Imperial College London

Imperial College London
Derpartment of Mathematics

State Space Modelling for Statistical Arbitrage

Philippe Remy

CID: 00993306

Supervised by Nikolas Kantas and Yanis Kiskiras

14th June 2015

This report is submitted as part requirement for the MSc Degree in Statistics at Imperial College London. It is substantially the result of my own work except where explicitly indicated in the text. The report will be distributed to the internal and external examiners, but thereafter may not be copied or distributed except with permission from the author.

Abstract

This project is aimed to investigate the practical benefit of using more complex modelling than what is currently standard practice in applications related to statistical arbitrage. The underlying assets will be modelled using appropriate mean-reverting time series or state space models. In order to fit these models to real data the project will involve using advanced particle methods such as Particle Markov Chain Monte Carlo. The primary aim of the project is to assess whether using more advanced modelling and model calibration will result to better performance than simple models used often in practise. This will be illustrated in numerical examples, where the computed portfolio is used for a realistic scenario obtained by popular trading platforms. Simulations will be mainly run in Matlab, but embedding C/C++ routines may be required to speed up computations. The project is a challenging computational Statistics application to finance and is this suitable for a student with an interest in finance, very good apptitute to computing and understanding of the material in the course related to Monte Carlo methods and Time Series.

Table of Contents

1. State Space Modelling

1.1. Bootstrap Particle Filter

The bootstrap particle filter is an iterative method for carrying out Bayesian inference for dynamic state space (partially observed Markov process) models, sometimes also known as hidden Markov models (HMMs). Here, an unobserved Markov process, $x_0, x_1, \ldots, x_T$ governed by a transition kernel $p(x_{t+1}|x_t)$ is partially observed via some measurement model $p(y_t|x_t)$ leading to data $y_1, \ldots, y_T$. The idea is to make inference for the hidden states $x_{0:T}$ given the data $y_{1:T}$. The method is a very simple application of the importance resampling technique. At each time, t, we assume that we have a (approximate) sample from $p(x_t|y_{1:t})$ and use importance resampling to generate an approximate sample from $p(x_{t+1}|y_{1:t+1})$.

More precisely, the procedure is initialised with a sample from $x_0^k \sim p(x_0)$, $k = 1, \ldots, M$ with uniform normalised weights $w'^k_0 = 1/M$. Then suppose that we have a weighted sample $\{x_t^k, w'^k_t | k = 1, \ldots, M\}$ from $p(x_t|y_{1:t})$. First generate an equally weighted sample by resampling with replacement M times to obtain $\{\tilde{x}_t^k | k = 1, \ldots, M\}$ (giving an approximate random sample from $p(x_t|y_{1:t})$). Note that each sample is independently drawn from $\sum_{i=1}^M w'^i_t \delta(x - x_t^i)$. Next propagate each particle forward according to the Markov process model by sampling $x_{t+1}^k \sim p(x_{t+1}|\tilde{x}_t^k)$, $k = 1, \ldots, M$ (giving an approximate random sample from $p(x_{t+1}|y_{1:t})$). Then for each of the new particles, compute a weight $w_{t+1}^k = p(y_{t+1}|x_{t+1}^k)$, and then a normalised weight $w'^k_{t+1} = w_{t+1}^k / \sum_i w_{t+1}^i$.

It is clear from our understanding of importance resampling that these weights are appropriate for representing a sample from $p(x_{t+1}|y_{1:t+1})$, and so the particles and weights can be propagated forward to the next time point. It is also clear that the average weight at each time gives an estimate of the marginal likelihood of the current data point given the data so far. So we define

$$\hat{p}(y_t|y_{1:t-1}) = \frac{1}{M} \sum_{k=1}^{M} w_t^k$$

and

$$\hat{p}(y_{1:T}) = \hat{p}(y_1) \prod_{t=2}^{T} \hat{p}(y_t|y_{1:t-1}).$$

Again, from our understanding of importance resampling, it should be reasonably clear that $\hat{p}(y_{1:T})$ is a consistent estimator of $p(y_{1:T})$. It is much less clear, but nevertheless true that this estimator is also unbiased. The standard reference for this fact is Del

Moral (2004), but this is a rather technical monograph. A much more accessible proof (for a very general particle filter) is given in Pitt et al (2011).

It should therefore be clear that if one is interested in developing MCMC algorithms for state space models, one can use a pseudo-marginal MCMC scheme, substituting in $_{\theta}(y_{1:T}) fromabootstrapparticlefilterinplaceof$p(y$_{1:T}|\theta$). This turns out to be a simple special case of the particle marginal Metropolis-Hastings (PMMH) algorithm described in Andreiu et al (2010).

In bootstrap particle filter, $\pi(x_k^{(i)}|x_{0:k-1}^{(i)}, y_{0:k}) = p(x_k^{(i)}|x_{k-1}^{(i)})$. When the transition prior probability is used as the importance function, the weights update formula is simplified:

$$w_k^{(i)} = w_{k-1}^{(i)} \frac{p(y_k|x_k^{(i)})p(x_k^{(i)}|x_{k-1}^{(i)})}{\pi(x_k^{(i)}|x_{0:k-1}^{(i)}, y_{0:k})} = w_{k-1}^{(i)} p(y_k|x_k^{(i)})$$

In the bootstrap particle filter, it is clear that the average weight at each time gives an estimate of the marginal likelihood of the current data point given the data so far:

$$p_\theta^N(y_t|y_{0:t-1}) = \frac{1}{N}\sum_{k=1}^{N} w_t^k$$

The marginal likelihood at time T is:

$$p_\theta(y_{0:T}) = p(y_0)\prod_{t=1}^{T} p(y_t|y_{1:t-1})$$

Again, from our understanding of importance resampling, it should be reasonably clear that $\hat{p}_\theta^N(y_{0:T})$ is a consistent estimator of $p_\theta(y_{0:T})$. It is much less clear, but nevertheless true that this estimator is also unbiased according to Del Moral (2004).

The marginal log likelihood and its estimator are:

$$\log(p_\theta(y_{0:t})) = \log(p(y_0)) + \sum_{t=1}^{t} \log\left(p(y_t|y_{1:t-1})\right)$$
$$\log(p_\theta^N(y_{0:t})) = -\log(N) + \sum_{t=1}^{t} \log\left(\frac{1}{N}\sum_{k=1}^{N} w_t^k\right)$$

Algorithm 1 Bootstrap Particle Filtering Algorithm (SIR)

procedure

    for i from 1 to N **do**

        $x_k^{(i)} \sim \pi(x_k|x_{0:k-1}^{(i)}, y_{0:k})$

        $w_k^{(i)} = \hat{w}_k^{(i-1)} p(y_k|x_k^{(i)})$

    end

    for i from 1 to N **do**

        $w_k^{(i)} = \hat{w}_k^{(i)} / \sum_{j=1}^{N} \hat{w}_k^{(j)}$

    end

    x_k = resampling(x_k, w_k)

    for i from 1 to N **do**

        $w_k^{(i)} = 1/N$

    end

return (x_k, w_k)

1.2. Stochastic Volatility

1.2.1. Simple SV Model

The following scalar non-linear Hidden Markov model is considered:

$$X_n = \rho X_{n-1} + \sigma V_n$$
$$Y_n = \beta \exp\left(\frac{X_n}{2}\right) W_n$$

where $W_n, V_n \sim \mathcal{N}(0,1)$ iid and $X_0 \sim \mathcal{N}\left(0, \frac{\sigma^2}{1-\rho^2}\right)$. The model is non-linear because the component Y_n has some non-additive noise and because it is non-linear with X_n. A bootstrap particle filter is derived with respect to this model. With little surprise, the SIR (Sequential Importance Sampling) Sequential Monte Carlo filter provides a better estimate than the SIS (Sequential Importance Sampling) filter. It is able to better track the peaks of the process $x^*_{0:T}$.

1.3. PMMH

In a more general context, a Metropolis Hastings scheme can be used to target $p(\theta|y)$ with the ratio:

$$\min\left(1, \frac{p(\theta^\star)}{p(\theta)} \times \frac{q(\theta|\theta^\star)}{q(\theta^\star|\theta)} \times \frac{p(y|\theta^\star)}{p(y|\theta)}\right)$$

where $q(\theta^\star|\theta)$ is the proposal density. In Hidden Markov Models, the marginal likelihood $p(y|\theta)$ is often intractable and the ratio is hard to compute. The simple likelihood-free scheme targets the full joint posterior $p(\theta, x|y)$. Usually $p(x|\theta)$ is tractable. For instance, in the linear Gaussian case, $x_{0:T}$ can be simulated when ρ and τ are known. The MH is built in two stages. First, a new θ^* is proposed from $q(\theta^\star|\theta)$ and then, x^* is sampled from $p(x^\star|\theta^\star)$. The pair $(\theta^\star, x^\star)$ is accepted with the ratio:

$$\min\left(1, \frac{p(\theta^\star)}{p(\theta)} \times \frac{q(\theta|\theta^\star)}{q(\theta^\star|\theta)} \times \frac{p(y|x^\star, \theta^\star)}{p(y|x, \theta)}\right)$$

At each step, the x^* is consistent of θ^* thanks to this proposed mecanism. The main drawback is that T must be really small to have a good acceptance rate. As a matter of fact, since the conditional likelihood $p(y|x^\star, \theta^\star)$ is a product of T terms over the path $x_{0:T}$, it becomes intractable very quickly as T increases. The sampled x^* must also be consistent with y. This is the reason why x^* should be sampled from $p(x^\star|\theta^\star, y)$. The ratio becomes:

$$\min\left(1, \frac{p(\theta^\star)}{p(\theta)} \frac{p(x^\star|\theta^\star)}{p(x|\theta)} \frac{f(\theta|\theta^\star)}{f(\theta^\star|\theta)} \frac{p(y|x^\star, \theta^\star)}{p(y|x, \theta)} \frac{p(x|y, \theta)}{p(x^\star|y, \theta^\star)}\right)$$

Using the basic marginal likelihood identity of Chib (1995), the ratio is simplified to:

$$\min\left(1, \frac{p(\theta^\star)}{p(\theta)} \frac{p(y|\theta^\star)}{p(y|\theta)} \frac{f(\theta|\theta^\star)}{f(\theta^\star|\theta)}\right)$$

Remarkably x is no more present and the ratio is exactly the same as the marginal scheme shown before. Indeed the ideal marginal scheme corresponds to PMMH when $N \to +\infty$. The likelihood-free scheme is obtained with just one particle in the filter. When N is intermediate, the PMMH algorithm is a trade-off between the ideal and the likelihood-free schemes, but is always likelihood-free when one bootstrap particle filter is used.

2. Then you get to the meaty stuff

> " The main sections should guide the reader through your results, analysing them and explaining them. It should show both your successes and your failures in trying to solve your problem (your unsuccessful attempts should be discussed, especially if you have ideas or explanations as to why they failed)."

2.1. With things like Methods

It turns out that doing mathematics properly requires paying attention to many subtle details. For example

$$2 = 2 + (1 - 1) = (2 - 1) + 1 = 1 + 1. \tag{2.1}$$

Beware of the different mathematical fonts, and note that often punctuation should be included at the end of equations. For example, if you understand that

$$0 = c \cdot o \cdot s^2 \cdot 0 = \int_{-\pi}^{\pi} cos^2(x)dx \neq \int_{-\pi}^{\pi} \cos^2(x)\mathrm{d}x = \pi,$$

then your equations will look much nicer.

> " Maths: In formulas use `\exp` and not `exp`."

2.1.1. organised into subsections if you want

Because well ordered thoughts are easier to follow.

or even subsubsections

And paragraphs. You most likely *don't* want to use the `\paragraph` command for this; just skip a line between paragraphs (or use `\par` or `\indent`).

You may wish to have a separate `*.tex` file for each chapter, and use the `\include{}` (or `\input{}`) commands, if that helps you keep things organised. You can then use `\includeonly{}` to only compile the chapter(s) you are currently working on.

3. And yet more meaty stuff

“ You have to submit a thesis, a substantial written thesis normally not exceeding 12000 words. This is a guideline: the appropriate length is a function of the project itself and its subject matter. Excess length disproportionate to the content may be penalised.

The thesis should be on A4-sized paper and typed (ideally using LaTeX), and words or paragraphs must not be crossed out. They should be in a simple binding; a ring or springback binder is suffcient. It is important that students sign the declaration “The work contained in this these is my own work unless otherwise stated". Each thesis should include (i) a brief summary, (ii) an introduction (iii) the main body of the thesis, and (iv) a bibliography.

Two printed copies of the thesis must be submitted to the MSc Administrator before the deadline listed in Section 3. An electronic copy of the thesis (one PDF document) must also be submitted via the Virtual Learning Environment. Late submission may be penalised and will normally delay consideration of the thesis to the following year. The thesis is worth 90% of the project mark.”

“ Figures: The best place is at the top or at the bottom of a page. If this is not possible they should go on a separate page. In LaTeX this can be achieved by

```
\begin{figure}[tbp]
...
\end{figure}
```

When generating plots from R it is usually best to export them as pdf or eps for inclusion in LaTeX. To get Greek letters, sub and superscripts into labels use eg `xlab=expression(alpha[5])`.”

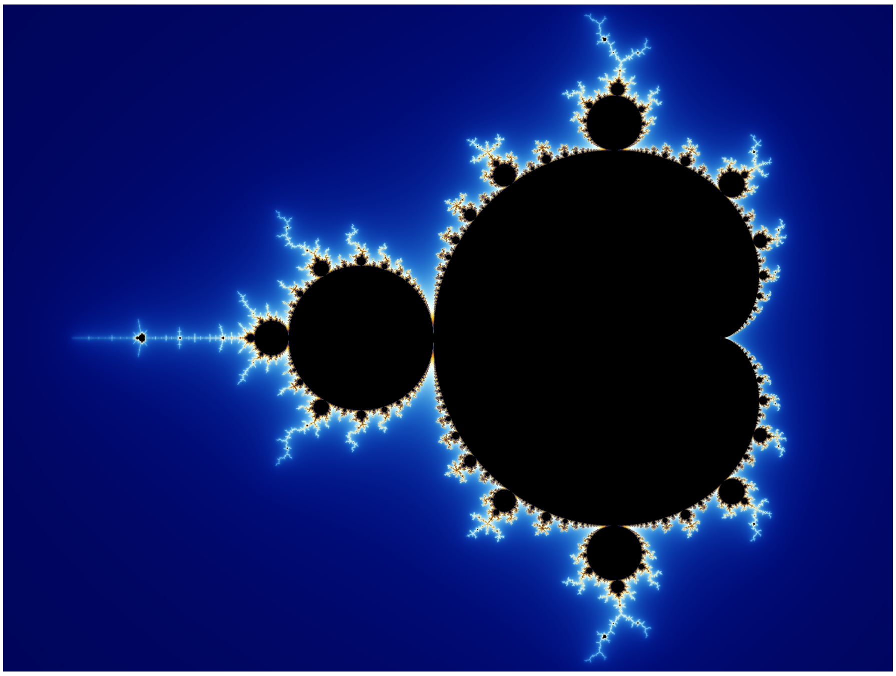

Figure 3.1.: “Graphs and simple diagrams (especially when they are neat) can sometimes be far more effective in presenting results than lots of numbers and/or lots of words.” This figure is by Irina Pechkareva.

4. Conclusion

It's up to you whether you want your introduction/conclusion to be numbered chapters or not. (un-numbered chapters can be created with `\chapter*{}`, but may require you to manually edit the table of contents (see the list of tables in the code comments or the bibliography for an example).

> " The conclusion section should summarise what you have learned. If you would have done more, given more time, you should indicate where your effort would have gone. If your work has raised any unsettled questions, you should address them and indicate what further work needs doing."

4.1. About Referencing

The numbers in this sentence are clickable and refer to equation (2.1) on page 8 and Figure 3.1.

> " List of references: Using author (year) style notation is good practice (the reader may know the paper, but she will definitively not know the number in your reference list. To achieve this in LaTeX you can use BibTeX together with the package natbib. If citing several references together, use `\citep{ref1,ref2,ref3}`. Use a coherent style - either all authors get their first full names or none gets their full names. Books need the name of the publisher, journal articles need the name of the journal. When using BibTeX for generating references, make sure that appropriate capitalisation is used, eg it should be Monte Carlo and not monte carlo. To achieve this in BibTeX, use `{M}onte {C}arlo`."

The bibliography below was generated by BibTeX using the natbib package. You should reference any material you use, e.g. journal articles (**?**). For example **?** explain the theory of generalized linear models in their book. Not citing your sources usually constitutes plagiarism (**?**). I recommend using `Ctrl+F` to search for question marks in your thesis, as these will often result from errors in referencing and cross-referencing (**?**).

There are various software options for managing references, which will allows you to automatically generate BibTeX files. Note however that some manual adjustments are typically required.

A. You probably want to put things like computer programs and long boring proofs in an appendix

> "Any programs in the appendices should be representative. A copy of every single version of every code is unnecessary. Programs should be documented with many comment lines and a discussion of the input necessary to drive them and the output resulting from them as appropriate. Large tables of results should be organised in reference form (as should large sets of graphs) with indices and tables of contents to guide the interested reader through them. Appendices do not count towards the word limit."

There are better ways to include code than the example below, e.g. through the listings package (especially if you intend to print in colour). Although you could write your thesis in Sweave, I would not recommend it since it will make compilation slower.

A.1. R code

```
addone <- function(x) {
# This function inputs a numeric variable x
# and outputs x+1.
# If x is a non-negative integer, then it outputs the next positive integer.
  return(x+1)
}
```

B. More Guidelines

B.1. Titlepage

> " The title page is your own design however it should include your name, CID, project title, supervisor's name. You may want to include the wording: "Submitted in partial fulfilment of the requirements for the MSc in Statistics of Imperial College London". You should not be using the Imperial crest, but you can use the Imperial logo:
>
> `http://www3.imperial.ac.uk/graphicidentity/applyingthegraphicidentity/usingthecrest`
>
> `http://www3.imperial.ac.uk/graphicidentity/applyingthegraphicidentity/usingthelogo`"

The logo has been included in the current title design but must be downloaded into the same folder as your `*.tex` file.

B.2. Declaration

> " The second page must contain a signed and dated plagiarism statement, "The work contained in this thesis is my own work unless otherwise stated". It is sufficient if you sign the hard copies."