importance Sampling: see, for instance Durbin and Koopman (97, Biom), (00, JRSSB). $g()$ Gaussian.

$$\begin{aligned} f(y|\theta) &= \int f(y|\alpha)g(\alpha|\theta)d\alpha \\ &= \int \frac{f(y|\alpha)}{g(y|\alpha)} g(y|\alpha)g(\alpha|\theta)d\alpha. \end{aligned}$$

Setting

$$\omega(\alpha) = \frac{f(y|\alpha)}{g(y|\alpha)}, \text{ and observing } g(y|\alpha)g(\alpha|\theta) = g(\alpha|\theta, y)g(y|\theta),$$

we get

$$\begin{aligned} f(y|\theta) &= \int f(y|\alpha)g(\alpha|\theta)d\alpha \\ &= g(y|\theta)\int \omega(\alpha)g(\alpha|\theta, y)d\alpha. \end{aligned}$$

So they estimate as

$$g(y|\theta)\frac{1}{M}\sum_{i=1}^{M}\omega(\alpha^i),$$

where $\alpha^i \sim g(\alpha|\theta, y)$. Problems for high dimensions as the Variance of $\omega(\alpha)$ rises exponentially (may not be finte to begin with). So variability of estimator rises exponentially with the number of observations over time.

Geyer's LR approach for general latent models:

$$\frac{f(y|\theta)}{f(y|\overline{\theta})} = \int \frac{f(\alpha|\theta)}{f(\alpha|\overline{\theta})} f(\alpha|y;\overline{\theta}) d\alpha.$$

Examples:

- SDE observed with noise

$$\begin{aligned} y(\tau_i) &\sim f(y(\tau_i) \mid x(\tau_i)) \\ dx(t) &= \mu(x(t))dt + \sigma(x(t))dW(t). \end{aligned}$$

  e.g, interest rate model + noise. Can simulate (arbitrarily accurately) from $f(x(\tau_{i+1}) \mid x(\tau_i))$ by exploiting Euler scheme.

  Let $\delta = (\tau_{i+1} - \tau_i)/M$, $z(0) = x(\tau_i)$

$$x(t+\delta)|x(t) \sim N(x(t) + \mu(x(t))\delta; \sigma^2(x(t))\delta), \tag{0·2}$$

- Full discrete time SV model, Shephard (1996)

$$
\begin{aligned}
y_t &= \mu + \beta\sigma_t^{\gamma} + \epsilon_t \sigma_t \chi_t, && (0\cdot3)\\
& && (0\cdot4)\\
\log \sigma_{t+1}^2 &= (1-\phi)\mu + \phi \log \sigma_t^2 + \eta_t, && (0\cdot5)\\
& && (0\cdot6)\\
\begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} &\sim NID\left\{0, \begin{pmatrix} 1 & \\ 0 & \sigma_\eta^2 \end{pmatrix}\right\} && (0\cdot7)\\
\chi_t &\sim Iga(\nu, 1). && (0\cdot8)
\end{aligned}
$$

- GARCH + noise model,

$$\begin{aligned} y_t|\alpha_t &\sim N(\alpha_t, \sigma^2), \\ \alpha_t|\sigma_t^2 &\sim N(0, \sigma_t^2) \\ \sigma_{t+1}^2 &= \beta_0 + \beta_1\alpha_t^2 + \beta_2\sigma_t^2. \end{aligned}$$

1. Likelihoods also used in Bayes factors,

$$f(y|M) = \frac{f(y|\theta)f(\theta)}{f(\theta|y)}$$

1 Efficient likelihood methods for general PFs:

general approach Pitt and Shephard(99). We then wish to sample from the following target density

$$\begin{aligned} f(\alpha_{t+1}, k|\mathcal{F}_{t+1}) &\propto f_1(y_{t+1}|\alpha_{t+1})f_2(\alpha_{t+1}|\alpha_t^k), \quad k = 1, ..., M \\ &\simeq \overline{g}(k, \alpha_{t+1}) = \overline{g}_1(y_{t+1}|\alpha_{t+1}, k)g_2(\alpha_{t+1}|\alpha_t^k) \\ &= \overline{g}(y_{t+1}|k)g(\alpha_{t+1}|k, y_{t+1}) = C.g(k, \alpha_{t+1}). \end{aligned} \tag{1·1}$$

So we now have a joint density $g(k, \alpha_{t+1})$,

$$\overline{g}(y_{t+1}|k) = \int \overline{g}(k, \alpha_{t+1})d\alpha_{t+1}, \quad g(\alpha_{t+1}|k, y_{t+1}) = \frac{\overline{g}(k, \alpha_{t+1})}{\overline{g}(y_{t+1}|k)},$$

and $C = \Sigma_{i=1}^{M} \overline{g}(y_{t+1}|i)$. So for our joint density $g(k, \alpha_{t+1})$, we have

$$g(k) = \frac{\overline{g}(y_{t+1}|k)}{\Sigma_{i=1}^{M} \overline{g}(y_{t+1}|i)}, \qquad g(\alpha_{t+1}|k) = g(\alpha_{t+1}|k, y_{t+1}). \tag{1·2}$$

Sample from our joint proposal density $g(k, \alpha_{t+1})$, R times we then allocate weights to the resulting samples $\left(\alpha_{t+1}^{j}, k^{j}\right)$, $j = 1, ..., R$,

$$\omega_j = \omega(\alpha_{t+1}^{j}, k^{j}), \qquad \pi_j = \frac{\omega_j}{\Sigma_{i=1}^{R} \omega_i}.$$

where

$$\omega(\alpha_{t+1}, k) = \frac{f_1(y_{t+1}|\alpha_{t+1}) f_2(\alpha_{t+1}|\alpha_t^k)}{\overline{g}_1(y_{t+1}|\alpha_{t+1}, k) g_2(\alpha_{t+1}|\alpha_t^k)}. \tag{1·3}$$

We require,

$$\widehat{f}(y_{t+1}|\mathcal{F}_t) = \int f_1(y_{t+1}|\alpha_{t+1}) \left\{ \sum_{k=1}^{M} f_2(\alpha_{t+1}|\alpha_t^k) \frac{1}{M} \right\} d\alpha_{t+1}. \qquad (1{\cdot}4)$$

This integral cannot, in general, be evaluated directly and so needs to be estimated.

THEOREM 1:

$$\widehat{f}(y_t|\mathcal{F}_{t-1}, \theta) = \left[\frac{1}{M} \sum_{i=1}^{M} \overline{g}(y_t|i) \right] E\left[\omega(\alpha_t; k)\right]$$

where $\omega(\alpha_t; k)$ is given by (1·3) and the expectation is with respect to $g(k, \alpha_t)$ given by (1·2).

This above result is useful practically because it means we can take the sample mean of the first stage weights and the sample mean of the second stage weights. So the likelihood $\widehat{f}(y_t|\mathcal{F}_{t-1}, \theta)$ is unbiassedly estimated as

$$\left[\frac{1}{M} \sum_{i=1}^{M} \overline{g}(y_t|i) \right] \left[\frac{1}{R} \sum_{j=1}^{R} \omega_j \right]. \qquad (1{\cdot}5)$$

Clearly when the ω_j have small variance (our criterion for an efficient particle filter) our estimator will be efficient statistically. Indeed, when the variance of the ω_j is 0, corresponding to full adaption, see Section **??**, then we will be exactly evaluating (**??**).

Can be regarded as a free bi-product of our auxiliary sampling scheme.

Example 1: AR(1) + noise model , GSSF, so use KF.

$$
\begin{array}{llll}
\text{meas eqn: } y_t & = & \alpha_t + \varepsilon_t & \varepsilon_t \sim N(0, 4.9) \\
\text{trans eqn: } \alpha_{t+1} & = & \beta + \phi(\alpha_t - \beta) + \eta_t, & \eta_t \sim N(0, 0.02)
\end{array}
$$

Do likelihood evaluation. $t = 1, ..., n$. $n = 550$, $\phi = 0.975$, $\beta = 0.5$.

Smooth Likelihood Estimation

It turns out that we may adjust the SIR procedure of GSS above quite easily to produce an entirely continuous likelihood surface.

The problem is inherent in the discreteness of the resampling:

At the sample stage (3) we replace the discrete distribution. we define a region i, R_i as follows: $R_i = [x^i, x^{i+1}]$, $k = 1, ..., R-1$. These regions form a partition of the sample space for x. We have different densities $g(x|i)$ within each region i, R_i. We shall assign $\Pr(i) = \frac{1}{2}(\pi_i + \pi_{i+1})$, $i = 2, ..., R-2$ and $\Pr(1) = \frac{1}{2}(2\pi_1 + \pi_2)$, $\Pr(R-1) = \frac{1}{2}(\pi_{R-1} + 2\pi_R)$. Clearly, these probabilities sum to 1. Within each region we shall define the conditional densities as follows,

$$g(x|i) = \frac{1}{(x^{i+1} - x^i)},\ x \in R_i,\ i = 2, ..., R-2,$$

and

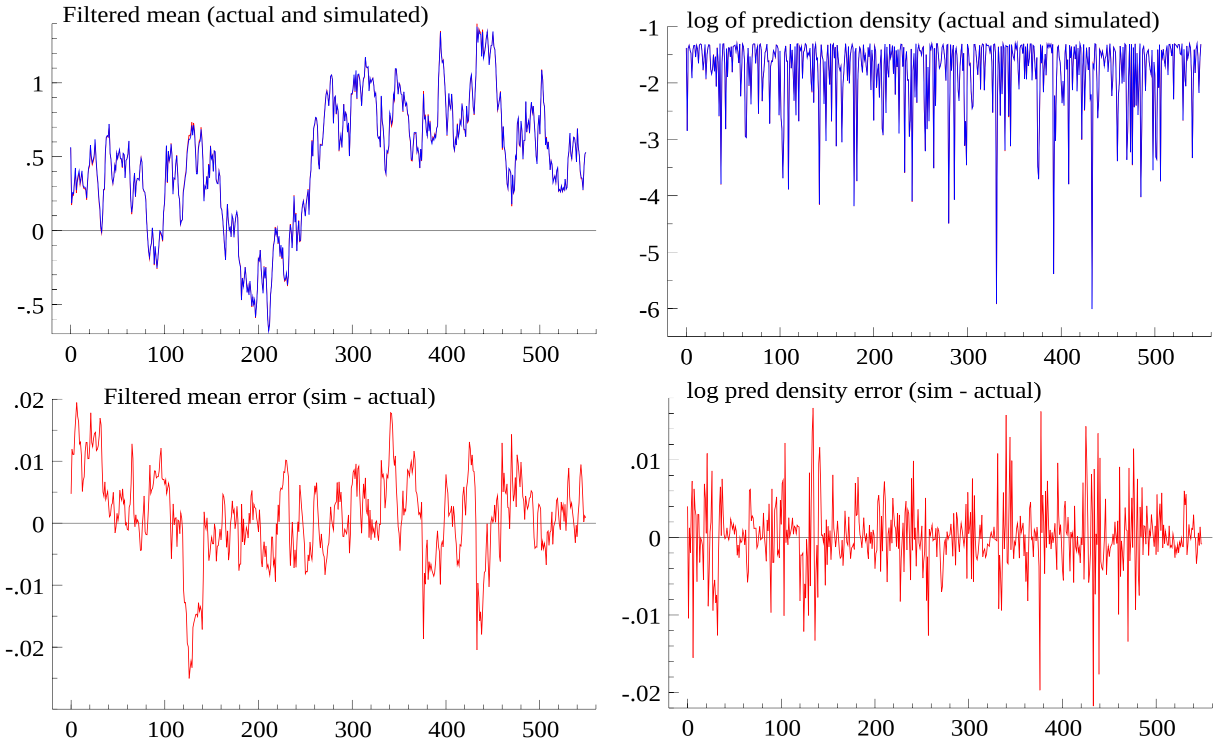

Figure 1: *Left: top is sample filter mean and true KF mean for AR(1) plus noise example. Beneath is the error. Right: top is sample estimate of log of prediction density. Beneath is error.* $T = 550$, $M = 3000$, $R = 4000$. *Auxiliary particle filter with stratification used.*

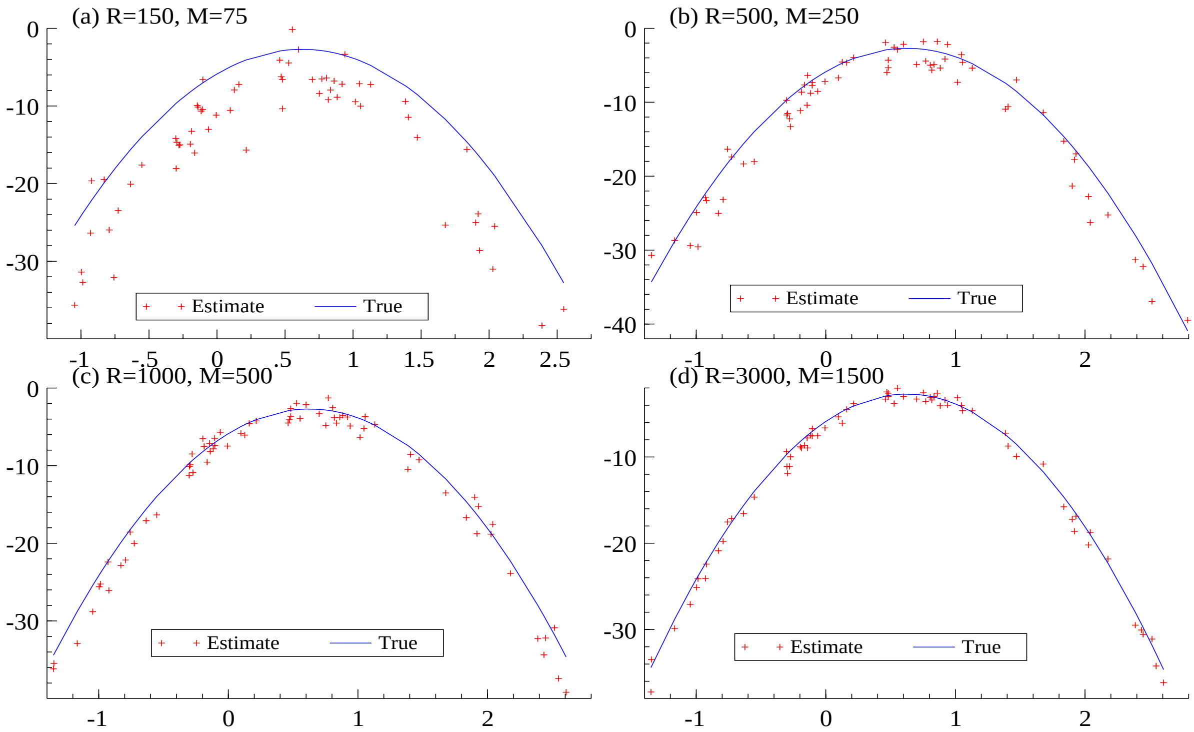

Figure 2: *Estimators of the relative log-likelihood computed via simulation estimators of the prediction decomposition. The graphs plot the estimated log-likelihood against the β, which means the true value is 0.5. Each value of β is used as a strata.*

$$g(x|1) = \begin{cases} \frac{\pi_1}{2\pi_1+\pi_2}, & x = x^1 \\ \\ \frac{\pi_1+\pi_2}{2\pi_1+\pi_2}\frac{1}{(x^2-x^1)}, & x \in R_1 \end{cases}$$

$$g(x|R-1) = \begin{cases} \frac{\pi_R}{\pi_{R-1}+2\pi_R}, & x = x^R \\ \\ \frac{\pi_{R-1}+\pi_R}{\pi_{R-1}+2\pi_R}\frac{1}{(x^R-x^{R-1})}, & x \in R_{R-1}. \end{cases}$$

Figure **??** shows a discrete cdf with the continuous interpolation for $R = 8$. Note that the continuous cdf passes through the mid-point of each step in the discrete cdf. As R becomes larger, the two cdfs become indistinguishable. The validity of the resampling method for SIR is preserved. Denoting our continuous cdf by $\widetilde{F}$ and the discrete SIR cdf by $\widehat{F}$, it can be seen that as $R \to \infty$,

$$\widetilde{F}(z) \to \widehat{F}(z) \to F(z),$$

where $F(z)$ is the true cdf. The justification for the convergence of $\widehat{F}(z)$ to $F(z)$ is given rather succinctly by Smith & Gelfand (1992).

The partitioning of the state space means that sampling from this continuous density is very efficient. We simply select the region i with $\Pr(i)$ and sample from $g(x|i)$. The form of $g(x|i)$ has been chosen to be linear. This ensures continuity and allows very fast sampling but there is no reason why a quadratic or cubic interpolation could not be used within each region. Indeed differentiability could be achieved by using a higher order interpolation.

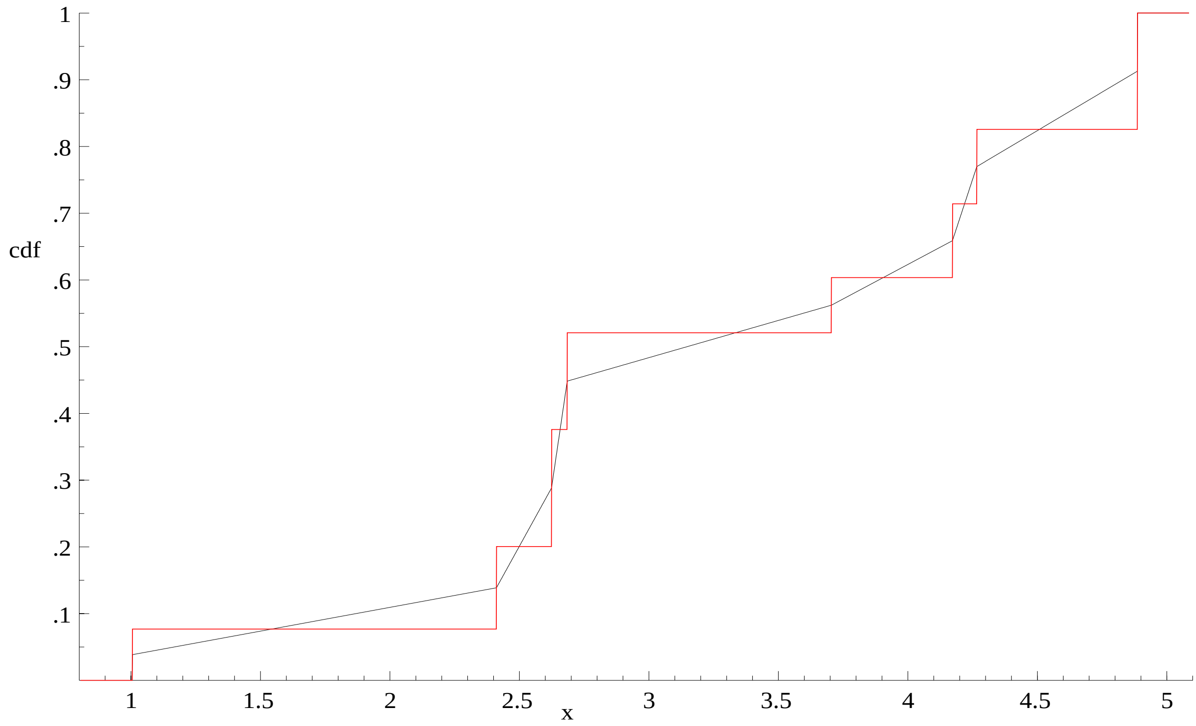

Figure 3: *cdf plot with 8.*

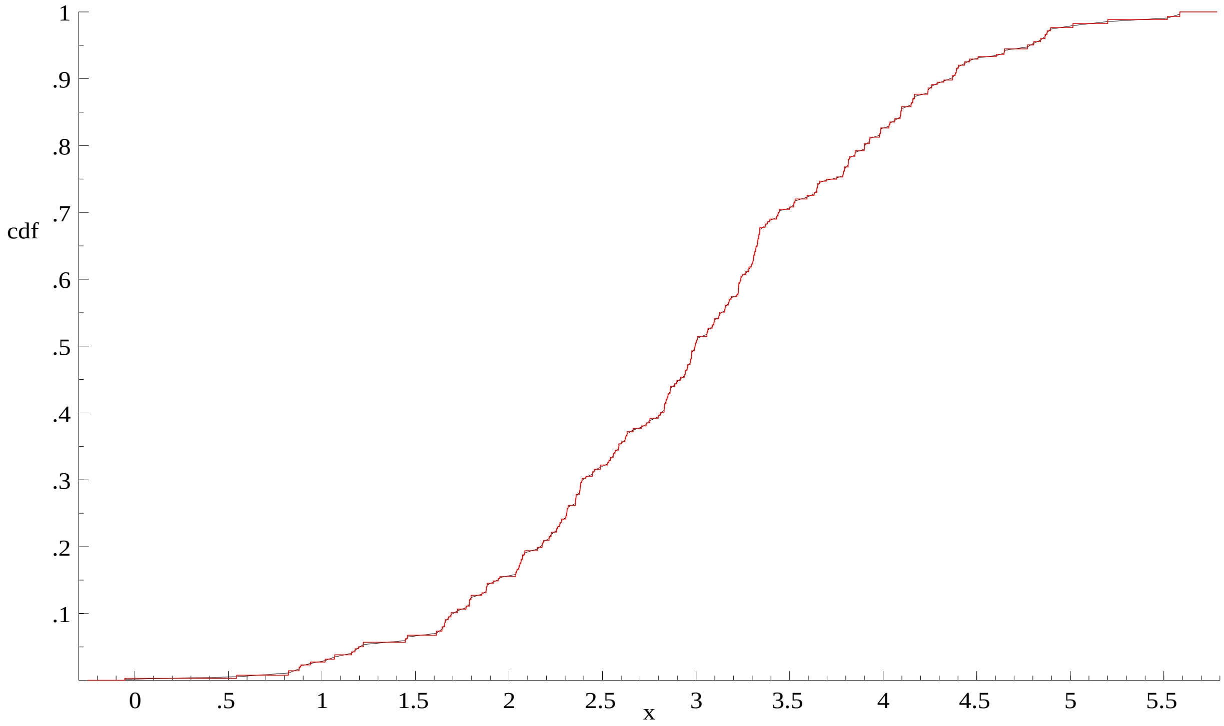

Figure 4: *cdf plot with 200.*

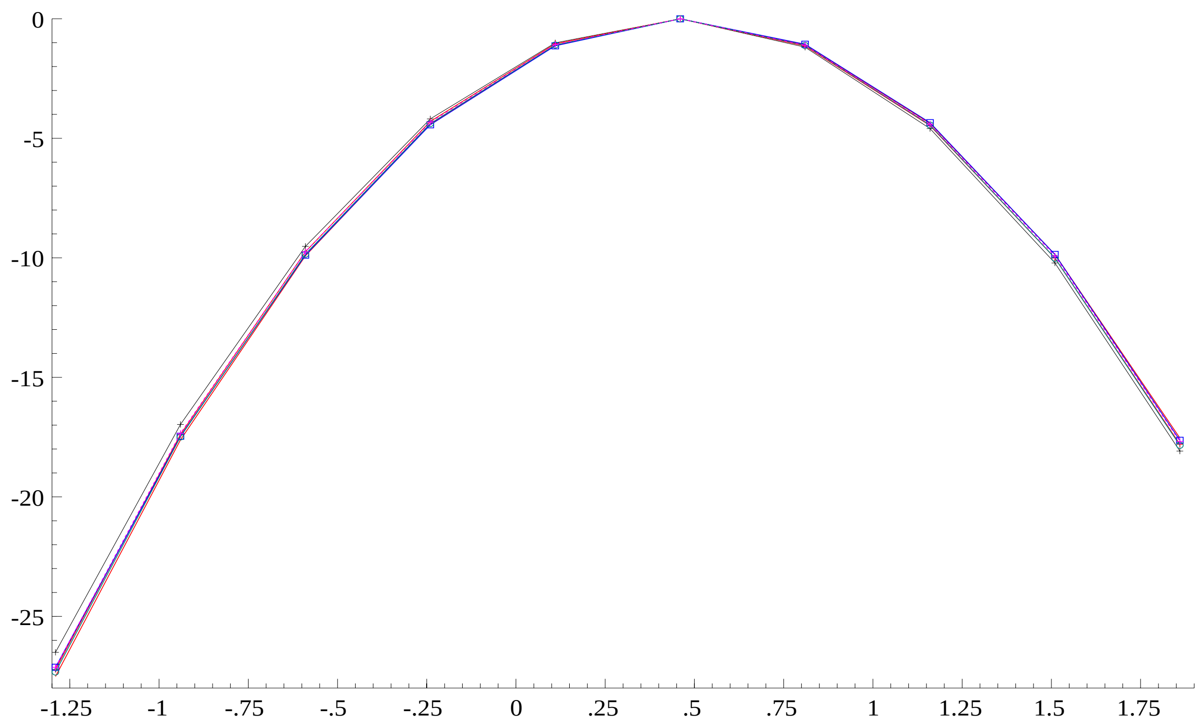

Figure 5: *Plot of the true log-likelihood (solid line), and five simulated log-likelihoods (symbols with lines) via the auxiliary particle mean trajectory for AR(1) plus noise example. Log-likelihoods plotted against the mean,* μ*.* $T = 550$*,* $M = 1500$*,* $R = 3000$*. Large scale left, medium scale right.*

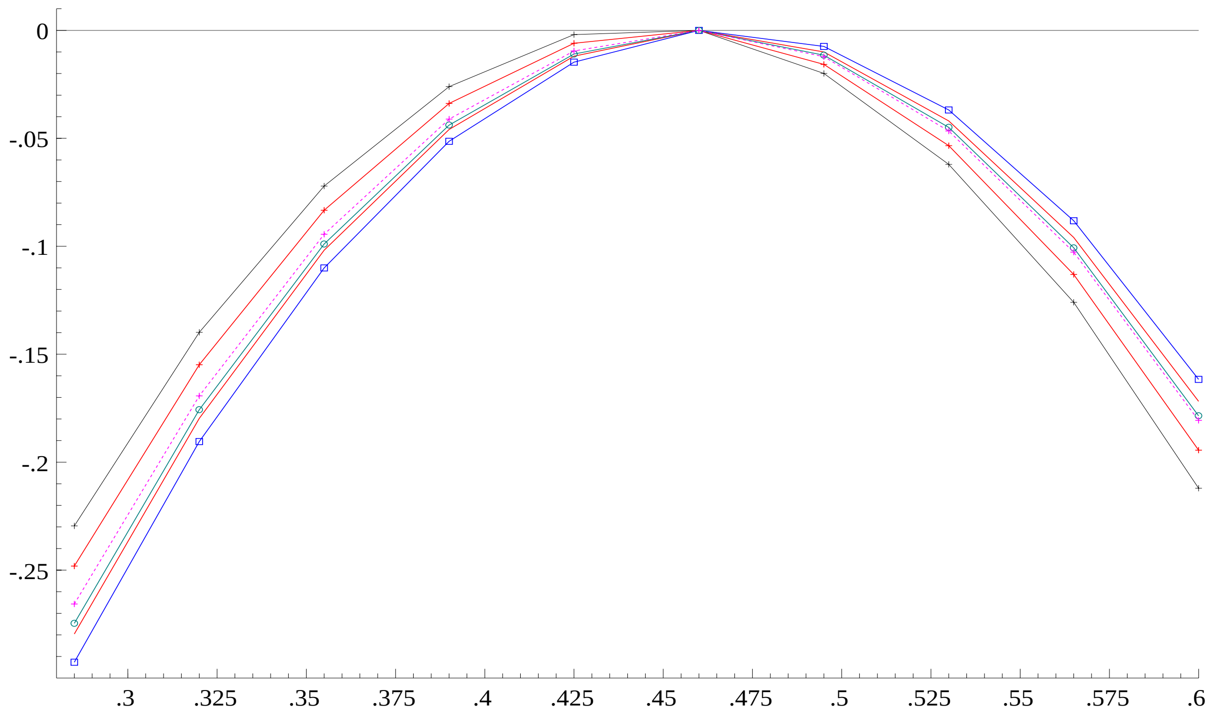

Figure 6: *Plot of the true log-likelihood (solid line), and five simulated log-likelihoods (symbols with lines) via the auxiliary particle mean trajectory for AR(1) plus noise example. Log-likelihoods plotted against the mean,* μ*.* $T = 550$*,* $M = 1500$*,* $R = 3000$*. Large scale left, medium scale right.*

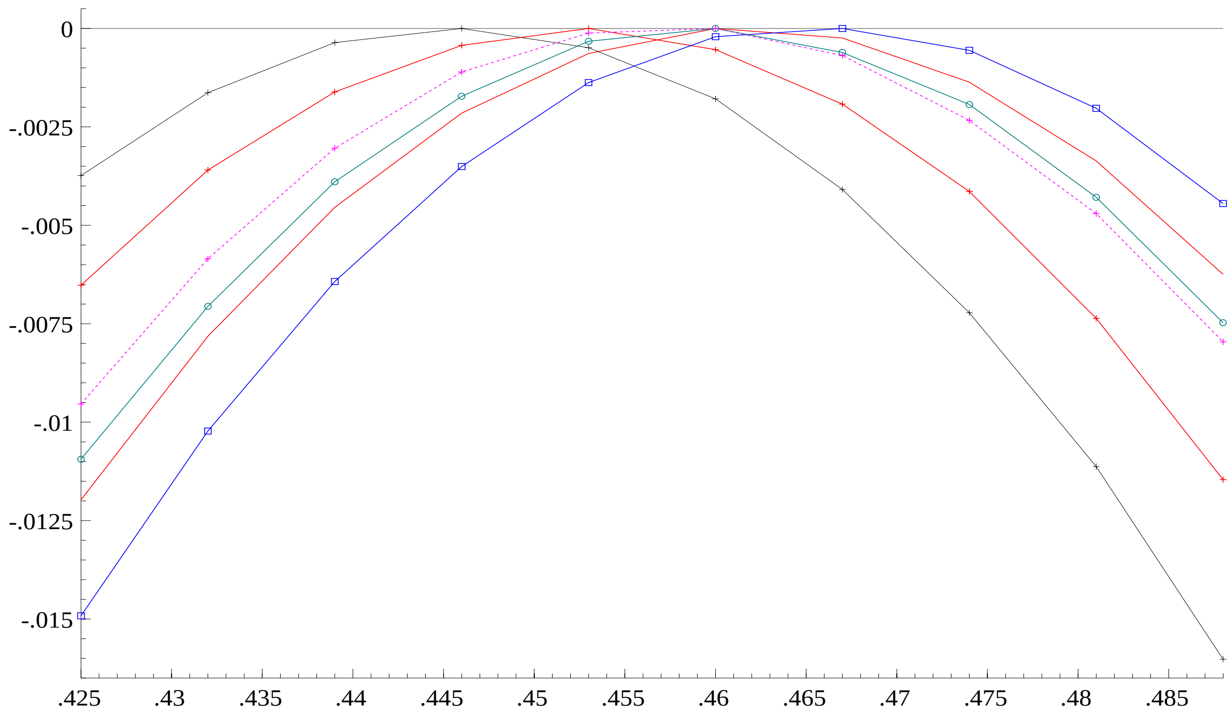

Figure 7: *Plot of the true log-likelihood (solid line), and five simulated log-likelihoods (symbols with lines) via the auxiliary particle mean trajectory for AR(1) plus noise example. Log-likelihoods plotted against the mean,* μ. $T = 550$, $M = 1500$, $R = 3000$. *Large scale left, medium scale right. Small scale.*

Results: AR(1) + noise model

To assess the performance of the ASIR_0 method we shall consider the AR(1) plus noise model. This is a linear state space form model the likelihood for which can be evaluated via the Kalman filter. The model is,

$$\begin{aligned} y_t &= \alpha_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_\varepsilon^2) \\ \alpha_{t+1} &= \mu + \phi(\alpha_t - \mu) + \eta_t, \quad \eta_t \sim N(0, \sigma_\eta^2). \end{aligned}$$

To mimic the stochastic volatility (SV) model, see Section **??** we have $\sigma_\varepsilon^2 = 2$, $\sigma_\eta^2 = 0.02$, $\phi = 0.975$ and $\mu = 0.5$. The choice of σ_η^2, ϕ and μ are chosen as typical values for the SV model, ϕ representing the persistence in variance, whilst σ_ε^2 is chosen from the curvature for the measurement density in the SV model (the second derivative of $\log f(y_t|\alpha_t)$ with respect to α_t).

A Fair test.

Recorded in Table 1 are the results for $T = 150$, using varying values of M and R. The average of the 50 simulated maximum likelihood estimates, the

50 variance estimates and the mean squared error are displayed for each set of M, R. The mean squared errors in Table 1 are small relative to the variation in the data, and become smaller as M, R increase. In addition the variance-covariance matrix is well estimated even for small M, R. These results are very encouraging.

The results for the case $T = 550$, Table 3, gives an insight into how the method might behave for the SV model for which the data is reasonably long.

Kalman Filter

$ML\,(\sigma_\eta,\,\mu,\,\phi)'$	$Var\,(\sigma_\eta,\,\mu,\,\phi)' \times 10^3$			
0.07543	3.814	-1.112	-2.02	
0.58276	-1.112	40.31	1.273	
0.96626	-2.02	1.273	2.050	

SIR Particle Filter; $M = 300, R = 400, SIM = 50.$

$\overline{ML_S}$	$\overline{Var} \times 10^3$			$MSE(ML_S) \times 10^4$
0.07574	3.155	-0.702	-1.661	0.182
0.58170	-0.702	35.02	0.898	2.217
0.96610	-1.85	0.898	1.846	0.103

SIR Particle Filter; $M = 1000, R = 1300, SIM = 50.$

$\overline{ML_S}$	$\overline{Var} \times 10^3$			$MSE(ML_S) \times 10^4$
0.07502	3.524	-0.934	-1.856	0.0628
0.58360	-0.934	35.75	1.129	0.5295
0.96623	-1.856	1.129	1.950	0.0450

SIRParticleFilter; $M = 3000, R = 4000, SIM = 50$

$\overline{ML_S}$	$\overline{Var} \times 10^3$			$MSE(ML_S) \times 10^4$
0.075200	3.693	-0.895	-1.937	0.0190
0.58177	-0.895	37.00	1.059	0.1495
0.96629	-1.937	1.059	1.988	0.0123

Table 1: *Performance of standard smooth SIR particle filter for* $T = 150$. The model is AR(1) + noise with truw parameters $\phi = 0.975$, $\sigma_\eta = \sqrt{0.02}$, $\mu = 0.5$. *Additionally the measurement noise is fixed at* $\sigma_\varepsilon = \sqrt{2}$.

Kalman Filter

$ML\,(\sigma_\eta,\,\mu,\,\phi)'$	$Var\,(\sigma_\eta,\,\mu,\,\phi)' \times 10^4$			
0.08080	7.670	1.113	−2.137	
0.45900	1.113	429.1	−0.8227	
0.98398	−2.137	−0.8227	1.223	

SIR Particle Filter; $M = 300, R = 400, SIM = 50.$

$\overline{ML_S}$	$\overline{Var} \times 10^4$			$MSE(ML_S) \times 10^4$
0.080774	7.702	−0.203	−2.156	0.1347
0.45380	−0.203	450.8	−0.8182	4.666
0.98395	−2.156	−0.8182	1.241	0.02287

SIRParticleFilter;$M = 1000, R = 1300, SIM = 50.$

$\overline{ML_S}$	$\overline{Var} \times 10^4$			$MSE(ML_S) \times 10^4$
0.080752	7.754	0.377	−2.155	0.0307
0.45965	0.377	437.3	−0.452	1.02
0.98398	−2.155	−0.452	1.224	0.00439

SIRParticleFilter; $M = 3000, R = 4000, SIM = 50$

$\overline{ML_S}$	$\overline{Var} \times 10^4$			$MSE(ML_S) \times 10^4$
0.08061	7.459	0.7897	−2.050	0.0101
0.45812	0.7897	437.0	−0.813	0.3253
0.98405	−2.050	−0.813	1.1882	0.00150

Table 2: *Performance of standard smooth SIR particle filter for* $T = 550$. The model is AR(1) + noise with truw parameters $\phi = 0.975$, $\sigma_\eta = \sqrt{0.02}$, $\mu = 0.5$. *Additionally the measurement noise is fixed at* $\sigma_\varepsilon = \sqrt{2}$.

1. **Stratified sampling**:

This stratification scheme is briefly described in **?**) and uses the suggestion of **?**). Explicitly, at the resampling stage, we produce stratified uniforms $\tilde{u}^1, ..., \tilde{u}^R$ by writing

$$\tilde{u}^k = \frac{(k-1)+u}{R}, \quad k = 1, ..., R \text{ where } \quad u \overset{iid}{\sim} UID(0,1).$$

That is we use a single uniform realisation to generate sorted stratified uniforms on $[0,1]$. An efficient method, see PS, for inverting the cdf is then used to produce the stratified sorted samples of our variables. **?**) justify using stratification ideas via sampling.

2. Bias correction:

Note that at present the log-likelihood will not be unbiassed. To correct this to first order we use the usual Taylor expansion method. Abstracting from likelihoods we have the large sample result that our estimated likelihood, $\overline{X}$ is unbiassed for the true likelihood, μ with and for large R we obtain,

$$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{R}\right).$$

We therefore have

$$E[\log \overline{X}] = \log \mu - \frac{1}{2}\frac{\sigma^2}{R\mu^2}.$$

Therefore we may bias correct by substituting μ as $\overline{X}$, obtaining

$$\widehat{\log} \mu = \log \overline{X} + \frac{1}{2}\frac{\widehat{\sigma}^2}{R\overline{X}^2}.$$

2 Full Adaption:

- We can draw directly from $f(\alpha_{t+1}|\alpha_t, y_{t+1})$.

- We can evaluate $f(y_{t+1}|\alpha_t)$.

  In fact the empirical prediction density, (**??**) is exactly obtained as

$$\widehat{f}(y_{t+1}|\mathcal{F}_t, \theta) = \frac{1}{M}\sum_{k=1}^{M} f_1(y_{t+1}|\alpha_t^k),$$

  where $\alpha_t^k \sim f(\alpha_t|\mathcal{F}_t)$.

$$f(\alpha_t^k|y_{t+1}, \mathcal{F}_t) \propto f(y_{t+1}|\alpha_t^k).$$

  Then smoothly sample to yield $\alpha_t^j \sim f(\alpha_t|\mathcal{F}_{t+1})$. Then sample $f(\alpha_{t+1}|\alpha_t, y_{t+1})$ to yield

$$\alpha_{t+1}^j \sim f(\alpha_{t+1}|\mathcal{F}_{t+1}).$$

Example 1: ARCH with error

An example of full adaption is for the ARCH model observed with Gaussian error. Consider the simplest Gaussian ARCH model (see, for example, Bollerslev et al. (1994) for a review) observed with independent Gaussian error. So we have

$$y_t|\alpha_t \sim N(\alpha_t, \sigma^2), \qquad \alpha_{t+1}|\alpha_t \sim N(0, \beta_0 + \beta_1\alpha_t^2).$$

It has received a great deal of attention in the econometric literature as it has some attractive multivariate generalizations: see the work by Diebold & Nerlove (1989), Harvey et al. (1992) and King et al. (1994). This model is exactly adaptable. It is clear to see that,

$$y_{t+1}|\alpha_t \ \sim \ N(0, \beta_0 + \beta_1\alpha_t^2 + \sigma^2), \qquad \alpha_{t+1}|\alpha_t, y_{t+1} \sim N(a, b^2),$$

where

$$b^2 = \frac{\sigma^2(\beta_0 + \beta_1\alpha_t^2)}{\beta_0 + \beta_1\alpha_t^2 + \sigma^2}, \qquad a = b^2\frac{y_{t+1}}{\sigma^2}.$$

As far as we know no likelihood methods exist in the literature for the analysis of this type of model (and its various generalizations) although a number of very good approximations have been suggested.

2·1 *Example 2: Probit/Tobit models*

$$\begin{aligned} y_t &= \begin{cases} 1, & \alpha_t > 0 \\ 0, & \alpha_t < 0 \end{cases} \\ \alpha_{t+1} &= \mu + \phi(\alpha_t - \mu) + \eta_t, \quad \eta_t \sim N(0, \sigma_\eta^2). \end{aligned}$$

We have marginally,

$$Pr(y_t = 1) = \Phi\left(\frac{\mu}{\sigma}\right)$$

where $\sigma^2 = \frac{\sigma_\eta^2}{1-\phi^2}$. This can be fully adapted as if $y_{t+1} = 1$,

$$\begin{aligned} \Pr(y_{t+1}|\alpha_t) &= \Phi\left(\frac{\mu + \phi(\alpha_t - \mu)}{\sigma_\eta}\right), \\ f(\alpha_{t+1}|y_{t+1}, \alpha_t) &= \mathrm{TN}_{>0}(\mu + \phi(\alpha_t - \mu); \sigma_\eta^2). \end{aligned}$$

if $y_{t+1} = 0$,

$$\begin{aligned} \Pr(y_{t+1}|\alpha_t) &= 1 - \Phi\left(\frac{\mu + \phi(\alpha_t - \mu)}{\sigma_\eta}\right), \\ f(\alpha_{t+1}|y_{t+1}, \alpha_t) &= \mathrm{TN}_{<0}(\mu + \phi(\alpha_t - \mu); \sigma_\eta^2). \end{aligned}$$

2·2 *Example 3: GARCH(1,1) +error (partial adaption)*

We can write this model in the following form,

$$
\begin{aligned}
y_t|\alpha_t &\sim N(\alpha_t, \sigma^2), \\
\alpha_t|\sigma_t^2 &\sim N(0, \sigma_t^2) \\
\sigma_{t+1}^2 &= \beta_0 + \beta_1\alpha_t^2 + \beta_2\sigma_t^2.
\end{aligned}
$$

We can equivalently write the above model as

$$
\begin{aligned}
y_t|\sigma_t^2 &\sim N(0, \sigma^2 + \sigma_t^2), \\
\alpha_t|\sigma_t^2, y_t &\sim N\left(\frac{b^2 y_t}{\sigma^2}; b^2\right), \\
\sigma_{t+1}^2 &= \beta_0 + \beta_1\alpha_t^2 + \beta_2\sigma_t^2,
\end{aligned}
$$

where $b^2 = \frac{\sigma^2 \sigma_t^2}{\sigma^2 + \sigma_t^2}$.

This is may be thought of as the "semi-adaptable" form of the model.

- $f(\sigma_{t+1}^2 | Y_t)$. Via $f(\alpha_t | y_t, \sigma_t^2)$, given above, R times, then we have R samples $\sigma_{t+1}^{2(i)}$, $i = 1, ..., M$ from $f(\sigma_{t+1}^2 | Y_t)$.

- Regarding these as being sorted in ascending order we now apply the smooth bootstrap method where we have weights $\omega_i = f(y_{t+1} | \sigma_{t+1}^{2(i)})$, $i = 1, ..., M$.

We illustrate this method by estimating the four parameters. We simulate a time series of length 500 and perform 100 different ML estimation procedures using the above method. The four parameters $(\beta_0, \beta_1, \beta_2, \sigma)'$ are set to $(0.01, 0.2, 0.75, 0.1)'$ in the single simulation. The procedure is then run 100 times with $M = 500$, $R = 600$. The results are shown in the table beneath:

– Variance of the simulated maximum likelihood are many hundreds of times smaller than the variance obtained by inverting the matrix of second derivatives at the mode.

—True values of the parameters lie well within thier 95% condidence limits.

—This suggests our approach is a fast, simple and reliable procedure for a problem for which a likelihood solution is, in general, non-trivial.

— likelihood ratio tests can be routinely undertaken.

GARCH plus noise model: $M = 500$, $R = 600$, $SIM = 100$, $n = 500$.					
$\overline{ML_S}(\beta_0, \beta_1, \beta_2, \sigma)'$	$\overline{Var} \times 10^4$				$SD(ML_S)$
0.01080	0.23094	-1.3474	0.10968	-2.4609	0.0002681
0.22563	-1.3474	49.540	-38.952	44.288	0.0039763
0.71723	0.10968	-38.952	39.018	-34.417	0.0028672
0.14436	-2.4609	44.288	-34.417	70.991	0.0060815

Table 3: *Monte Carlo results for GARCH + error model.*

GARCH $(\beta_0, \beta_1, \beta_2)'$		01 02 81 –> 12 29 82, T=500.			
Par	ML	Var $\times 10^{-4}$	Log-lik $= -548.5982$.		
β_0	0.00602	0.25517	0.16912	-0.65797	
β_1	0.02703	0.16912	1.1312	-1.3720	
β_2	0.96220	-0.65797	-1.3766	2.6044	
GARCH+error $(\beta_0, \beta_1, \beta_2, \sigma)'$		M=3000, 4000. Log-lik $= -545.2613$.			
β_0	0.0006446	0.0043743	-0.014907	0.00095946	-0.076571
β_1	0.12874	-0.014907	7.7253	-7.6715	2.3985
β_2	0.86911	0.00095946	-7.6715	7.6919	-2.1834
σ	0.55315	-0.076571	2.3985	-2.1834	7.2245

Table 4: *Estimation results for the firearms homcide dataset.*

Levy processes in continuous time

Barndorff-Nielsen, Shephard(2001). Continous representations of marginal processes

Without loss of generality, we restrict ourselves to the $\Gamma(\nu, 1)$ marginal.

We have EXPLICIT process: Walker (2000),

$$\sigma^2(t) \;=\; \exp(-\lambda t)\sigma^2(0) + \exp(-\lambda t)\varepsilon(\lambda t),$$

where $\varepsilon(\lambda t)$ an independent rv,

$$\varepsilon(\lambda t) \;\sim\; \mathrm{Ga}(z, 1), \quad z \;\sim\; \mathrm{PoGa}\left(\nu, \tfrac{1}{\exp(\lambda t)-1}\right),$$

We obtain $\Pr(\varepsilon(\lambda t) = 0) = \exp(-\nu\lambda t)$. The conditional density of $\varepsilon(\lambda t)$, given that it is greater than 0, is

$$f_{\varepsilon|\varepsilon>0}(x) = \frac{1}{1-\exp(-\nu\lambda t)} \sum_{z=1}^{\infty} \mathrm{Ga}(x|z;1)\mathrm{PoGa}\left(z|\nu;1/(\exp(\lambda t)-1)\right).$$

We can illustrate by simulating a dataset of size 1550, $M = 3000, R = 4200$.

$$y_n|\sigma_n^2 \sim N\left(0, \frac{\sigma_n^2}{b}\right).$$

We take a unit sampling interval, $b = 1$, $\nu = 3$ and $\lambda = 0.02$.

References

Bollerslev, T., Engle, R. F., & Nelson, D. B. (1994). ARCH models. In Engle, R. F. & McFadden, D., editors, *The Handbook of Econometrics, Volume 4*, pages 2959–3038. North-Holland.

Carlin, B. P., Polson, N. G., & Stoffer, D. (1992). A Monte Carlo approach to nonnormal and nonlinear state-space modelling. *J. American Statistical Association* **87**, 493–500.

Diebold, F. X. & Nerlove, M. (1989). The dynamics of exchange rate volatility: a multivariate latent factor ARCH model. *J. Applied Econometrics* **4**, 1–21.

Gourieroux, C. a. A. M. & Renault, E. (1993). Indirect inference. *J. Applied Econometrics* **8**, S85–S118. Reprinted as pp. 101–34 in van Dijk, H.K., A. Monfort and B.W. Brown(1995), *Econometric Inference Using Simulation Techniques*, Chichester: Wiley.

Harvey, A. C., Ruiz=, E., & Sentana, E. (1992). Unobserved component time series models with ARCH disturbances. *J. Econometrics* **52**, 129–158.

JACQUIER, E., POLSON, N. G., & ROSSI, P. E. (1994). Bayesian analysis of stochastic volatility models (with= discussion). *J. Business and Economic Statist.* **12**, 371–417.

KING, M., SENTANA=, E., & WADHWANI, S. (1994). Volatility and links between national stock markets. *Econometrica* **62**, 901–933.

SHEPHARD, N. (1996). Statistical aspects of ARCH and stochastic volatility. In Cox, D. R., Hinkley, D. V., & Barndorff-Nielson, O. E., editors, *Time Series Models in Econometrics, Finance and Other Fields*, pages 1–67. Chapman & Hall, London.

SMITH, A. F. M. & GELFAND, A. E. (1992). Bayesian statistics without tears: a sampling-resampling= perspective. *American Statistican* **46**, 84–88.

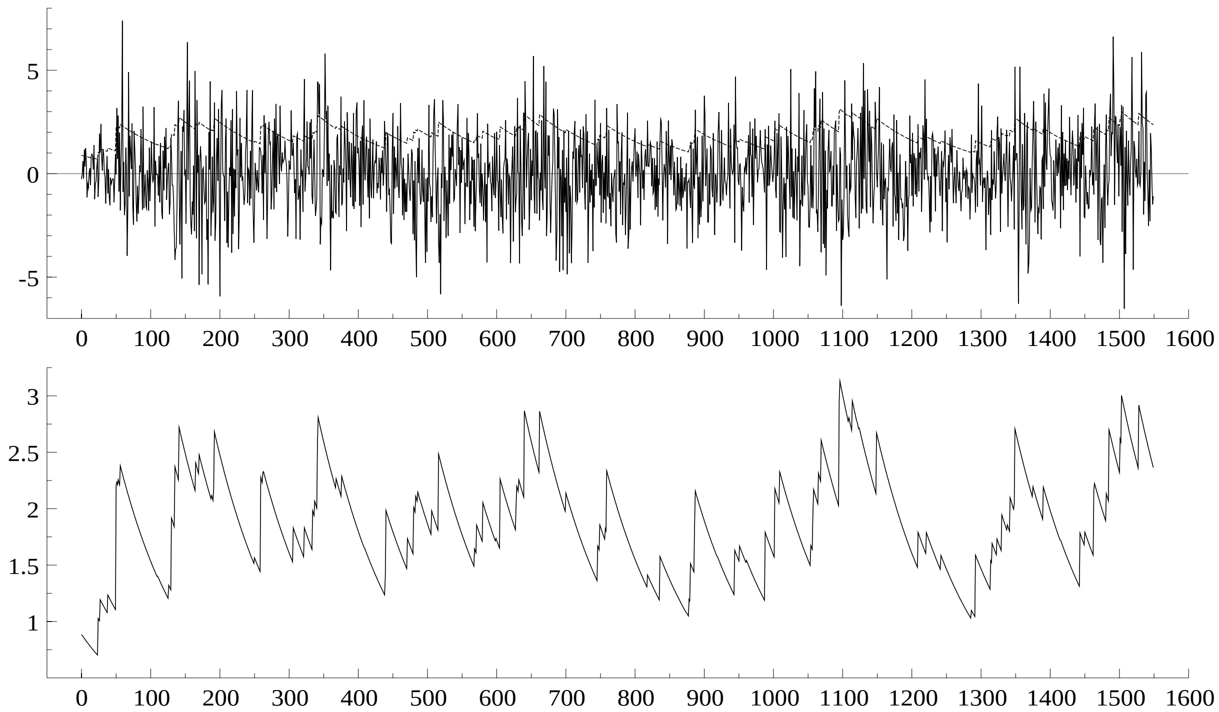

Figure 8: SD and process of returns.

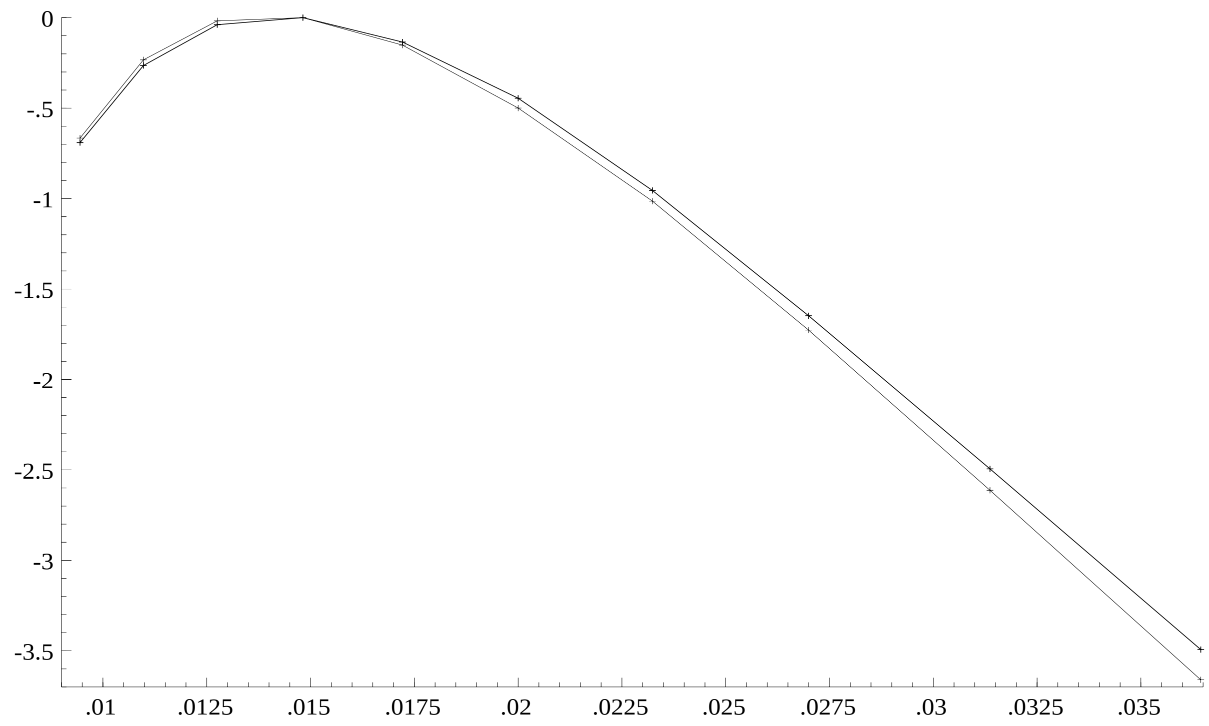

Figure 9: *Lambda, profile plot.*

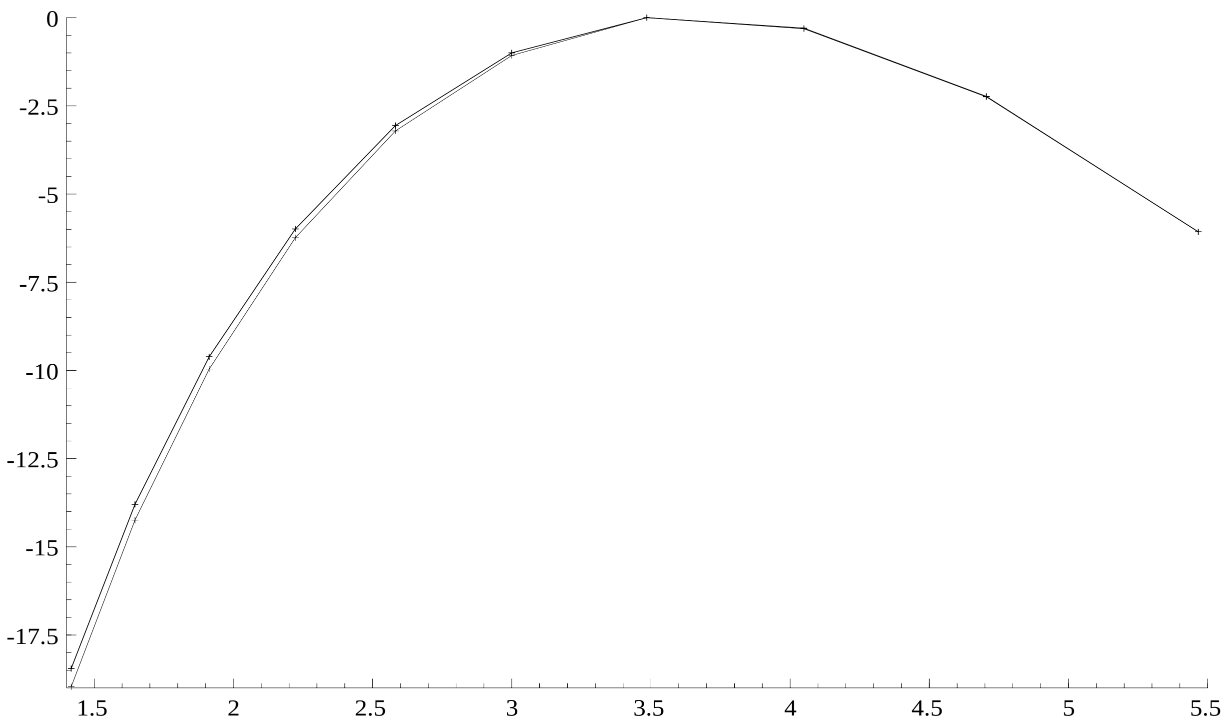

Figure 10: a, *profile plot.*

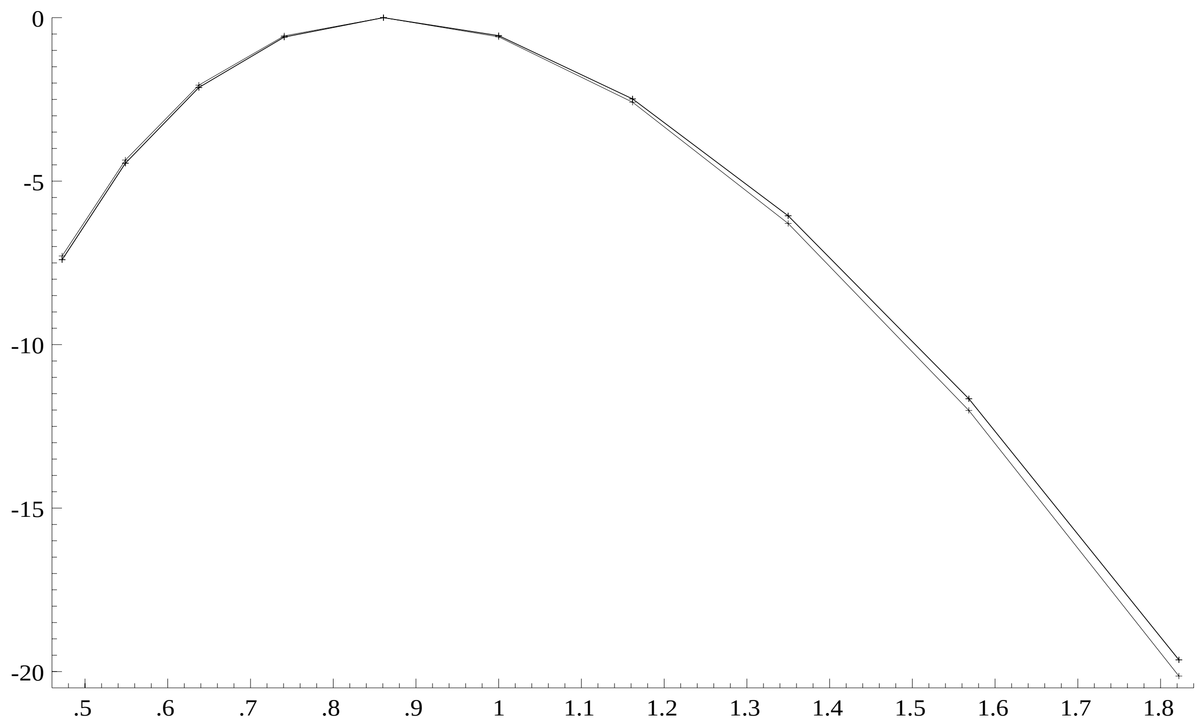

Figure 11: *b, profile plot.*