Mixture models and stick-breaking processes

(Applications in Financial Econometrics)

Maria Kalli

PhD Candidate, IMSAS,
University of Kent

7 August 2007

Kick off

- Financial Time Series
- Mixture Models
- Stick Breaking Processes SBP
- Slice Sampler

Kick off

- Financial Time Series
- Mixture Models
- Stick Breaking Processes SBP
- Slice Sampler

Kick off

- Financial Time Series
- Mixture Models
- Stick Breaking Processes SBP
- Slice Sampler

Kick off

- Financial Time Series
- Mixture Models
- Stick Breaking Processes SBP
- Slice Sampler

Characteristics of Financial Time Series

- Non Constant Volatility
  1. clustering
  2. continuous evolution
  3. jumps are rare
  4. varies within a fixed range
- Outliers, which lead to heavy tails
- GARCH is one of the dominant models because:
  1. it captures heavy tails and varying volatility
  2. the marginal mean and variance are constant

Characteristics of Financial Time Series

- Non Constant Volatility
  1. clustering
  2. continuous evolution
  3. jumps are rare
  4. varies within a fixed range
- Outliers, which lead to heavy tails
- GARCH is one of the dominant models because:
  1. it captures heavy tails and varying volatility
  2. the marginal mean and variance are constant

Characteristics of Financial Time Series

- Non Constant Volatility
  1. clustering
  2. continuous evolution
  3. jumps are rare
  4. varies within a fixed range
- Outliers, which lead to heavy tails
- GARCH is one of the dominant models because:
  1. it captures heavy tails and varying volatility
  2. the marginal mean and variance are constant

Characteristics of Financial Time Series

- Non Constant Volatility
  1. clustering
  2. continuous evolution
  3. jumps are rare
  4. varies within a fixed range

- Outliers, which lead to heavy tails

- GARCH is one of the dominant models because:
  1. it captures heavy tails and varying volatility
  2. the marginal mean and variance are constant

Characteristics of Financial Time Series

- Non Constant Volatility
  1. clustering
  2. continuous evolution
  3. jumps are rare
  4. varies within a fixed range

- Outliers, which lead to heavy tails

- GARCH is one of the dominant models because:
  1. it captures heavy tails and varying volatility
  2. the marginal mean and variance are constant

Characteristics of Financial Time Series

- Non Constant Volatility
  1. clustering
  2. continuous evolution
  3. jumps are rare
  4. varies within a fixed range
- Outliers, which lead to heavy tails
- GARCH is one of the dominant models because:
  1. it captures heavy tails and varying volatility
  2. the marginal mean and variance are constant

Characteristics of Financial Time Series

- Non Constant Volatility
  1. clustering
  2. continuous evolution
  3. jumps are rare
  4. varies within a fixed range
- Outliers, which lead to heavy tails
- GARCH is one of the dominant models because:
  1. it captures heavy tails and varying volatility
  2. the marginal mean and variance are constant

Stick Breaking Processes (SBP)

- Bayesian Mixture Modeling: The hierarchical structure

1. $y|\theta \sim f(y|\theta)$
2. $\theta|G \sim G$
3. $G \sim SBP(G_0)$

Where level (1) is the likelihood function, level (2) represents the unknown distribution of the components and level (3) places the prior on G, which is normally based on some stick breaking process (SBP) with a centering measure G_0

Stick Breaking Processes (SBP)

- Bayesian Mixture Modeling: The hierarchical structure

1. $y|\theta \sim f(y|\theta)$
2. $\theta|G \sim G$
3. $G \sim SBP(G_0)$

Where level (1) is the likelihood function, level (2) represents the unknown distribution of the components and level (3) places the prior on G, which is normally based on some stick breaking process (SBP) with a centering measure G_0

Stick Breaking Processes (SBP)

- Bayesian Mixture Modeling: The hierarchical structure

1. $y|\theta \sim f(y|\theta)$
2. $\theta|G \sim G$
3. $G \sim SBP(G_0)$

Where level (1) is the likelihood function, level (2) represents the unknown distribution of the components and level (3) places the prior on G, which is normally based on some stick breaking process (SBP) with a centering measure G_0

Stick Breaking Processes (SBP)

- Bayesian Mixture Modeling: The hierarchical structure

  1. $y|\theta \sim f(y|\theta)$
  2. $\theta|G \sim G$
  3. $G \sim SBP(G_0)$

  Where level (1) is the likelihood function, level (2) represents the unknown distribution of the components and level (3) places the prior on G, which is normally based on some stick breaking process (SBP) with a centering measure G_0

Stick Breaking Processes (SBP)

Sethuraman in his 1994 paper represented the Dirichlet Process (DP) in this way. Later extended to the Pitman- Yor process (see Pitman and Yor 1997).

The set up is : $G = \sum_j w_j \delta_{\theta_j}(.)$. Where δ_{θ_j} is the dirac measure at θ_j. The components, θ_j 's are iid from G_0. Each w_j represents the weight/proportion of component θ_j.

The w_j's are constructed as follows :
$w_j = v_j(1 - v_{j-1})...(1 - v_1)$ *where* $v_j \sim Beta(\alpha_j, \beta_j)$
e.g for Dirichlet Process $\alpha_j = 1$, *and* $\beta_j = c$ for Pitman-Yor Process $\alpha_j = 1 - a$, *and* $\beta_j = b + ja$

Stick Breaking Processes (SBP)

Sethuraman in his 1994 paper represented the Dirichlet Process (DP) in this way. Later extended to the Pitman- Yor process (see Pitman and Yor 1997).

The set up is : $G = \sum_j w_j \delta_{\theta_j}(.)$. Where δ_{θ_j} is the dirac measure at θ_j. The components, θ_j 's are iid from G_0. Each w_j represents the weight/proportion of component θ_j.

The w_j's are constructed as follows :
$w_j = v_j(1 - v_{j-1})...(1 - v_1)$ *where* $v_j \sim Beta(\alpha_j, \beta_j)$
e.g for Dirichlet Process $\alpha_j = 1$, *and* $\beta_j = c$ for Pitman-Yor Process $\alpha_j = 1 - a$, *and* $\beta_j = b + ja$

Stick Breaking Processes (SBP)

Sethuraman in his 1994 paper represented the Dirichlet Process (DP) in this way. Later extended to the Pitman- Yor process (see Pitman and Yor 1997).

The set up is : $G = \sum_j w_j \delta_{\theta_j}(.)$. Where δ_{θ_j} is the dirac measure at θ_j. The components, θ_j 's are iid from G_0. Each w_j represents the weight/proportion of component θ_j.

The w_j's are constructed as follows :
$w_j = v_j(1 - v_{j-1})...(1 - v_1)$ *where* $v_j \sim Beta(\alpha_j, \beta_j)$
e.g for Dirichlet Process $\alpha_j = 1$, *and* $\beta_j = c$ for Pitman-Yor Process $\alpha_j = 1 - a$, *and* $\beta_j = b + ja$

Stick Breaking Processes (SBP)

Sethuraman in his 1994 paper represented the Dirichlet Process (DP) in this way. Later extended to the Pitman- Yor process (see Pitman and Yor 1997).

The set up is : $G = \sum_j w_j \delta_{\theta_j}(.)$. Where δ_{θ_j} is the dirac measure at θ_j. The components, θ_j 's are iid from G_0. Each w_j represents the weight/proportion of component θ_j.

The w_j's are constructed as follows :
$w_j = v_j(1 - v_{j-1})...(1 - v_1)$ *where* $v_j \sim Beta(\alpha_j, \beta_j)$
e.g for Dirichlet Process $\alpha_j = 1$, *and* $\beta_j = c$ for Pitman-Yor Process $\alpha_j = 1 - a$, *and* $\beta_j = b + ja$

Mixture Models with SBP

The model just described gives : $f(y) = \sum_{j=1}^{\infty} w_j f(y|\theta_j)$

θ_j is a mixture component, w_j is the probability attached to each component and j is the component index (it can be infinitely countable). $f(y|\theta_j)$ is some chosen continuous density.

Mixture Models with SBP

The model just described gives : $f(y) = \sum_{j=1}^{\infty} w_j f(y|\theta_j)$

θ_j is a mixture component, w_j is the probability attached to each component and j is the component index (it can be infinitely countable). $f(y|\theta_j)$ is some chosen continuous density.

The Slice Sampler

- Since the proposed model adopts SBP methods, a way to deal with the infinite sum is to introduce slice sampling. It is a natural way of truncating.
- It is a Gibbs sampler, which with the help of an auxiliary variable, helps us sample from a non standard density, $\pi(y)$ *on* R^d
- The draws are then from a joint density $\pi(y, u) = I_{\{\pi(y)>u\}}$. Considering only the y values we have draws from $\pi(y)$

The Slice Sampler

- Since the proposed model adopts SBP methods, a way to deal with the infinite sum is to introduce slice sampling. It is a natural way of truncating.
- It is a Gibbs sampler, which with the help of an auxiliary variable, helps us sample from a non standard density, $\pi(y)$ *on* R^d
- The draws are then from a joint density $\pi(y, u) = I_{\{\pi(y) > u\}}$. Considering only the y values we have draws from $\pi(y)$

The Slice Sampler

- Since the proposed model adopts SBP methods, a way to deal with the infinite sum is to introduce slice sampling. It is a natural way of truncating.
- It is a Gibbs sampler, which with the help of an auxiliary variable, helps us sample from a non standard density, $\pi(y)$ *on* R^d
- The draws are then from a joint density $\pi(y, u) = I_{\{\pi(y) > u\}}$. Considering only the y values we have draws from $\pi(y)$

The Slice Sampler enters the pitch

Under the infinite sum construction of the SBP

$$f(y) = \sum_j w_j \, f(y|\theta_j)$$

Introducing the auxiliary variable u and considering

$f(y, u|\theta, w) = \sum_j I\{u < w_j\} \, f(y|\theta_j)$ we have a finite sum.

Introducing auxiliary variable κ we have :

$$f(y, u, \kappa) = I\{u < w_\kappa\} \, f(y|\theta_\kappa) = w_\kappa \, U(u|0, w_\kappa) \, f(y|\theta_\kappa)$$

The Slice Sampler enters the pitch

Under the infinite sum construction of the SBP

$$f(y) = \sum_j w_j \, f(y|\theta_j)$$

Introducing the auxiliary variable u and considering

$$f(y, u|\theta, w) = \sum_j I\{u < w_j\} \, f(y|\theta_j)$$ we have a finite sum.

Introducing auxiliary variable κ we have :

$$f(y, u, \kappa) = I\{u < w_\kappa\} \, f(y|\theta_\kappa) = w_\kappa \, U(u|0, w_\kappa) \, f(y|\theta_\kappa)$$

The Slice Sampler enters the pitch

Under the infinite sum construction of the SBP

$$f(y) = \sum_j w_j f(y|\theta_j)$$

Introducing the auxiliary variable u and considering

$$f(y, u|\theta, w) = \sum_j I\{u < w_j\} f(y|\theta_j)$$ we have a finite sum.

Introducing auxiliary variable κ we have :

$$f(y, u, \kappa) = I\{u < w_\kappa\} f(y|\theta_\kappa) = w_\kappa U(u|0, w_\kappa) f(y|\theta_\kappa)$$

The Slice Sampler enters the pitch

Under the infinite sum construction of the SBP

$$f(y) = \sum_{j} w_j \, f(y|\theta_j)$$

Introducing the auxiliary variable u and considering

$$f(y,u|\theta,w) = \sum_{j} I\{u < w_j\} \, f(y|\theta_j) \text{ we have a finite sum.}$$

Introducing auxiliary variable κ we have :

$$f(y,u,\kappa) = I\{u < w_\kappa\} \, f(y|\theta_\kappa) = w_\kappa \, U(u|0,w_\kappa) \, f(y|\theta_\kappa)$$

The Hierarchical Structure of the model is:

1. $y|(\theta,\kappa) \sim f(y|\theta_\kappa)$
2. $u|\kappa, w \sim U(0, w_\kappa)$
3. $f(\kappa|w) = w_\kappa$

The n observation model likelihood will be:

$$\ell(w,\kappa,\theta|y) = \prod_{i=l}^{n} f(y_i, u_i, \kappa_i|\theta, w) = \prod_{i=l}^{n} w_{\kappa_i}\, U(u_i|0, w_{\kappa_i})\, f(y_i|\theta_{\kappa_i})$$

The Hierarchical Structure of the model is:

1. $y|(\theta,\kappa) \sim f(y|\theta_\kappa)$
2. $u|\kappa, w \sim U(0, w_\kappa)$
3. $f(\kappa|w) = w_\kappa$

The n observation model likelihood will be:

$$\ell(w,\kappa,\theta|y) = \prod_{i=l}^{n} f(y_i,u_i,\kappa_i|\theta,w) = \prod_{i=l}^{n} w_{\kappa_i}\, U(u_i|0,w_{\kappa_i})\, f(y_i|\theta_{\kappa_i})$$

The Hierarchical Structure of the model is:

1. $y|(\theta,\kappa) \sim f(y|\theta_\kappa)$
2. $u|\kappa, w \sim U(0, w_\kappa)$
3. $f(\kappa|w) = w_\kappa$

The n observation model likelihood will be:

$$\ell(w,\kappa,\theta|y) = \prod_{i=l}^{n} f(y_i, u_i, \kappa_i|\theta, w) = \prod_{i=l}^{n} w_{\kappa_i} U(u_i|0, w_{\kappa_i}) f(y_i|\theta_{\kappa_i})$$

The Hierarchical Structure of the model is:

1. $y|(\theta,\kappa) \sim f(y|\theta_\kappa)$
2. $u|\kappa, w \sim U(0, w_\kappa)$
3. $f(\kappa|w) = w_\kappa$

The n observation model likelihood will be:

$$\ell(w,\kappa,\theta|y) = \prod_{i=l}^{n} f(y_i, u_i, \kappa_i|\theta, w) = \prod_{i=l}^{n} w_{\kappa_i}\, U(u_i|0, w_{\kappa_i})\, f(y_i|\theta_{\kappa_i})$$

The Data Kernel: $y|(\theta, \kappa, u, w) \sim f(y|\theta_\kappa)$

The choice is open to the researcher based on her/his knowledge of the data she/he is analyzing. Since I am dealing with stock returns my choices are the following:

1. A uniform kernel preserves the symmetry and unimodality the financial data exhibit. The choice is : $f(y|\theta_\kappa) = U(-\theta_\kappa, \theta_\kappa)$
2. The flexibility of this kernel choice allows for the inclusion of skewness.The kernel then becomes: $f(y|\theta_\kappa, \lambda) = U(-\theta_\kappa e^{-\lambda}, \theta_\kappa e^{\lambda})$
3. Finally the model is complete with the inclusion of the regression part in the form of $e^{\beta x_i}$ *where* $x_i = y_{i-1}^2$. Therefore the kernel is: $f(y|\theta_\kappa, \lambda, \beta) = U(-\theta_\kappa e^{-\lambda+\beta x_i}, \theta_\kappa e^{\lambda+\beta x_i})$ ARCH(1) type

The Data Kernel: $y|(\theta,\kappa,u,w) \sim f(y|\theta_\kappa)$

The choice is open to the researcher based on her/his knowledge of the data she/he is analyzing. Since I am dealing with stock returns my choices are the following:

1. A uniform kernel preserves the symmetry and unimodality the financial data exhibit. The choice is : $f(y|\theta_\kappa) = U(-\theta_\kappa,\theta_\kappa)$
2. The flexibility of this kernel choice allows for the inclusion of skewness.The kernel then becomes: $f(y|\theta_\kappa,\lambda) = U(-\theta_\kappa e^{-\lambda},\theta_\kappa e^{\lambda})$
3. Finally the model is complete with the inclusion of the regression part in the form of $e^{\beta x_i}$ *where* $x_i = y_{i-1}^2$. Therefore the kernel is: $f(y|\theta_\kappa,\lambda,\beta) = U(-\theta_\kappa e^{-\lambda+\beta x_i},\theta_\kappa e^{\lambda+\beta x_i})$ ARCH(1) type

The Data Kernel: $y|(\theta, \kappa, u, w) \sim f(y|\theta_\kappa)$

The choice is open to the researcher based on her/his knowledge of the data she/he is analyzing. Since I am dealing with stock returns my choices are the following:

1. A uniform kernel preserves the symmetry and unimodality the financial data exhibit. The choice is : $f(y|\theta_\kappa) = U(-\theta_\kappa, \theta_\kappa)$
2. The flexibility of this kernel choice allows for the inclusion of skewness.The kernel then becomes:
$f(y|\theta_\kappa, \lambda) = U(-\theta_\kappa e^{-\lambda}, \theta_\kappa e^{\lambda})$
3. Finally the model is complete with the inclusion of the regression part in the form of $e^{\beta x_i}$ *where* $x_i = y_{i-1}^2$. Therefore the kernel is:
$f(y|\theta_\kappa, \lambda, \beta) = U(-\theta_\kappa e^{-\lambda+\beta x_i}, \theta_\kappa e^{\lambda+\beta x_i})$ ARCH(1) type

The Data Kernel: $y|(\theta, \kappa, u, w) \sim f(y|\theta_\kappa)$

The choice is open to the researcher based on her/his knowledge of the data she/he is analyzing. Since I am dealing with stock returns my choices are the following:

1. A uniform kernel preserves the symmetry and unimodality the financial data exhibit. The choice is : $f(y|\theta_\kappa) = U(-\theta_\kappa, \theta_\kappa)$
2. The flexibility of this kernel choice allows for the inclusion of skewness.The kernel then becomes: $f(y|\theta_\kappa, \lambda) = U(-\theta_\kappa e^{-\lambda}, \theta_\kappa e^{\lambda})$
3. Finally the model is complete with the inclusion of the regression part in the form of $e^{\beta x_i}$ *where* $x_i = y_{i-1}^2$. Therefore the kernel is: $f(y|\theta_\kappa, \lambda, \beta) = U(-\theta_\kappa e^{-\lambda+\beta x_i}, \theta_\kappa e^{\lambda+\beta x_i})$ ARCH(1) type

Why slice-sampling helps

- Sampling the v_j's and in consequence the w_j's involves the infinite sum.
- Introducing the auxiliary variable u, and imposing the condition $u < w_j$, the sets $A_w(u) = \{j : w_j > u\}$ are formed. This means that only a finite number of the v's is needed to move on to the next iteration.
  - All the v_j's and θ_j's whose index is not in an $A_w(u)$ set are drawn independently from their respective priors

So the sampler is:

1. $\underline{\theta}$ | everything else
2. $\underline{\text{v}}$ | everything else
3. $\underline{\kappa}$ | everything else
4. $\underline{\text{u}}$ | everything else

Why slice-sampling helps

- Sampling the v_j's and in consequence the w_j's involves the infinite sum.
- Introducing the auxiliary variable u, and imposing the condition $u < w_j$, the sets $A_w(u) = \{j : w_j > u\}$ are formed. This means that only a finite number of the v's is needed to move on to the next iteration.
  - All the v_j's and θ_j's whose index is not in an $A_w(u)$ set are drawn independently from their respective priors

So the sampler is:

1. $\underline{\theta}$ | everything else
2. $\underline{v}$ | everything else
3. $\underline{\kappa}$ | everything else
4. $\underline{u}$ | everything else

Why slice-sampling helps

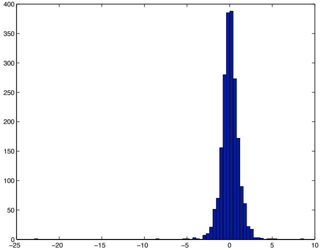
- Sampling the v_j's and in consequence the w_j's involves the infinite sum.
- Introducing the auxiliary variable u, and imposing the condition $u < w_j$, the sets $A_w(u) = \{j : w_j > u\}$ are formed. This means that only a finite number of the v's is needed to move on to the next iteration.
  - All the v_j's and θ_j's whose index is not in an $A_w(u)$ set are drawn independently from their respective priors

So the sampler is:

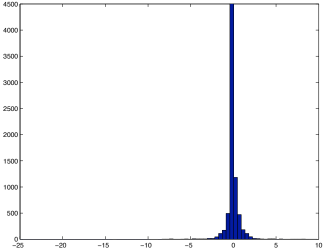
1. $\underline{\theta}$ | everything else
2. $\underline{\text{v}}$ | everything else
3. $\underline{\kappa}$ | everything else
4. $\underline{\text{u}}$ | everything else

SP data set

The real data

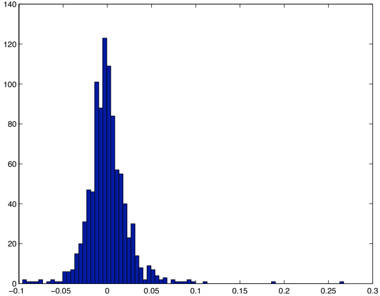


Predictive after 8000 iterations

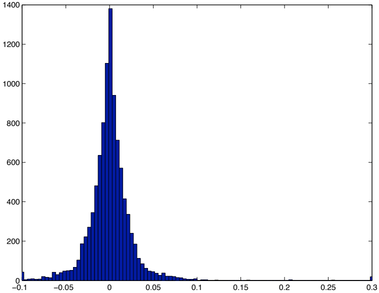


CSE data set

The real data

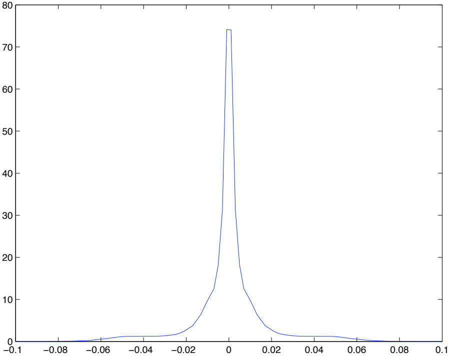


Predictive after 10000 iterations



CSE data set

Density estimate:



Conclusions

1. The Pitman-Yor stick breaking process with the uniform kernel is a flexible model.
2. The slice sampler is a simple and efficient method of truncation for any Stick Breaking Process Mixture.
3. Based on the results so far the predictive distributions have captured the characteristics of the actual data sets.

Conclusions

1. The Pitman-Yor stick breaking process with the uniform kernel is a flexible model.
2. The slice sampler is a simple and efficient method of truncation for any Stick Breaking Process Mixture.
3. Based on the results so far the predictive distributions have captured the characteristics of the actual data sets.

Conclusions

1. The Pitman-Yor stick breaking process with the uniform kernel is a flexible model.
2. The slice sampler is a simple and efficient method of truncation for any Stick Breaking Process Mixture.
3. Based on the results so far the predictive distributions have captured the characteristics of the actual data sets.

Future Work

- Proceed to inference analysis
- Introduce more covariates
- Localize the skewness,i.e. link it to each observation
- Explore the bivariate case

Future Work

- Proceed to inference analysis
- Introduce more covariates
- Localize the skewness,i.e. link it to each observation
- Explore the bivariate case

Future Work

- Proceed to inference analysis
- Introduce more covariates
- Localize the skewness,i.e. link it to each observation
- Explore the bivariate case

Future Work

- Proceed to inference analysis
- Introduce more covariates
- Localize the skewness,i.e. link it to each observation
- Explore the bivariate case