Econ2146: Financial econometrics
Background material 3: Sequential learning

version 0.2

Neil Shephard

Economics & Statistics Departments,
Harvard University

March 2015

Literature: Sequential learning. Linear case

- Harvey (1989),
  - Monograph on Kalman filtering. Harvey (1993) more gentle introduction.
- West and Harrison (1989)
  - Connections with Bayesian work.
- Whittle (1983b). Classic book on signal extraction. Control books:
  - Whittle (1982), Whittle (1983a), Whittle (1990) and Whittle (1996)
- Durbin and Koopman (2012)
  - Monograph introducing Kalman filtering and some non-Gaussian systems.
- Diebold and Rudebusch (2013). Modelling bond yields.
- Hansen and Sargent (2007)
  - Risk sensitive control to macro problems, applying Whittle's books.
- Hansen and Sargent (2014)
  - Introduction to macro models with control & state space models.

Non-linear case I

- Creal (2012)
  - Introduction to sequential Monte Carlo, mixing theory with empirical illustration. Great place for economics students to start learning scope and basic skills.
- Douc et al. (2014)
  - Strong time series book on non-linear models.
- Doucet and Johansen (2011),
  - high modern quality review
- Liu (2001)
  - Strong on sequential Monte Carlo estimation.
- Johannes and Polson (2009)
  - review of some finance work
- Doucet et al. (2001)

Non-linear case II

- Sequential Monte Carlo, collection of early review articles.

- Johannes et al. (2009)
  - use of methods on SV with jumps
- Andrieu et al. (2010)
  - sequential Monte Carlo and estimated likelihoods
- Crisan and Rozovskii (2011)
  - book of reviews on non-linear filtering

Time series and prediction decomposition

- Think about a time series

$$y_1, y_2, \ldots, y_T.$$

- Natural filtration is $\mathcal{F}_t$, info up to and including time t.
- Then

$$\begin{aligned} f(y_1, y_2, \ldots, y_T|\mathcal{F}_0) &= f(y_1|\mathcal{F}_0)f(y_2, \ldots, y_T|\mathcal{F}_1) \\ &= \prod_{t=1}^{T} f(y_t|\mathcal{F}_{t-1}). \end{aligned}$$

- Call the prediction decomposition.
  - Prediction plays the same role as independence in time series.
- Likewise

$$\log f(y_1, y_2, \ldots, y_T|\mathcal{F}_0) = \sum_{t=1}^{T} \log f(y_t|\mathcal{F}_{t-1}).$$

Time series and likelihood I

- So if the model is indexed by parameters, the log-likelihood is

$$\log L(\theta) = \sum_{t=1}^{T} \log f(y_t | \mathcal{F}_{t-1}, \theta).$$

- Maximum likelihood (ML) estimator is

$$\begin{aligned}
\widehat{\theta} &= \arg\max_{\theta \in \Theta} L(\theta) \\
&= \arg\max_{\theta \in \Theta} \log L(\theta) = \arg\max_{\theta \in \Theta} \frac{1}{T} \sum_{t=1}^{T} \log f(y_t | \mathcal{F}_{t-1}, \theta).
\end{aligned}$$

Time series and likelihood II

- Assume the unconditional expectation $\mathrm{E}\left\{\log f(y_t|\mathcal{F}_{t-1},\theta)\right\}$ exists uniformly over $\theta \in \Theta$. We write

$$\theta^* = \arg\max \mathrm{E}\left\{\log f(y_t|\mathcal{F}_{t-1},\theta)\right\},$$

  noting this is an unconditional expectation.
  - θ^* is called the "pseudo-true" value in econometrics.
  - For correctly specified models θ^* is the "true value," generating the data. Clarify in a moment.
  - If $\log f(y_t|\mathcal{F}_{t-1},\theta)$ is uniformly (over $\theta \in \Theta$) ergodic then $\widehat{\theta} \xrightarrow{p} \theta^*$ as $T \to \infty$.

Time series and likelihood I

$$\theta^* = \arg\max \mathrm{E}\left\{\log f(y_t|\mathcal{F}_{t-1},\theta)\right\}, \tag{1}$$

- From this typically then (range of suppose of the data does not depend upon θ)

$$\left.\frac{\partial \mathrm{E}\left\{\log f(y_t|\mathcal{F}_{t-1},\theta)\right\}}{\partial\theta}\right|_{\theta=\theta^*} = \mathrm{E}\left\{\left.\frac{\partial \log f(y_t|\mathcal{F}_{t-1},\theta)}{\partial\theta}\right|_{\theta=\theta^*}\right\} = 0.$$

Score has a zero unconditional expectation at θ^*.

Time series and likelihood II

- If the model is correct, i.e. data is generated as $f(y_t|\mathcal{F}_{t-1},\theta^*)$, then

$$
\begin{aligned}
& \mathrm{E}\left\{\frac{\partial \log f(y_t|\mathcal{F}_{t-1},\theta)}{\partial\theta}|\mathcal{F}_{t-1}\right\} \\
= & \int \frac{\partial \log f(y_t|\mathcal{F}_{t-1},\theta)}{\partial\theta} f(y_t|\mathcal{F}_{t-1},\theta^*)\mathrm{d}y_t \\
= & \int \frac{1}{f(y_t|\mathcal{F}_{t-1},\theta)} \frac{\partial f(y_t|\mathcal{F}_{t-1},\theta)}{\partial\theta} f(y_t|\mathcal{F}_{t-1},\theta^*)\mathrm{d}y_t.
\end{aligned}
$$

- So

$$
\begin{aligned}
\mathrm{E}\left\{\frac{\partial \log f(y_t|\mathcal{F}_{t-1},\theta^*)}{\partial\theta}|\mathcal{F}_{t-1}\right\} &= \int \frac{\partial f(y_t|\mathcal{F}_{t-1},\theta^*)}{\partial\theta}\mathrm{d}y_t \qquad (2) \\
&= 0,
\end{aligned}
$$

a vector of 0s (assumes support does not depend upon θ).

Time series and likelihood III

- Much tighter:

$$\theta^* = \arg\max \mathrm{E}\left\{\log f(y_t|\mathcal{F}_{t-1}, \theta)|\mathcal{F}_{t-1}\right\}.$$

- So identification is being generated automatically each time period.

The score and time series I

- For time series scores always sum up

$$\frac{\partial \log L(\theta)}{\partial \theta} = \sum_{t=1}^{T} \frac{\partial \log f(y_t|\mathcal{F}_{t-1}, \theta)}{\partial \theta}.$$

- But the individual scores

$$\left.\frac{\partial \log f(y_t|\mathcal{F}_{t-1}, \theta)}{\partial \theta}\right|_{\theta=\theta^*}$$

is a MD seq wrt $\mathcal{F}_t$ due to (2) if the model is correct & so

$$\begin{aligned} M_t &= \sum_{s=1}^{t} \frac{\partial \log f(y_s|\mathcal{F}_{s-1}, \theta^*)}{\partial \theta}, \quad t = 1, 2, \ldots \\ &= M_{t-1} + \frac{\partial \log f(y_t|\mathcal{F}_{t-1}, \theta^*)}{\partial \theta}, \end{aligned}$$

is a martingale wrt $\mathcal{F}_t$, with $M_0 = 0$.

The score and time series II

- Some incorrect models also have MD scores. Significant simplification. Depends on the details of

$$\mathrm{E}\left\{\frac{\partial \log f(y_t|\mathcal{F}_{t-1},\theta)}{\partial\theta}|\mathcal{F}_{t-1}\right\}.$$

  e.g. some AR, ARCH models.

Time series and likelihood

$$\frac{\partial \log L(\theta^*)}{\partial \theta} = M_T = \sum_{s=1}^{T} \frac{\partial \log f(y_t|\mathcal{F}_{t-1},\theta^*)}{\partial \theta}$$

- So if they exist

$$\begin{aligned}
\mathrm{Var}\left(\frac{\partial \log L(\theta^*)}{\partial \theta}|\mathcal{F}_0\right) &= \mathrm{Var}\left(M_T|\mathcal{F}_0\right) = \sum_{t=1}^{T} \mathrm{Var}\left(M_t - M_{t-1}|\mathcal{F}_0\right) \\
&= \sum_{t=1}^{T} \mathrm{Var}\left(\frac{\partial \log f(y_t|\mathcal{F}_{t-1},\theta)}{\partial \theta}|\mathcal{F}_0\right) \\
&= \sum_{t=1}^{T} \mathrm{E}\left[\mathrm{Var}\left(\frac{\partial \log f(y_t|\mathcal{F}_{t-1},\theta)}{\partial \theta}|\mathcal{F}_{t-1}\right)|\mathcal{F}_0\right]
\end{aligned}$$

Time series and likelihood

- Or, randomizing over initial conditions,

$$
\begin{aligned}
\mathrm{Var}\left(\frac{\partial \log L(\theta^*)}{\partial\theta}\right) &= \sum_{t=1}^{T} \mathrm{Var}\left\{\frac{\partial \log f(y_t|\mathcal{F}_{t-1},\theta^*)}{\partial\theta}\right\} \\
&= \sum_{t=1}^{T} \mathrm{E}\left\{\mathrm{Var}\left(\frac{\partial \log f(y_t|\mathcal{F}_{t-1},\theta)}{\partial\theta}|\mathcal{F}_{t-1}\right)\right\}.
\end{aligned}
$$

Time series and likelihood I

- If the model is correct then this extends to the information equality

$$\mathrm{Var}\left(\frac{\partial \log f(y_t|\mathcal{F}_{t-1},\theta^*)}{\partial\theta}|\mathcal{F}_{t-1}\right) = -\mathrm{E}\left(\frac{\partial^2 \log f(y_t|\mathcal{F}_{t-1},\theta^*)}{\partial\theta\partial\theta'}|\mathcal{F}_{t-1}\right).$$

- Hence

$$\mathrm{Var}\left(\frac{\partial \log L(\theta^*)}{\partial\theta}|\mathcal{F}_0\right) = -\sum_{t=1}^{T}\mathrm{E}\left\{\frac{\partial^2 \log f(y_t|\mathcal{F}_{t-1},\theta^*)}{\partial\theta\partial\theta'}|\mathcal{F}_0\right\},$$

and

$$\mathrm{Var}\left(\frac{\partial \log L(\theta^*)}{\partial\theta}\right) = -\sum_{t=1}^{T}\mathrm{E}\left\{\frac{\partial^2 \log f(y_t|\mathcal{F}_{t-1},\theta^*)}{\partial\theta\partial\theta'}\right\}.$$

- Obviously all models are wrong.

Time series overview I

There are basically two ways to build time series models

- Specify a predictive model

$$y_t|\mathcal{F}_{t-1}^{y}$$

  explicitly, e.g. AR, MA, ARCH, etc. Sometimes called in econ "reduced form models" or in stats "observation drive models".

- "State space models" specify the conditional random variables

$$y_t|\alpha_t, \mathcal{F}_{t-1}^{y,\alpha} = y_t|\alpha_t, \mathcal{F}_{t-1}^{y}, \quad \alpha_{t+1}|\alpha_t, y_t, \mathcal{F}_{t-1}^{y,\alpha} = \alpha_{t+1}|\alpha_t, \mathcal{F}_{t-1}^{y},$$

  where $\mathcal{F}_{t-1}^{y}$ does not include the history of the latent α_t. Sometimes called parameter driven models.

Time series overview II

- Typically the simpler version

$$y_t|\alpha_t, \quad \alpha_{t+1}|\alpha_t$$

  is used. These models are often called a "Hidden Markov model" (HMM).
  - I use the names interchangeably.

State space definition I

$$y_t|\alpha_t, \quad \alpha_{t+1}|\alpha_t$$

Typically call the first "measure equation", second "transition equation".

- Statistical specification: in terms of measures,

  $$\mu_t(A) = \Pr\left(y_t \in A|\alpha_t\right), \quad \gamma_t(A) = \Pr\left(\alpha_{t+1} \in A|\alpha_t\right),$$

  or in terms of densities for continuous random variables

  $$f(y_t|\alpha_t), \quad f(\alpha_{t+1}|\alpha_t),$$

  or for discrete random variables

  $$\Pr(y_t|\alpha_t), \quad \Pr(\alpha_{t+1}|\alpha_t).$$

State space definition II

- In some discussions the model is written as dynamic equations

$$y_t = h_t(\alpha_t, u_t), \quad \alpha_{t+1} = g_t(\alpha_t, v_t),$$

  where u_t, v_t are independent (multivariate) uniforms.
- We will work with the statistical versions.

Example

Suppose

$$y_t|\alpha_t \sim N(\alpha_t, \sigma_\varepsilon^2), \quad \alpha_{t+1}|\alpha_t \sim N(\alpha_t, \sigma_\eta^2).$$

Called a "random walk plus noise model" or "local level model". Think of data $\{y_s\}_{s=1,2,\ldots}$ as imperfect measure of earnings growth $\{\alpha_s\}_{s=1,2,\ldots}$.

State space definition III

Example

Suppose

$$y_t|\alpha_t \sim N(0, \exp(\alpha_t)), \quad \alpha_{t+1}|\alpha_t \sim N(\mu + \phi\,(\alpha_t - \mu)\,, \sigma_\eta^2),$$

this is a discrete time log-normal stochastic volatility model. Discrete version of the Hull and White (1987) continuous time model.

Filtering, smoothing and prediction

Now think about information sets: write $\mathcal{F}_t$ as the history of $\{y_s\}_{s=1,2,...}$ up to time t.

- We learn, or "filter",

$$\alpha_t | \mathcal{F}_t,$$

  or forecast

$$\alpha_{t+s} | \mathcal{F}_t, \quad s = 1, 2, ...,$$

  or "smooth"

$$\alpha_t | \mathcal{F}_T, \quad t < T.$$

Bayesian updating I

$$f(y_t|\alpha_t), \quad f(\alpha_{t+1}|\alpha_t),$$

Assume we know

$$f\left(\alpha_t|\mathcal{F}_t\right).$$

- Then we carry out Bayesian updating as

$$\begin{aligned} f\left(\alpha_{t+1}|\mathcal{F}_t\right) &= \int f(\alpha_{t+1}|\alpha_t)\mathrm{d}F\left(\alpha_t|\mathcal{F}_t\right) \\ f\left(\alpha_{t+1}|\mathcal{F}_{t+1}\right) &\propto f(y_{t+1}|\alpha_{t+1})f\left(\alpha_{t+1}|\mathcal{F}_t\right) \\ f(y_{t+1}|\mathcal{F}_t) &= \int f(y_{t+1}|\alpha_{t+1})\mathrm{d}F\left(\alpha_{t+1}|\mathcal{F}_t\right). \end{aligned}$$

- This is trivial to state. Typically need numerical methods to compute.
- Use method to process through time, updating 1 period at a time. Initialized at some

$$f\left(\alpha_1|\mathcal{F}_0\right).$$

Smoothing

Also in economics we see the use of smoothing,

$$\alpha_1, ..., \alpha_T | \mathcal{F}_T.$$

Example

The Hodrick and Prescott (1997) filter used in macro to detrend GDP etc is actually a smoother based on the model $\alpha_t = (\beta_t, \beta_{t-1})$

$$y_t|\alpha_t \sim N(\beta_t, \sigma^2_\varepsilon), \quad \beta_{t+1}|\alpha_t \sim N(\beta_t + \left(\beta_t - \beta_{t-1}\right), \lambda\sigma^2_\varepsilon).$$

In statistics this is called a cubic spline model and has a very long history.

Stochastic volatility and local level model

Example

Stock and Watson (2007) use

$$\begin{aligned} y_t|\sigma^2_{t\varepsilon},\sigma^2_{t\eta},\beta_t &\sim N(\beta_t,\sigma^2_{t\varepsilon}), \quad \beta_{t+1}|\sigma^2_{t\varepsilon},\sigma^2_{t\eta},\beta_t \sim N(\beta_t,\sigma^2_{t\eta}), \\ \log\sigma^2_{t+1\varepsilon}|\sigma^2_{t\varepsilon},\sigma^2_{t\eta},\beta_t &\sim N(\log\sigma^2_{t\varepsilon},\omega^2_{\varepsilon}), \\ \log\sigma^2_{t+1\eta}|\sigma^2_{t\varepsilon},\sigma^2_{t\eta},\beta_t &\sim N(\log\sigma^2_{t\eta},\omega^2_{\eta}), \end{aligned}$$

to study how predictable inflation is over the short run. Here $\alpha_t = \left(\sigma^2_{t\varepsilon},\sigma^2_{t\eta},\beta_t\right)'$. The $\sigma^2_{t\varepsilon},\sigma^2_{t\eta}$ processes are stochastic vol. They report smoothing results the volatility processes.

A yield curve model I

- Nelson and Siegel (1987) interpolate yield curve using surface

$$c(\tau) = \beta_0 + \beta_1 \left(\frac{1 - e^{-\lambda\tau}}{\lambda t} \right) + \beta_2 \left(\frac{1 - e^{-\lambda\tau}}{\lambda t} - e^{-\lambda\tau} \right), \quad \tau, \lambda > 0,$$

  where τ notes maturity. $\beta_0, \beta_1, \beta_2, \lambda$ are parameters.

- Components
  - 1, a level
  - $\frac{1-e^{-\lambda\tau}}{\lambda t}$ starts at 1, decays to 0, called a slope.
  - $\frac{1-e^{-\lambda\tau}}{\lambda t} - e^{-\lambda\tau}$ starts at 0, goes up and then decays, called a curvature.

- Diebold and Rudebusch (2013) allow

$$c_t(\tau) = L_t + S_t \left(\frac{1 - e^{-\lambda\tau}}{\lambda t} \right) + C_t \left(\frac{1 - e^{-\lambda\tau}}{\lambda t} - e^{-\lambda\tau} \right),$$

  where $\{L_t, S_t, C_t\}$ follow a stochastic process, e.g. a Gaussian stationary VAR.

A yield curve model II

- Then $\alpha_t = (L_t, S_t, C_t)'$, and using yields from zero coupon bonds $(y_{1t}, ..., y_{dt})$ at maturities $(\tau_1, ..., \tau_d)$, we have a model

$$
\begin{aligned}
y_t &= Z_t \alpha_t + \varepsilon_t, \\
Z_t &= \begin{pmatrix} z_{1t} \\ z_{2t} \\ z_{dt} \end{pmatrix}, z_{it} = \left(1, \frac{1 - e^{-\lambda \tau_i}}{\lambda \tau_i}, \frac{1 - e^{-\lambda \tau_i}}{\lambda \tau_i} - e^{-\lambda \tau_i}\right).
\end{aligned}
$$

- Diebold and Rudebusch (2013) show how to impose no arbitrage on this model.

Core issue with smoothing I

- Store results from the filter $f(\alpha_t|\mathcal{F}_T)$, $t = 1, 2, ..., T$. Have $f(\alpha_T|\mathcal{F}_T)$.
- Then, noting

$$\begin{aligned} f(\alpha_{t-1}|\mathcal{F}_T) &= \int f(\alpha_{t-1}, \alpha_t|\mathcal{F}_T)\mathrm{d}\alpha_t \\ &= \int f(\alpha_t|\mathcal{F}_T) f(\alpha_{t-1}|\alpha_t, \mathcal{F}_T)\mathrm{d}\alpha_t. \end{aligned}$$

- Smooth working backwards: $t = T, T-1, T-2, ...$

Core issue with smoothing II

- The remaining task is

$$
\begin{aligned}
f(\alpha_t|\alpha_{t+1},\mathcal{F}_T) &= f(\alpha_t|\alpha_{t+1},\mathcal{F}_t)\\
&= \frac{f(\alpha_t|\mathcal{F}_t)f(\alpha_{t+1}|\alpha_t)}{\int f(\alpha_t|\mathcal{F}_t)f(\alpha_{t+1}|\alpha_t)\mathrm{d}\alpha_t}\\
&\propto f(\alpha_t|\mathcal{F}_t)f(\alpha_{t+1}|\alpha_t). \qquad (3)
\end{aligned}
$$

Hence the need to store the filtering density.

Smoothing density

- In modern time series, it is useful to know the joint smoothing density. This is

$$
\begin{aligned}
f(\alpha_1, ..., \alpha_T | \mathcal{F}_T) &= f(\alpha_T | \mathcal{F}_T) f(\alpha_1, ..., \alpha_{T-1} | \mathcal{F}_T) \\
&= f(\alpha_T | \mathcal{F}_T) \prod_{t=1}^{T-1} f(\alpha_t | \alpha_{t+1}, \mathcal{F}_T) \\
&= f(\alpha_T | \mathcal{F}_T) \prod_{t=1}^{T-1} f(\alpha_t | \alpha_{t+1}, \mathcal{F}_t).
\end{aligned}
$$

- This result goes back to Carter and Kohn (1994), Frühwirth-Schnatter (1994). See also de Jong and Shephard (1995) and Durbin and Koopman (2002).
- "Smoothing density backward decomposition".

General issues

$$
\begin{aligned}
f(\alpha_{t+1}|\mathcal{F}_t) &= \int f(\alpha_{t+1}|\alpha_t)\mathrm{d}F(\alpha_t|\mathcal{F}_t) \\
f(\alpha_{t+1}|\mathcal{F}_{t+1}) &\propto f(y_{t+1}|\alpha_{t+1})f(\alpha_{t+1}|\mathcal{F}_t) \\
f(y_{t+1}|\mathcal{F}_t) &= \int f(y_{t+1}|\alpha_{t+1})\mathrm{d}F(\alpha_{t+1}|\mathcal{F}_t).
\end{aligned}
$$

- In few cases can these integrals be solved. The cases I know of are
  - Gaussian case: Kalman filter
  - α_t having finite support: Baum filter (in economics this is usually called the Hamilton filter)
  - A small number of gamma/beta transition equations. See the work on stochastic volatility of Shephard (1994) and the multivariate generalization in Uhlig (1997).

Linear Gaussian state space

$$y_t|\alpha_t, \quad \alpha_{t+1}|\alpha_t, \quad t = 1, 2, ..., T.$$

- Gaussian linear state space model is

$$\begin{aligned} y_t &= Z_t\alpha_t + G_{\varepsilon t}\varepsilon_t, \quad \alpha_{t+1} = T_t\alpha_t + H_{\varepsilon t}\eta_t, \\ \varepsilon_t &\sim N(0, I), \quad \eta_t \sim N(0, I), \\ \Sigma_{\varepsilon t} &= G_{\varepsilon t}G'_{\varepsilon t}, \quad \Sigma_{\eta t} = H_{\varepsilon t}H'_{\varepsilon t}, \\ \alpha_1|\mathcal{F}_0 &\sim N(a_{1|0}, P_{1|0}), \end{aligned}$$

  while we take $(\varepsilon_s \perp\!\!\!\perp \eta_t)$ for all t, s.
- Here $(Z_t, \Sigma_{\varepsilon t})$ are known (up to come parameters) given $\mathcal{F}_{t-1}$ (it is adapted to $\mathcal{F}_{t-1}$), while $(T_t, \Sigma_{\eta t})$ is known given $\mathcal{F}_t$.
- Most models have $(Z_t, \Sigma_{\varepsilon t}, T_t, \Sigma_{\eta t})$ being sparse indexed by some parameters and not depending upon $\mathcal{F}_t$.

ARMA models and state space representation

- Markov assumption on the state looks restrictive. But can expand the dimension.

Examples

$y_t = \phi y_{t-1} + u_t + \kappa u_{t-1}$, $u_t \sim N(0, \sigma^2)$, so let

$$\begin{aligned} y_t &= (1, 0)\, \alpha_t \\ \alpha_{t+1} &= \begin{pmatrix} \phi & 1 \\ 0 & 0 \end{pmatrix} \alpha_t + \begin{pmatrix} 1 \\ \kappa \end{pmatrix} \sigma \eta_t, \quad \eta_t \overset{iid}{\sim} N(0,1). \end{aligned}$$

So $\alpha_{2t} = \sigma \kappa \eta_{t-1}$,

$$\begin{aligned} y_t &= \alpha_{1t} = \phi \alpha_{1t-1} + \sigma \eta_{t-1} + \alpha_{2t-1} \\ &= \phi \alpha_{1t-1} + \sigma \eta_{t-1} + \sigma \kappa \eta_{t-2}, \end{aligned}$$

as required.

Long memory models I

- Outside this framework are fractional models. But can be approximated by sums of short-memory processes (e.g. Granger (1980), ...).
- Suppose

$$y_t = \sum_{j=1}^{D} \beta_{jt},$$

where β_{jt} are independent short memory processes

$$\beta_{jt} = \phi_j \beta_{jt-1} + \sqrt{1-\phi_j^2}\sigma u_{jt},$$

where $|\phi_j| < 1$, $u_{jt} \overset{iid}{\sim} N(0,1)$.
- Then the stationary distribution is $\beta_{jt} \sim N(0,\sigma^2)$.

Long memory models II

- So

$$\begin{aligned}
\mathrm{Cor}(y_t, y_{t-s}) &= \frac{\sum_{j=1}^{D} \mathrm{Cov}(\beta_{jt}, \beta_{jt-s})}{\sum_{j=1}^{D} \mathrm{Var}(\beta_{jt})} = \frac{\sum_{j=1}^{D} \sigma_j^2 \phi_j^s}{\sum_{j=1}^{D} \sigma_j^2} \\
&= \sum_{j=1}^{D} w_j \left(\phi_j\right)^s, \quad w_j = \sigma_j^2 / \sum_{i=1}^{D} \sigma_i^2, \\
&= \mathrm{E}_M(\phi^s),
\end{aligned}$$

where M is a discrete probability measure $\left\{w_j, \phi_j\right\}_{j=1,2,\ldots,D}$.

- Here

$$\mathrm{E}\left(\phi^s\right) = \int_0^1 \phi^s \mathrm{d}F(\phi).$$

Long memory models III

- Assume $\phi \in [0, 1)$, can write it as $\phi = \exp(-\lambda)$, $\lambda > 0$. Then

$$\mathrm{E}\left(\phi^s\right) = \int_0^1 e^{-\lambda s} \mathrm{d}F(\lambda).$$

- In the gamma case, for example,

$$\begin{aligned} \mathrm{E}\left(\phi^s\right) &= \int_0^1 e^{-\lambda s} \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda} \mathrm{d}\lambda = \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^1 \lambda^{\alpha-1} e^{-(\beta+s)\lambda} \mathrm{d}\lambda \\ &= \beta^\alpha (\beta + s)^{-\alpha} \propto \left(1 + \frac{s}{\beta}\right)^{-\alpha}. \end{aligned}$$

- So for large s,

$$\frac{\log \mathrm{E}\left(\phi^s\right)}{\log s} \to -\alpha, \quad \text{or} \quad \mathrm{E}\left(\phi^s\right) \simeq c \left(\frac{1}{s}\right)^\alpha.$$

Long memory models IV

- That is $\mathrm{E}\left(\phi^{s}\right)$ is hyperbolic if $\alpha \leq 1$, so has long-memory, i.e. $\sum_{s=1}^{\infty} |\mathrm{Cor}(y_t, y_{t-s})| = \infty$.
- But this is easily roughly handled in the state space model as

$$\alpha_t = \left(\beta_{1t}, \beta_{2t}, ..., \beta_{Dt}\right)', \quad Z_t = \iota'.$$

  Do exactly produce the result we need an infinite dimensional state.
- Use in finance includes Calvet and Fisher (2001), Calvet and Fisher (2002) and Calvet and Czellar (2014).

Kalman filtering

- We write

$$\alpha_t|\mathcal{F}_s \sim N(a_{t|s}, P_{t|s}),$$

so trivially

$$y_t|\mathcal{F}_{t-1} \sim N(Z_t a_{t|t-1}, F_t), \quad \text{where} \quad F_t = Z_t P_{t|t-1} Z_t' + \Sigma_{\varepsilon t}.$$

- In the Gaussian case

$$\begin{aligned} f\left(\alpha_{t+1}|\mathcal{F}_t\right) &= \int f(\alpha_{t+1}|\alpha_t)\mathrm{d}F\left(\alpha_t|\mathcal{F}_t\right) \\ f\left(\alpha_{t+1}|\mathcal{F}_{t+1}\right) &\propto f(y_{t+1}|\alpha_{t+1})f\left(\alpha_{t+1}|\mathcal{F}_t\right) \\ f(y_{t+1}|\mathcal{F}_t) &= \int f(y_{t+1}|\alpha_{t+1})\mathrm{d}F\left(\alpha_{t+1}|\mathcal{F}_t\right). \end{aligned}$$

can be solved.

Kalman filter and proof I

Algorithm: Kalman (1960) filter. Assuming $Z_t P_{t|t-1} Z_t' + \Sigma_{\varepsilon t} > 0$ (the LHS is positive definite) then

$$
\begin{aligned}
a_{t|t} &= a_{t|t-1} + P_{t|t-1} Z_t' F_t^{-1} (y_t - Z_t a_{t|t-1}) \\
P_t &= P_{t|t-1} - P_{t|t-1} Z_t' F_t^{-1} Z_t P_{t|t-1} \\
a_{t+1|t} &= T_t a_t, \quad P_{t+1|t} = T_t P_{t|t} T_t' + \Sigma_{\eta t}
\end{aligned}
$$

Proof. The last two lines are trivial results. Now $\alpha_t | \mathcal{F}_{t-1} \sim N(a_{t|t-1}, P_{t|t-1})$, so

$$
\begin{pmatrix} \alpha_t \\ y_t \end{pmatrix} | \mathcal{F}_{t-1} \sim N\left(\begin{pmatrix} a_{t|t-1} \\ Z_t a_{t|t-1} \end{pmatrix}, \begin{pmatrix} P_{t|t-1} & P_{t|t-1} Z_t' \\ Z_t P_{t|t-1} & Z_t P_{t|t-1} Z_t' + \Sigma_{\varepsilon t} \end{pmatrix} \right).
$$

Then recall that

$$
\begin{pmatrix} Y \\ X \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_Y \\ \mu_X \end{pmatrix}, \begin{pmatrix} \Sigma_{YY} & \Sigma_{YX} \\ \Sigma_{XY} & \Sigma_{XX} \end{pmatrix} \right).
$$

Kalman filter and proof II

Then if $\Sigma_{XX} > 0$ then the usual posterior from a regression delivers:

$$Y|X \sim N(\mu_Y + \Sigma_{YX}\Sigma_{XX}^{-1}(X - \mu_X), \Sigma_{YY} - \Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}).$$

Now apply this to the state space problem, we need that $Z_t P_{t|t-1} Z_t' + \Sigma_{\varepsilon t} > 0$. Then

$$\begin{aligned} \alpha_t|\mathcal{F}_t \sim & N(a_{t|t-1} + P_{t|t-1}Z_t'\left(Z_t P_{t|t-1} Z_t' + \Sigma_{\varepsilon t}\right)^{-1}\left(y_t - Z_t a_{t|t-1}\right), \\ & P_{t|t-1} - P_{t|t-1}Z_t'\left(Z_t P_{t|t-1} Z_t' + \Sigma_{\varepsilon t}\right)^{-1} Z_t P_{t|t-1}), \end{aligned}$$

which is the stated result.

Smoothing I

- Recall the notation

$$\alpha_t|\mathcal{F}_T \sim N(a_{t|T}, P_{t|T}).$$

I am going to assume $(Z_t, \Sigma_{\varepsilon t}, T_t, \Sigma_{\eta t})$ are non-stochastic (I do not know how to do smoothing if they are stochastic).

- **Algorithm: The Rauch et al. (1965) moment smoothing (fixed lag)**. Let $R_t^* = P_{t|t} T_t' P_{t+1|t}^{-1}$, then

$$\begin{aligned} a_{t|T} &= a_{t|t} + R_t^* \left(a_{t+1|T} - a_{t+1|t} \right), \quad t = T-1, ..., 1, \\ P_{t|T} &= P_{t|t} + R_t^* \left(P_{t+1|T} - P_{t+1|t} \right) R_t^{*\prime}. \end{aligned} \tag{4}$$

Smoothing II

Proof.
Recall (3) which stated that

$$f(\alpha_t|\alpha_{t+1},\mathcal{F}_T) \propto f(\alpha_t|\mathcal{F}_t)f(\alpha_{t+1}|\alpha_t),$$

while

$$\left(\begin{array}{c}\alpha_t\\ \alpha_{t+1}\end{array}\right)|\mathcal{F}_t \sim N\left(\left(\begin{array}{c}a_{t|t}\\ a_{t+1|t}\end{array}\right),\left(\begin{array}{cc}P_{t|t} & P_{t|t}T_t'\\ T_tP_{t|t} & T_tP_{t|t}T_t'+\Sigma_{\eta t}\end{array}\right)\right),$$

so, if $T_tP_{t|t}T_t'+\Sigma_{\eta t}=P_{t+1|t}>0$, then

$$\begin{aligned}\alpha_t|\alpha_{t+1},\mathcal{F}_T \quad &\sim \quad N(a_{t|t}+P_{t|t}T_t'P_{t+1|t}^{-1}\left(\alpha_{t+1}-a_{t+1|t}\right), \\ &\qquad P_{t|t}-P_{t|t}T_t'P_{t+1|t}^{-1}T_tP_{t|t}).\end{aligned} \tag{5}$$

Then

$$p(\alpha_t|\mathcal{F}_T)=\int f(\alpha_{t+1}|\mathcal{F}_T)f(\alpha_t|\alpha_{t+1},\mathcal{F}_T)\mathrm{d}\alpha_{t+1},$$

Smoothing III

so

$$\begin{aligned}\alpha_t|\mathcal{F}_T \sim\ & N(a_{t|t} + P_{t|t}T_t'P_{t+1|t}^{-1}\left(a_{t+1|T} - a_{t+1|t}\right), \\ & P_{t|t} - P_{t|t}T_t'P_{t+1|t}^{-1}T_tP_{t|t} + P_{t|t}T'P_{t+1|t}^{-1}P_{t+1|T}P_{t+1|t}^{-1}TP_{t|t}).\end{aligned}$$

Thus

$$\begin{aligned}R_t^*T_tP_{t|t} &= P_{t|t}T_t'P_{t+1|t}^{-1}T_tP_{t|t} = P_{t|t}T_t'P_{t+1|t}^{-1}P_{t+1|t}P_{t+1|t}^{-1}T_tP_{t|t} \\ &= R_t^*P_{t+1|t}^{-1}R_t^{*\prime}.\end{aligned}$$

then

$$\begin{aligned}a_{t|T} &= a_{t|t} + R_t^*\left(a_{t+1|T} - a_{t+1|t}\right) \\ P_{t|T} &= P_{t|t} + R_t^*\left(P_{t+1|T} - P_{t+1|t}\right)R_t^{*\prime},\end{aligned}$$

as required.

Smoothing IV

- This is often called the Kalman smoother, but Kalman did not discuss smoothing. This result seems to be due to Rauch et al. (1965), I will use their name.
- There are faster moment smoothers. The state of the art is discussed in Durbin and Koopman (2012). This matters if the state vector is of vector high dimensional (e.g. some factor models).

Simulation smoothing I

- Recall the smoothing density

$$f(\alpha_1, ..., \alpha_T|\mathcal{F}_T) = f(\alpha_T|\mathcal{F}_T) \prod_{t=1}^{T-1} f(\alpha_t|\alpha_{t+1}, \mathcal{F}_t),$$

  and that if $R_t^* = P_{t|t} T_t' P_{t+1|t}^{-1}$ then (5) can be written as

$$\alpha_t|\alpha_{t+1}, \mathcal{F}_T \sim N(a_{t|t} + R_t^* \left(\alpha_{t+1} - a_{t+1|t}\right), \left(I - R_t^* T_t\right) P_{t|t}).$$

- Then we can sample the path

$$\alpha_1, ..., \alpha_T|\mathcal{F}_T$$

  using the simulation smoothing algorithm.
- **The Carter and Kohn (1994) and Frühwirth-Schnatter (1994) simulation smoothing algorithm**:

Simulation smoothing II

- Perform filter. Store: $\left\{a_{t+1|t}, P_{t|t}, R_t^*\right\}_{t=1,2,\ldots,T-1}$. Set $t = T$.
- Simulate $\alpha_t|\mathcal{F}_t$.
- Let $t = t - 1$ and simulate from

$$\alpha_t|\alpha_{t+1}, \mathcal{F}_T \sim N(a_{t|t} + R_t^* \left(\alpha_{t+1} - a_{t+1|t}\right), (I - R_t^* T_t) P_{t|t})$$

  until $t = 1$.

de Jong and Shephard (1995) and Durbin and Koopman (2002) have faster algorithms.

An alternative simulation smoother I

- The Durbin and Koopman (2002) simulation smoothing algorithm is very pretty and fast for repeated use.
- It can be expressed widely: let

$$y = a + Bu, \quad u \sim N(0, I).$$

- Now $u|y \sim N(\widehat{u}, U)$, (e.g. if BB' were invertible then $u|y = N(B' (BB')^{-1} (y - a), I - B' (BB')^{-1} B)$, more generally some generalized inverses are necessary) where $\widehat{u}$ is affine in y and U does not depend upon y.
- Now simulate

$$u^{(1)} \sim N(0, I), \quad \text{calculate} \quad y^{(1)} = a + Bu^{(1)}.$$

An alternative simulation smoother II

- Compute $\widehat{u}^{(1)} = \mathrm{E}(u^{(1)}|y^{(1)})$ and build $\widetilde{u}^{(1)} = \widehat{u} + u^{(1)} - \widehat{u}^{(1)}$. Then

$$\widetilde{u}^{(1)}|y \sim N(\widehat{u}, U) \stackrel{L}{=} u|y.$$

- Proof.

$$\mathrm{E}\left(\widetilde{u}^{(1)}|y\right) = \widehat{u} + \mathrm{E}(u^{(1)}|y) - \mathrm{E}\left(\widehat{u}^{(1)}|y\right) = 0,$$

  and, noting $\widehat{u}$ is known given y,

$$\mathrm{Var}\left(\widetilde{u}^{(1)}|y\right) = \mathrm{E}\left\{\left(u^{(1)} - \widehat{u}^{(1)}\right)\left(u^{(1)} - \widehat{u}^{(1)}\right)'|y\right\} = U.$$

- Remark: This algorithm just needs code to compute $\mathrm{E}(u|y)$ fast. This is just the expectation smoother (4), or faster versions which are available. This algorithm applies more widely than Gaussian linear state space models.

Bayesian inference I

Fearnhead (2011) reviews.

- Two strategies
  - Use Kalman filter to compute

    $$f(\theta|y,\mathcal{F}_0) \propto p(\theta|\mathcal{F}_0) f(y|\theta,\mathcal{F}_0) = p(\theta|\mathcal{F}_0) \prod_{t=1}^{T} f(y_t|\theta,\mathcal{F}_{t-1}),$$

    explore using the H-M algorithm.
  - MCMC sample

    $$f(\alpha,\theta|y,\mathcal{F}_0),$$

    and discard draws α draws.
- Former, nothing new to say.

Bayesian inference II

- Latter, one approach is to sample the block

$$f(\alpha|\theta, y, \mathcal{F}_0)$$

using the simulation smoother and then to make H-M moves for θ using

$$f(\theta|y, \alpha, \mathcal{F}_0) \propto p(\theta|\mathcal{F}_0) f(y|\alpha, \theta, \mathcal{F}_0).$$

Discrete state Markov chain I

- The other large class of applications focuses on the case where $\alpha_t \in S$, a discrete set of support. The most celebrated case is $\alpha_t \in \{0, 1\}$.
- This model is huge in genetics, where α_t is the DNA code, $\alpha_t \in \{A, C, G, T\}$, e.g. Li and Durbin (2011).
- In economics this kind of model is associated with Hamilton (1989). S could represent recession, not recession. High vol, low vol.
- Analyzed by Leonard Baum in the 1960s in a series of papers. Includes Baum-Welch algorithm (EM algorithm for HMM) and Viterbi algorithm (max joint smoothing distribution $\Pr(\alpha_1, \alpha_2, ..., \alpha_T | \mathcal{F}_T)$).

Discrete state Markov chain II

- The discrete case filtering is trivial

$$
\begin{aligned}
\Pr(\alpha_{t+1}|\mathcal{F}_t) &= \sum_{\alpha_t \in S} f(\alpha_{t+1}|\alpha_t)\Pr(\alpha_t|\mathcal{F}_t) \\
\Pr(\alpha_{t+1}|\mathcal{F}_{t+1}) &\propto f(y_{t+1}|\alpha_{t+1})\Pr(\alpha_{t+1}|\mathcal{F}_t) \\
f(y_{t+1}|\mathcal{F}_t) &= \sum_{\alpha_{t+1} \in S} f(y_{t+1}|\alpha_{t+1})\Pr(\alpha_{t+1}|\mathcal{F}_t).
\end{aligned}
$$

Non-linear, non-Gaussian problems I

$$
\begin{aligned}
f(\alpha_{t+1}|\mathcal{F}_t) &= \int f(\alpha_{t+1}|\alpha_t)\mathrm{d}F(\alpha_t|\mathcal{F}_t) \\
f(\alpha_{t+1}|\mathcal{F}_{t+1}) &\propto f(y_{t+1}|\alpha_{t+1})f(\alpha_{t+1}|\mathcal{F}_t) \\
f(y_{t+1}|\mathcal{F}_t) &= \int f(y_{t+1}|\alpha_{t+1})\mathrm{d}F(\alpha_{t+1}|\mathcal{F}_t).
\end{aligned}
$$

- In general these integrals cannot be solved.
- Early lit used numerical integration rules, e.g. Kitagawa (1987). Does not work well on higher dimensional state problems. Dead-end.
- Modern method is "sequential Monte Carlo" or "particle filtering".
- Up to date resources on particle filters is available at `http://www.stats.ox.ac.uk/~doucet/smc_resources.html`
- More theory focused: Doucet and Johansen (2011)
- More economics and finance focused: Creal (2012).

Basic particle filter I

- You can trace an earlier history, but modern subject starts with Gordon et al. (1993).
- Write, generically, the path

$$\alpha_{1:t} = (\alpha_1, ..., \alpha_t)$$

- Idea: we draw a path from

$$\begin{aligned} f(\alpha_{1:t}|\mathcal{F}_t) &= \frac{g(\alpha_{1:t})}{L_t}, \\ g(\alpha_{1:t}) &= f(\alpha_{1:t}) f(y_{1:t}|\alpha_{1:t}), \\ L_t &= \int g(\alpha_{1:t}) \mathrm{d}\alpha_{1:t} = f(y_{1:t}|\mathcal{F}_0) \end{aligned}$$

with a sample of size M,

$$\left\{ \alpha_{1:t}^{(j)} \right\}_{j=1,2,...,M}.$$

Basic particle filter I

- Could do MCMC at each step through time.
- Instead work by induction. Same as Kalman filter but with paths.
- Similar to bootstrap. Use samples to represent distributions.
- Resampling plays a crucial and surprising role.

Importance sampling I

- $\left\{\alpha_{1:t}^{(j)}\right\}_{j=1,2,\ldots,M}$ i.i.d. from

$$f(\alpha_{1:t}|\mathcal{F}_t) = \frac{g(\alpha_{1:t})}{L_t}, \quad g(\alpha_{1:t}) = f(\alpha_{1:t})f(y_{1:t}|\alpha_{1:t}).$$

  noting $L_t = f(y_{1:t}|\mathcal{F}_0)$.
- $g(\alpha_{1:t})$ is known, but sampling is not easy.
- Importance sampling, $\left\{\alpha_{1:t}^{(j)}\right\}_{j=1,2,\ldots,M}$ i.i.d. from $q(\alpha_{1:t}|\mathcal{F}_t)$.
- Then weight is

$$W_{t,M}^{j} = \frac{w_t^j}{\sum_{i=1}^{M} w_t^i}, \quad w_t^j = \frac{g(\alpha_{1:t}^{(j)})}{q(\alpha_{1:t}^{(j)}|\mathcal{F}_t)}.$$

Importance sampling II

- Record the estimated normalizing constant

$$\begin{aligned}\widehat{L}_{t,M} &= \frac{1}{M}\sum_{j=1}^{M} w_t^j \xrightarrow{p} \int \left\{ \frac{g(\alpha_{1:t})}{q(\alpha_{1:t}|\mathcal{F}_t)} \right\} q(\alpha_{1:t}|\mathcal{F}_t)\mathrm{d}\alpha_{1:t} \\ &= \int g(\alpha_{1:t})\mathrm{d}\alpha_{1:t} = f(y_{1:t}|\mathcal{F}_0) = L_t.\end{aligned}$$

  - So is consistent as $M \to \infty$.
  - It is trivially simulation unbiased

$$\mathrm{E}_U(\widehat{L}_{t,M}|\mathcal{F}_t) = L_t$$

- Of course

$$\widehat{f}(\alpha_{1:t}|\mathcal{F}_t) = \frac{1}{\widehat{L}_{t,M}} g(\alpha_{1:t}).$$

- Notice that

$$
\begin{aligned}
\mathrm{Var}_U\left(\widehat{L}_{t,M}|\mathcal{F}_t\right) &= \mathrm{E}_U\left(\widehat{L}^2_{t,M}|\mathcal{F}_t\right) - L_t^2 \\
&= \int \left\{ \frac{g(\alpha_{1:t})}{q(\alpha_{1:t}|\mathcal{F}_t)} \right\}^2 q(\alpha_{1:t}|\mathcal{F}_t)\mathrm{d}\alpha_{1:t} - L_t^2 \\
&= L_t^2 \int \left\{ \frac{f(\alpha_{1:t}|\mathcal{F}_t)}{q(\alpha_{1:t}|\mathcal{F}_t)} \right\}^2 q(\alpha_{1:t}|\mathcal{F}_t)\mathrm{d}\alpha_{1:t} - L_t^2,
\end{aligned}
$$

- Typically written compactly as

$$
\mathrm{Var}_U\left(\frac{\widehat{L}_{t,M}}{L_t}|\mathcal{F}_t\right) = \int \left\{ \frac{f(\alpha_{1:t}|\mathcal{F}_t)}{q(\alpha_{1:t}|\mathcal{F}_t)} \right\}^2 q(\alpha_{1:t}|\mathcal{F}_t)\mathrm{d}\alpha_{1:t} - 1.
$$

Properties of importance sampling

So summarizing

Fact

Importance sampler: $\widehat{L}_{t,M} = \frac{1}{M}\sum_{j=1}^{M} w_t^j$ *has the properties*

$$\mathrm{E}(\widehat{L}_{t,M}) = L_t, \quad \frac{\mathrm{Var}\left(\widehat{L}_{t,M}|\mathcal{F}_t\right)}{L_t^2} = \frac{1}{M}\int \frac{f(\alpha_{1:t}|\mathcal{F}_t)^2}{q(\alpha_{1:t}|\mathcal{F}_t)}\mathrm{d}\alpha_{1:t} - 1.$$

Example

Doucet and Johansen (2011): Suppose under f: $\alpha_{1:t} \sim N(0, I_t)$, and

$$g(\alpha_{1:t}) = \left(\sqrt{2\pi}\right)^{-t} \prod_{s=1}^{t} \exp(-\alpha_s^2/2), \quad z_t = 1.$$

under q: $\alpha_{1:t} \sim N(0, \sigma^2 I_t)$, then

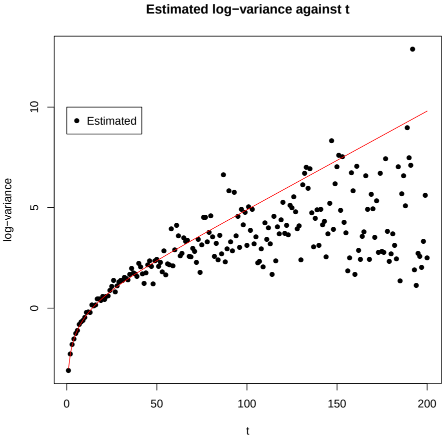
$$\begin{aligned} M\frac{\mathrm{Var}\left(\widehat{L}_{t,M}\right)}{L_t^2} &= \left\{\left(\sqrt{2\pi\sigma^2}\right)^t \int \exp\left(-\frac{2\sigma^2-1}{2\sigma^2}\sum_{s=1}^{t}\alpha_s^2\right) \mathrm{d}\alpha_{1:t} - 1\right\} \\ &= \left\{\left(\frac{\sigma^4}{(2\sigma^2-1)}\right)^{t/2} - 1\right\}. \end{aligned}$$

Hence if $\sigma^2 > 1/2$ and $\sigma^2 \neq 1/2$, grows exp with t. If $\sigma^2 = 1.2$ (stronger tails than target) then $M\mathrm{Var}\left(\widehat{L}_{t,M}\right)/L_t^2 \simeq (1.103)^{t/2}$, which is 1.9×10^{21} when $t = 1,000$. Even good step-by-step approximations fall down.

```
logf <- function(alpha) { # log posterior
 logp = log(dnorm(alpha,0.0,1.0))

 logp # return logL
}
logq <- function(alpha,sigma) { # log proposal density
 logp = log(dnorm(alpha,0.0,sigma))

 logp
}
T =200 # max t
sigma = sqrt(1.2)
M = 1000 # simulation size
mStore = array(0, dim=c(T,5)) # store answers
mLogWeight<- array(0, dim=c(M,1)) # storage for result
sigma = 1.2
for (t in (1:T)){
 for (i in (1:M)){
```

```
 alpha = rnorm(t,0,sigma) # propose
 logw = sum(logf(alpha) - logq(alpha,sigma))
 mLogWeight[i] = logw # storage for results
 }

 aVar = ((((sigma^4)/((2.0*sigma*sigma)-1.0))^(0.5*t))-1.0)
 mStore[t,] =
c(mean(exp(mLogWeight)),1.0,log(var(exp(mLogWeight))),t,log(aV
}
pdf("ImportanceLogVar.pdf")
plot(mStore[,3],main="Estimated log-variance against
t",ylab="log-variance",xlab="t",col=c(1),pch = c(16))
lines(mStore[,5],type="l",col=2)
legend(x=0,y=10,c("Estimated"),col=c(1),pch = c(16))
dev.off()
```

Simulation of Gaussian random walk, $M = 1,000$



Importance sampler's, against t,

$$\log\left\{\left(\frac{\sigma^4}{(2\sigma^2-1)}\right)^{t/2} - 1\right\}.$$

Sequential Importance Sampling I

To solve the exponential problem we first reexpress importance sampling. Alone this does not help.

- In sequential calculations we step through the data.
- Build a sequential proposal:

$$
\begin{aligned}
q(\alpha_{1:t+1}|\mathcal{F}_{t+1}) &= q(\alpha_{1:t}|\mathcal{F}_t)q(\alpha_{t+1}|\alpha_{1:t},\mathcal{F}_{t+1}) \\
&= q(\alpha_{1:t-1}|\mathcal{F}_{t-1})q(\alpha_t|\alpha_{1:t-1},\mathcal{F}_t)q(\alpha_{t+1}|\alpha_{1:t},\mathcal{F}_{t+1}) \\
&= q(\alpha_1|\mathcal{F}_1)\prod_{s=1}^{t} q(\alpha_{s+1}|\alpha_{1:s},\mathcal{F}_{s+1}).
\end{aligned}
$$

- Now recall $g(\alpha_{1:t}) = f(\alpha_{1:t})f(y_{1:t}|\alpha_{1:t})$. Write

$$
g(\alpha_{t+1}|\alpha_{1:t}) = \frac{g(\alpha_{1:t+1})}{g(\alpha_{1:t})} = f(y_{t+1}|\alpha_{t+1})f(\alpha_{t+1}|\alpha_t).
$$

Sequential Importance Sampling II

- Then the importance sampling weights are

$$
\begin{aligned}
w^j_{t+1} &= \frac{g(\alpha^{(j)}_{1:t+1})}{q(\alpha^{(j)}_{1:t+1}|\mathcal{F}_{t+1})} = \frac{g(\alpha^{(j)}_1)}{q(\alpha^{(j)}_1|\mathcal{F}_1)} \prod_{s=1}^{t} \frac{g(\alpha^{(j)}_{t+1}|\alpha^{(j)}_{1:t})}{q(\alpha^{(j)}_{s+1}|\alpha^{(j)}_{1:s}, \mathcal{F}_{s+1})} \\
&= w^j_t c^j_{t+1}, \quad c^j_{t+1} = \frac{g(\alpha^{(j)}_{t+1}|\alpha^{(j)}_{1:t})}{q(\alpha^{(j)}_{t+1}|\alpha^{(j)}_{1:t}, \mathcal{F}_{t+1})},
\end{aligned}
$$

which can be computed sequentially.

Sequential Importance Sampling I

Can propose paths $\alpha_{1:t+1}$ by simulating:

1. Start with a sample $\left\{\alpha_0^{*j}\right\}_{j=1,2,\ldots,M}$ from $f(\alpha_0|\mathcal{F}_0)$. Set $t=0$.
2. Simulate α_{t+1} from $q(\alpha_{t+1}|\alpha_{1:t},\mathcal{F}_{t+1})$, recording

$$w_{t+1}^j = w_t^j c_{t+1}^j, \quad c_{t+1}^j = \frac{g(\alpha_{t+1}^{(j)}|\alpha_{1:t}^{(j)})}{q(\alpha_{t+1}^{(j)}|\alpha_{1:t}^{(j)},\mathcal{F}_{t+1})}.$$

3. Set $t=t+1$, goto 2.

Sequential Importance Sampling II

Then

$$\widehat{L}_{t,M} = \frac{1}{M}\sum_{j=1}^{M} w_t^j \xrightarrow{p} \int g(\alpha_{1:t})\mathrm{d}\alpha_{1:t} = L_t,$$

so

$$\frac{\widehat{L}_{t,M}}{\widehat{L}_{t-1,M}} = \frac{\sum_{j=1}^{M} w_{t,M}^j}{\sum_{j=1}^{M} w_{t-1,M}^j} \xrightarrow{p} \frac{L_t}{L_{t-1}} = \frac{f(y_{1:t}|\mathcal{F}_0)}{f(y_{1:t-1}|\mathcal{F}_0)} = f(y_t|\mathcal{F}_{t-1}).$$

- Now recall the notation $g(\alpha_{t+1}|\alpha_{1:t})$:

$$\begin{aligned} g(\alpha_{t+1}|\alpha_{1:t}) &= f(y_{t+1}|\alpha_{t+1})f(\alpha_{t+1}|\alpha_{1:t}) \\ &= f(\alpha_{t+1}|\alpha_{1:t}, y_{t+1})f(y_{t+1}|\alpha_{1:t}). \end{aligned}$$

  Hence if we select

$$q(\alpha_{t+1}|\alpha_{1:t}, \mathcal{F}_{t+1}) = f(\alpha_{t+1}|\alpha_{1:t}, y_{t+1})$$

Sequential Importance Sampling III

then

$$c_{t+1}^{j} = \frac{g(\alpha_{t+1}^{(j)}|\alpha_{1:t}^{(j)})}{q(\alpha_{t+1}^{(j)}|\alpha_{1:t}^{(j)}, \mathcal{F}_{t+1})} = f(y_{t+1}|\alpha_{1:t}^{(j)}).$$

- Called "exact adaption" in this literature.
- Typical sequential importance sampling strategy is to produce

$$q(\alpha_{s+1}|\alpha_{1:s}, \mathcal{F}_{s+1}) \simeq f(\alpha_{s+1}|\alpha_{1:s}, y_{t+1}).$$

- Sequential importance sampling is a special case of importance sampling, so gets exponentially worse as t increases.
- Need a new idea.

Resampling I

- The above uses the proposals

$$q(\alpha_{1:t}|\mathcal{F}_t) = q(\alpha_1|\mathcal{F}_1)\prod_{s=1}^{t-1} q(\alpha_{s+1}|\alpha_{1:s}, \mathcal{F}_{s+1}). \tag{6}$$

- But we have a reasonable estimate of the distribution function

$$F_{t,M}(x) = \sum_{j=1}^{M} W_{t,M}^{j} 1(\alpha_{1:t}^{(j)} \leq x). \tag{7}$$

  As $M \to \infty$ has the correct distribution function!
- We can sample from this instead of the proposal density (6).
- This is called **Sequential Monte Carlo**.
- A typical way to resample is Sample, importance resampling (SIR).

Generic particle filter

Algorithm:

1. Start with a sample $\left\{\alpha_0^{*j}\right\}_{j=1,2,\ldots,M}$ from $f(\alpha_0|\mathcal{F}_0)$. Set $t=1$.
2. Simulate $q(\alpha_t^j|\alpha_{t-1}^{*j},\mathcal{F}_t)$, and compute, where $g(\alpha_t|\alpha_{t-1}) = f(y_t|\alpha_t)f(\alpha_t|\alpha_{t-1})$,

$$w_t^j = \frac{g(\alpha_t^j|\alpha_{t-1}^j)}{q(\alpha_t^j|\alpha_{t-1}^{*j},\mathcal{F}_t)}, \quad W_t^j = \frac{w_t^j}{\sum_{i=1}^M w_t^i}.$$

   Record

$$\frac{\widehat{L}_t}{L_{t-1}} = \widehat{f}(y_t|\mathcal{F}_{t-1}) = \frac{1}{M}\sum_{j=1}^M w_t^j.$$

   Resample $\left\{W_t^j,\alpha_t^j\right\}$. Call these draws $\left\{\frac{1}{M},\alpha_t^{*j}\right\}$.
3. Set $t=t+1$ and goto 2 again.

Basic particle filter

- $q(\alpha_t^j|\alpha_{t-1}^{*j}, \mathcal{F}_t)$, how to?
- Most famous is the Gordon et al. (1993) sampler:

$$q(\alpha_t|\alpha_{t-1}, \mathcal{F}_t) = f\left(\alpha_t|\alpha_{t-1}\right),$$

  simulate transition equation. Often called bootstrap particle filter.
- Weight becomes

$$w_t(\alpha_{1:t}) = \frac{g(\alpha_{1:t})}{q(\alpha_t|\alpha_{t-1}, \mathcal{F}_t)} = \frac{f(y_t|\alpha_t)f(\alpha_t|\alpha_{t-1})}{f\left(\alpha_t|\alpha_{t-1}\right)} = f(y_t|\alpha_t).$$

- Don't need transition den & easy to code!
- But the proposal is blind :— does not use y_t. Fix in a moment.
- Generic properties?

Generic properties of basic particle filter

Fact

State space models with $q(\alpha_t|\alpha_{t-1},\mathcal{F}_t)$ if multinomial resampling is used at each step, then

$$\mathrm{E}_u\left(\frac{\widetilde{L}_{t,M}}{L_t}|\mathcal{F}_t\right)=1$$

and

$$M\times\mathrm{Var}_u\left(\frac{\widetilde{L}_{t,M}}{L_t}|\mathcal{F}_t\right)\to\int\frac{f(\alpha_1|\mathcal{F}_t)^2}{q(\alpha_1|\mathcal{F}_t)}\mathrm{d}\alpha_1-1$$
$$+\sum_{s=2}^{t}\left\{\int\frac{f(\alpha_{s-1},\alpha_s|\mathcal{F}_t)^2}{f(\alpha_{s-1}|\mathcal{F}_{s-1})q(\alpha_s|\alpha_{s-1},\mathcal{F}_s)}\mathrm{d}\alpha_{s-1:s}-1\right\}.$$

- The proof of unbiasedness is difficult, it is due to Moral (2004). A more accessible version appears in Pitt et al. (2012).

Generic properties of basic particle filter

$$\int \frac{f(\alpha_{s-1}, \alpha_s|\mathcal{F}_t)^2}{f(\alpha_{s-1}|\mathcal{F}_{s-1})q(\alpha_s|\alpha_{s-1}, \mathcal{F}_s)} \mathrm{d}\alpha_{s-1:s}$$

- The variance is coming from

  $$\frac{f(\alpha_{s-1}, \alpha_s|\mathcal{F}_t)^2}{f(\alpha_{s-1}|\mathcal{F}_{s-1})} = \frac{f(\alpha_s|\mathcal{F}_t)^2 f(\alpha_{s-1}|\alpha_s, \mathcal{F}_{s-1})^2}{f(\alpha_{s-1}|\mathcal{F}_{s-1})} = \frac{f(\alpha_{1:s}|\mathcal{F}_t)^2}{f(\alpha_{1:s-1}|\mathcal{F}_{s-1})}.$$

  Like doing t lots of independent importance sampling of

  $$\begin{aligned} & \frac{1}{L_t} \int g(\alpha_{1:s}|\mathcal{F}_t) \mathrm{d}\alpha_{1:s} \\ = & \int \frac{f(\alpha_{1:s}|\mathcal{F}_t)}{f(\alpha_{1:s-1}|\mathcal{F}_{s-1})q(\alpha_s|\alpha_{s-1}, \mathcal{F}_s)} \\ & f(\alpha_{1:s-1}|\mathcal{F}_{s-1})q(\alpha_s|\alpha_{s-1}, \mathcal{F}_s) \mathrm{d}\alpha_{1:s} \end{aligned}$$

  where the importance sampling uses the correct distribution up to time $s-1$.
- This correct distribution is generated by a reset caused by resampling.

Generic properties of basic particle filter

Why this form?

$$\frac{\widehat{L}_t}{L_t} = \prod_{s=1}^{t} \frac{\widehat{l}_s}{l_s}, \quad l_s = f(y_s|\mathcal{F}_{s-1}), \quad \widehat{l}_s = \widehat{f}(y_s|\mathcal{F}_{s-1})$$

In the simple case where

$$\mathrm{E}_u(\widehat{l}_s|\mathcal{F}_t) = l_s,$$

and $\widehat{l}_t/l_t$ are independent over t conditional on $\mathcal{F}_t$. Then

$$\begin{aligned}
\mathrm{Var}_u\left(\frac{\widehat{L}_t}{L_t}|\mathcal{F}_t\right) &= \prod_{s=1}^{t} \mathrm{E}_u\left\{\left(\frac{\widehat{l}_s}{l_s}\right)^2 |\mathcal{F}_t\right\} - 1 \\
&= \prod_{s=1}^{t}\left[\mathrm{Var}_u\left\{\left(\frac{\widehat{l}_s}{l_s}\right) |\mathcal{F}_t\right\} + 1\right] - 1 \\
&= \prod_{s=1}^{t}\left(\frac{\sigma_s^2}{M} + 1\right) - 1 \simeq \frac{1}{M}\sum_{s=1}^{t}\sigma_s^2 - 1.
\end{aligned}$$

Properties I

Example

Continued. Here the proposal $q(\alpha_s|\alpha_{s-1}, \mathcal{F}_s) = q(\alpha_s)$. Then with multinomial resampling:

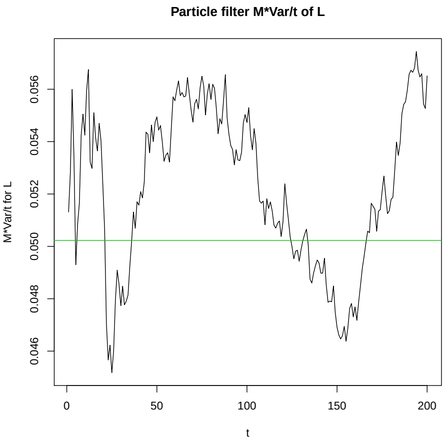
$$
\begin{aligned}
M \times \mathrm{Var}\left(\frac{\widetilde{L}_{t,M}}{L_t}\right) &\simeq \sum_{s=1}^{t}\left\{\int \frac{f(\alpha_s)^2}{q(\alpha_s)}\mathrm{d}\alpha_s - 1\right\} \\
&= t\left\{\left(\frac{\sigma^4}{(2\sigma^2-1)}\right)^{1/2} - 1\right\}.
\end{aligned}
$$

Hence the error accrues linearly not exponentially.

Generic properties of basic particle filter

```
iRep = 500
sigma = 1.2
T = 200
M = 50
mlogL = array(0, dim=c(iRep,T))
for (i in (1:iRep)){
 for (t in (1:T)){
 alpha = rnorm(M,0,sigma)
 logw = logf(alpha) - logq(alpha,sigma)
 w1 = exp(logw-max(logw))
 Wstar = w1/sum(w1)
 sample(alpha,M,replace=T,prob=Wstar) # irrelevant here

 mlogL[i,t] = max(logw) + log(mean(w1))
 }
 mlogL[i,] = cumsum(mlogL[i,]) # cumulate the logL
}
X = exp(mlogL) # compute L
```

Simulation of Gaussian random walk



Importance sampler's, against t,

$$\frac{M}{t} \times \mathrm{Var}\left(\frac{\widetilde{L}_{t,M}}{L_t}\right) \simeq \left\{\left(\frac{\sigma^4}{(2\sigma^2-1)}\right)^{1/2} - 1\right\}.$$

Properties I

- Now analyze a stochastic volatility process:

$$
\begin{aligned}
y_t &= \varepsilon_t e^{\alpha_t/2}, \quad t = 1, 2, ..., T, \\
\alpha_{t+1} &= \mu + \phi(\alpha_{t+1} - \mu) + \eta_t.
\end{aligned}
$$

Use simulated data with $T = 500$, $\mu = -0.5$, $\phi = 0.98$, $\sigma = 0.13$.

```
logf <- function(y,alpha) { # log posterior
 logp = log(dnorm(y,0.0,exp(0.5*alpha)))
 logp # return logL
}
T = 500
mu = -0.5
phi =0.98
sigma = 0.13
y = simSV(T,mu,phi,sigma)
```

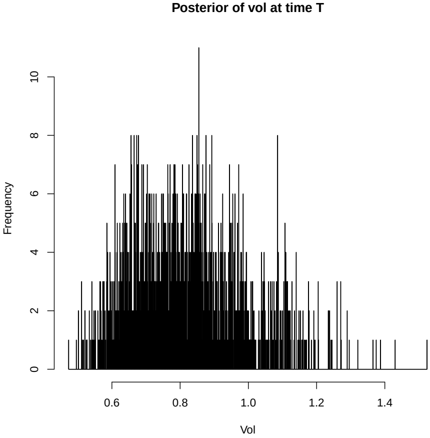
Properties II

```
iRep = 250
T = 500
M = 2000
mlogL = array(0, dim=c(iRep,T))
mlogLmax = array(0, dim=c(1,T))
for (i in (1:iRep)){
 alpha = mu + rnorm(M,0.0,sigma/sqrt(1.0-(phi^2)))
 for (t in (1:T)){
 alpha = mu + phi*(alpha-mu) + sigma*rnorm(M,0,1)
 logw = logf(y[t],alpha)
 w1 = exp(logw-max(logw))
 Wstar = w1/sum(w1)
 alpha = sample(alpha,M,replace=T,prob=Wstar) # irr here
 mlogL[i,t] = max(logw) + log(mean(w1))
 }
 mlogL[i,] = cumsum(mlogL[i,]) # cumulate the logL
```

Properties III

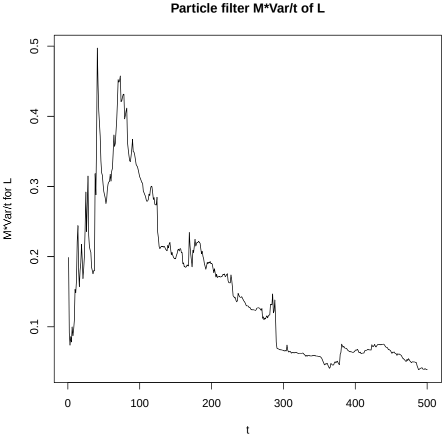
```
}
pdf("ParticleVolEst.pdf")
hist(exp(0.5*alpha),breaks=M,main="Posterior of vol at time
T",xlab="Vol")
dev.off()
for (t in (1:T)){
 mlogLmax[t] = max(mlogL[,t])
 mlogL[,t] = mlogL[,t] - mlogLmax[t]
}
X = exp(mlogL) # compute L
pdf("SVParticle.pdf")
plot(M*apply(X,2,var)/(1:T),ylab="M*Var/t for
L",xlab="t",main="Particle filter M*Var/t of L",type="l")
dev.off()
```

Samples from vol posterior at time T



Histogram of samples from $e^{\alpha_T/2}|\mathcal{F}_T$.

Samples from vol posterior at time T



Scaled estimated variance of the estimated likelihood at time t.

Properties I

- Typically can estimate posterior moments

$$\sum_{j=1}^{M} W_t^j h(\alpha_t^{(j)}) \xrightarrow{p} \mathrm{E}\left\{h(\alpha_t)|\mathcal{F}_t\right\},$$

unbiasedly and with error which is order $O(M^{-1})$ and does not increase with t. This needs some mixing assumptions on the smoothing distribution

$$\frac{1}{2}\int \left|f(\alpha_t|\alpha_k,\mathcal{F}_t) - f(\alpha_t|\alpha_k',\mathcal{F}_t)\right| \mathrm{d}\alpha_t \le \beta^{t-k}, \quad \beta \in [0,1)\,,$$

uniformly over (α_k, α_k') if $t-k$ is large enough.

Better resampling I

- Resample $\left\{W_t^j, \alpha_t^j\right\}_{j=1,2,\dots,M}$. Call these draws $\left\{\alpha_t^{*j}\right\}_{j=1,2,\dots,M}$.
- Although they kill silly particles, they add unnecessary randomness.
- Replace some of the multinomial sampling by residual sampling

$$
\begin{aligned}
n_t^j &= \left\lfloor M W_t^j \right\rfloor, \quad v_{t,j} = M W_t^j - n_{t,j} \in [0,1], \\
N_t &= \sum_{j=1}^{M} n_t^j, \quad V_{t,j} = \frac{v_{t,j}}{\sum_{i=1}^{M} v_{t,i}}.
\end{aligned}
$$

- Then we do multinomial sampling with probability $\left\{V_{t,j}, \alpha_t^j\right\}_{j=1,2,\dots,M}$ to produce $M - N_t$ draws and add systematically n_t^j copies of α_t^j to them for $j = 1, 2, \dots, M$.
- Has the advantage that if the weights are even, it does not resample.
- Alternatives: stratified sampling & systematic sampling.

Basic particle filter I

- Like to use:

$$
\begin{aligned}
q(\alpha_t|\alpha_{t-1}, \mathcal{F}_t) &= f(\alpha_t|\alpha_{t-1}, \mathcal{F}_t) \\
&\propto f(y_t|\alpha_t) f(\alpha_t|\alpha_{t-1}).
\end{aligned}
$$

  In specific models can find good approximations, e.g. approximate using normal distribution, e.g. extended Kalman filter, or a heavier tail alternative.

- Most well known approach is the Pitt and Shephard (1999) "auxiliary particle filter."

Basic particle filter I

- It starts with

$$
\begin{aligned}
f(\alpha_t|\mathcal{F}_t) &\propto f(y_t|\alpha_t)f(\alpha_t|\mathcal{F}_{t-1}) \\
&\simeq f(y_t|\alpha_t)\left\{\frac{1}{M}\sum_{j=1}^{M} f(\alpha_t|\alpha_{t-1}^{j})\right\}.
\end{aligned}
$$

- Bootstrap particle filter takes each j and then draws $\alpha_t|\alpha_{t-1}^{j}$ and reweight using the likelihood.
- Auxiliary particle filter does not sample each j equally, but boosts one which are likely to deliver high $f(y_t|\alpha_t)$.

Basic particle filter II

- More abstractly, can sample from above by sampling from

$$g(\alpha_t, j) \propto f(y_t|\alpha_t) f(\alpha_t|\alpha_{t-1}^j),$$

has required marginal distribution. e.g. SIR, sample j with probability

$$\pi_t^j \propto f(y_t|\alpha_t = \alpha_{t-1}^j),$$

then α_t^j through $\alpha_t|\alpha_{t-1}^j$, then evaluate weight

$$w_{t,j} = \frac{f(y_t|\alpha_t^j)}{\pi_j}.$$

Basic particle filter I

- This suggests APFs use two levels of resampling.
- But actually you can do it once. Before resampling at time t, we have

$$\left\{ W_t^j, \alpha_{1:t}^{(j)} \right\}.$$

- Usually we resample from above, but instead use weights

$$w_t^{*j} = W_t^j \pi_{t+1}^j, \quad \pi_{t+1}^j = f(y_{t+1} | \alpha_{t+1} = \alpha_t^j).$$

- Then resample, producing an unequal measure

$$\left\{ \frac{1}{\pi_{t+1}^j}, \alpha_t^{*j} \right\},$$

from $\alpha_t | \mathcal{F}_t$. This can be simulated forward for each j.

Particle filter and MCMC I

- Particle filters deliver unbiased estimates of the likelihood.
- Can put this estimated likelihood inside the H-M algorithm.
- Gives the correct posterior distribution.
- Focus of intense recent research:
  - Andrieu et al. (2010)
  - Flury and Shephard (2011)
  - Pitt et al. (2012)
  - Gerber and Chopin (2014), quasi-Monte Carlo based particle filters

Basic particle filter I

- From Flury and Shephard (2011).
- The stock returns and stochastic volatility factor are assumed to follow the processes

$$\begin{array}{rcl} y_t & = & \mu + \exp\left\{\beta_0 + \beta_1 \alpha_t\right\} \varepsilon_t, \\ \alpha_{t+1} & = & \phi \alpha_t + \eta_t \end{array} \qquad \left(\begin{array}{c} \varepsilon_t \\ \eta_t \end{array}\right) \overset{i.i.d.}{\sim} N\left(0, \left(\begin{array}{cc} 1 & \rho \\ \rho & 1 \end{array}\right)\right),$$

  where $\alpha_0 \sim N\left(0, \left(1-\phi^2\right)^{-1}\right)$.
- Have $\theta = (\mu, \beta_0, \beta_1, \phi, \rho)'$.
- The likelihood is not available.

Basic particle filter I

- We assume a Gaussian prior given by $\theta \sim N(\theta_0, I_5)$ where $\theta_0 = (0.036, -0.286, 0.077, 0.984, -0.794)'$.
- Any proposals for $\phi, \rho \notin (-1, 1)$ are automatically rejected.
- We are using the following random walk proposals $\Delta\mu_i = 0.017\nu_{1,i}$, $\Delta\beta_{0,i} = 0.104\nu_{2,i}$, $\Delta\beta_{1,i} = 0.010\nu_{3,i}$, $\Delta\phi_i = 0.004\nu_{4,i}$, $\Delta\rho_i = 0.067\nu_{5,i}$, where $\nu_{j,i} \overset{i.i.d.}{\sim} N(0, 1)$ for $j = 1, \ldots, 5$ and $i = 1, \ldots, N$.

Basic particle filter I

- We use $3,271$ daily observations from 03.01.1995 until 31.12.2007 of the end-of-day level of the S&P500 Composite Index (NYSE/AMEX only) from CRSP.
- Daily returns are defined as $y_t = 100\,(\log P_t - \log P_{t-1})$.

Basic particle filter I

- This analysis is based on $M = 2,000$ and $N = 100,000$.
- Convergence looks pretty fast for this algorithm, with very modest inefficiency factors.
- It is computationally demanding as each simulated likelihood evaluation is quite expensive in terms of time, but the coding effort is very modest indeed.
- The implementation of this method is quite simple and seems competitive with other procedures put forward in the literature.

Summary of posterior I

$$\begin{aligned} y_t &= \mu + \exp\left\{\beta_0 + \beta_1 \alpha_t\right\} \varepsilon_t, \\ \alpha_{t+1} &= \phi \alpha_t + \eta_t \end{aligned} \qquad \begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} \overset{i.i.d.}{\sim} N\left(0, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right),$$

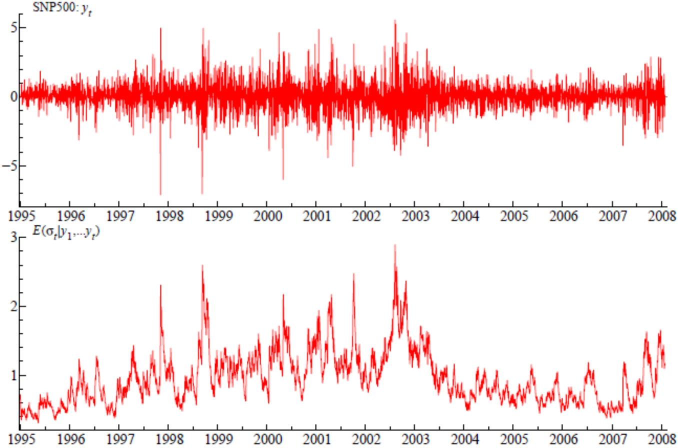
	mean	mcse	Pr	posterior covariance and correlation					inef
μ	0.042	0.000	0.410	0.000	-0.686	-0.137	0.304	0.145	15
β_0	-0.141	0.001	0.395	-0.001	0.006	0.061	-0.222	-0.044	12
β_1	0.080	0.000	0.398	0.000	0.000	0.000	-0.688	-0.015	18
ϕ	0.982	0.000	0.424	0.000	0.000	0.000	0.000	-0.054	16
ρ	-0.742	0.000	0.427	0.000	0.000	0.000	0.000	0.002	6.4

Summary of output I

- The most important aspect of the estimated parameters is the extremely strong negative statistical leverage parameter ρ
- The high negative posterior cor between μ & β_0 and β_1 & ϕ.
- ρ is not importantly correlated with the other parameters in the model.

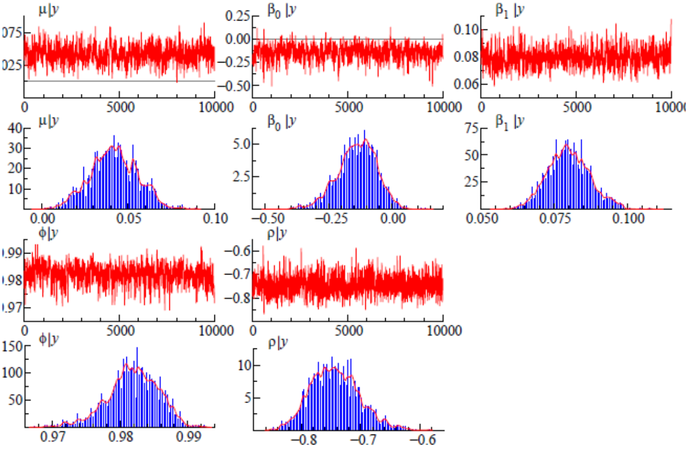
Samples from vol posterior at time T

S&P500 and the filtered volatility $\mathrm{E}\left\{\exp\left(\beta_0 + \beta_1 \alpha_t\right) | \mathcal{F}_t\right\}$



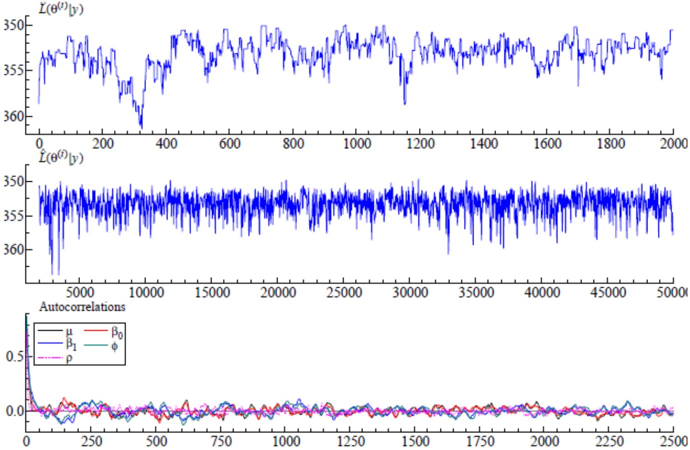
Samples from vol posterior at time T

Particle based MCMC paths



Samples from vol posterior at time T

Particle based MCMC paths of logL



Summary

- Linear Gaussian problem is solved by the Kalman filter.
- Finite support is solved by Baum (Hamilton) filter
- Sequential Monte Carlo is a recent advance on importance sampling
- Can be applied outside time series
- Allows learning on non-Gaussian state space models
- MCMC + particle filters yield likelihood inference
- Completes extension of Kalman filter to general problems

References I

Andrieu, C., A. Doucet, and R. Holenstein (2010). Particle Markov chain Monte Carlo methods (with discussion). *Journal of the Royal Statistical Society, Series B 72*, 1–33.

Calvet, L. and V. Czellar (2014). Accurate methods for approximate Bayesian computation filtering. *Journal of Financial Econometrics*. Forthcoming.

Calvet, L. and A. Fisher (2001). Forecasting multifractal volatility. *Journal of Econometrics 105*, 27–58.

Calvet, L. and A. Fisher (2002). Multifractality in asset returns: theory and evidence. *Review of Economics and Statistics 84*, 381–406.

Carter, C. K. and R. Kohn (1994). On Gibbs sampling for state space models. *Biometrika 81*, 541–553.

Creal, D. (2012). A survey of sequential Monte Carlo methods for economics and finance. *Econometric Reviews 31*, 245–296.

References II

Crisan, D. and B. Rozovskii (2011). *The Oxford Handbook of Nonlinear Filtering*. Oxford: Oxford University Press.

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

Diebold, F. X. and G. D. Rudebusch (2013). *Yield curve modelling and forecasting: the dyanmics Nelson-Siegel approach*. Princeton University Press.

Douc, R., E. Moulines, and D. S. Stoffer (2014). *Nonlinear time series: theory, methods and applications with R examples*. Boca Raton: Chapman and Hall.

Doucet, A., N. de Freitas, and N. J. Gordon (Eds.) (2001). *Sequential Monte Carlo Methods in Practice*. New York: Springer-Verlag.

Doucet, A. and A. Johansen (2011). A tutorial on particle filtering and smoothing: fifteen years later. In D. Crisan and B. Rozovsky (Eds.), *The Oxford Handbook of Nonlinear filtering*. Oxford University Press.

References III

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

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

Fearnhead, P. (2011). Mcmc and state-space models. pp. 513–529. Chapman and Hall.

Flury, T. and N. Shephard (2011). Bayesian inference based only on simulated likelihood: particle filter analysis of dynamic economic models. *Econometric Theory 27*, 933–956.

Frühwirth-Schnatter, S. (1994). Data augmentation and dynamic linear models. *Journal of Time Series Analysis 15*, 183–202.

Gerber, M. and N. Chopin (2014). Sequential quasi-Monte Carlo. *Journal of the Royal Statistical Society, Series B*. Forthcoming.

References IV

Gordon, N. J., D. J. Salmond, and A. F. M. Smith (1993). A novel approach to nonlinear and non-Gaussian Bayesian state estimation. *IEE-Proceedings F 140*, 107–113.

Granger, C. W. J. (1980). Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics 14*, 227–238.

Hamilton, J. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica 57*, 357–384.

Hansen, L. P. and T. J. Sargent (2007). *Robustness*. Princeton University Press.

Hansen, L. P. and T. J. Sargent (2014). *Recursive Models of Dynamic Linear Economies*. Princeton University Press.

Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.

References V

Harvey, A. C. (1993). *Time Series Models* (2 ed.). Hemel Hempstead: Harvester Wheatsheaf.

Hodrick, R. J. and E. C. Prescott (1997). Postwar U.S. business cycles: an empirical investigation. *Journal of Money, Credit, and Banking 24*, 1–16.

Hull, J. and A. White (1987). The pricing of options on assets with stochastic volatilities. *Journal of Finance 42*, 281–300.

Johannes, M. and N. Polson (2009). Particle filtering. In T. G. Andersen, R. A. Davis, J.-P. Kreiss, and T. Mikosch (Eds.), *Handbook of Financial Time Series*, pp. 1015–1029. Springer.

Johannes, M., N. Polson, and J. Stroud (2009). Optimal filtering of jump-diffusions: Extracting latent states from asset prices. *Review of Financial Studies 22*, 2259–2299.

References VI

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering, Transactions ASMA, Series D 82*, 35–45.

Kitagawa, G. (1987). Non-Gaussian state space modelling of non-stationary time series. *Journal of the American Statistical Association 82*, 503–514.

Li, H. and R. Durbin (2011). Inference of human population history from individual whole-genome sequences. *Nature 475*, 493–496.

Liu, J. S. (2001). *Monte Carlo Strategies in Scientific Computing*. New York: Springer.

Moral, P. D. (2004). *Feynman-Kac Formulae: Genealogical and Interacting Particle Systems with Applications*. New York: Springer.

Nelson, C. R. and A. F. Siegel (1987). Parsimonious modeling of yield curves. *Journal of Business 60*, 473–489.

References VII

Pitt, M. K. and N. Shephard (1999). Filtering via simulation: auxiliary particle filter. *Journal of the American Statistical Association 94*, 590–599.

Pitt, M. K., R. Silva, P. Giordani, and R. Kohn (2012). On some properties of Markov chain Monte Carlo simulation methods based on the particle filter. *Journal of Econometrics 171*, 134–151.

Rauch, H. E., F. Tung, and C. T. Striebel (1965). *Maximum likelihood estimates of linear dynamic systems*, Volume 3.

Shephard, N. (1994). Local scale model: state space alternative to integrated GARCH processes. *Journal of Econometrics 60*, 181–202.

Stock, J. H. and M. W. Watson (2007). Why has U.S. inflation become harder to forecast? *Journal of Money, Credit, and Banking 39*, 3–34.

Uhlig, H. (1997). Bayesian vector autoregressions with stochastic volatility. *Econometrica 65*, 59–73.

West, M. and J. Harrison (1989). *Bayesian Forecasting and Dynamic Models*. New York: Springer-Verlag.

Whittle, P. (1982). *Optimisation over Time*, Volume 1. Chichester: Wiley.

Whittle, P. (1983a). *Optimisation over Time*, Volume 2. Chichester: Wiley.

Whittle, P. (1983b). *Prediction and Regulation* (2 ed.). Oxford: Blackwell. 1st edition, 1963.

Whittle, P. (1990). *Risk-sensitive Optimal Control*. Chichester: Wiley.

Whittle, P. (1996). *Optimal Control; Basics and Beyond*. Chichester: Wiley.