Bayesian Modeling of Conditional Distributions

John Geweke
University of Iowa
Indiana University
Department of Economics
February 27, 2007
Outline

- Motivation
- Model description
- Methods of inference
- Earnings example
- Asset returns example
- Conclusion
The general case and an example

\[
\begin{align*}
\{ x_t, y_t \}_{n \times 1} & \text{ i.i.d.} \\
p(y_t | x_t) &= ?
\end{align*}
\]

Example: Earnings

\[
\begin{align*}
y_t & \quad \text{Earnings or wage of individual } t \\
x_{1t} & \quad \text{Experience or age of individual } t \\
x_{2t} & \quad \text{Education of individual } t
\end{align*}
\]
A second example: Asset returns

\[ y_t = \text{Return on asset in period } t \]

\[ y_t \mid (y_1, \ldots, y_{t-1}) \sim? \]

The covariates are

- \( x_{1t} \)  \( \text{Return on asset in period } t - 1, \ x_{1t} = y_{t-1} \)

- \( x_{2t} \)  \( \text{Recent volatility,} \)

\[
g \cdot x_{2,t-1} + (1 - g) |a_{t-1}|^\kappa = \sum_{s=0}^{\infty} g^s |y_{t-2-s}|^\kappa
\]

\((g = 0.95, \ \kappa = 1)\)
The common structure of normal mixture models

\[ y_t \quad \text{Variable of interest} \]

\[ x_t \quad \text{Vector of covariates} \quad n \times 1 \]

\[ \tilde{s}_t \quad \text{Latent state, } \tilde{s}_t \in \{1, \ldots, m\} \]

\[ y_t \mid (x_t, \tilde{s}_t = j) \sim N \left( \beta_j' x_t, \sigma_j^2 \right) \]

\[ y_t \mid (x_t, \tilde{s}_t = j) \sim N \left[ \left( \beta + \alpha_j \right)' x_t, \left( h \cdot h_j \right)^{-1} \right] \]
A simple special case:

The normal mixture model with no covariates

\[
x_t \begin{pmatrix} 1 \\ 1 \end{pmatrix} = 1
\]

\( \tilde{s}_t \) i.i.d., \( P(\tilde{s}_t = j) = p_j \)

\[
y_t \mid (x_t = 1, \tilde{s}_t = j) \sim N\left(\beta_j, \sigma_j^2\right)
\]

\[
y_t \mid (x_t = 1, \tilde{s}_t = j) \sim N\left[\beta + \alpha_j, (h \cdot h_j)^{-1}\right]
\]
Bayesian Modeling of Conditional Distributions

- Model description
- Mixture modeling
What is new in this work

- Begin with the same normal mixture model

\[ y_t \mid (x_t, \tilde{s}_t = j) \sim N(\beta'_j x_t, \sigma^2_j) \]

- Determination of latent states \( \tilde{s}_t \):
  - \( \tilde{w}_t = \Gamma x_t + \zeta_t; \quad \zeta_t \overset{iid}{\sim} N(0, I_m) \)
  - \( \tilde{s}_t = j \) iff \( \tilde{w}_{tj} \geq \tilde{w}_{ti} \ \forall \ i = 1, \ldots, m \)
  - \( \iff \tilde{s}_t = \arg\max_{i=1,\ldots,m} (\tilde{w}_{ti}) \)
Bayesian Modeling of Conditional Distributions

- Model description
- Literature context

## Literature context

- Bayesian nonparametric regression (e.g. Wahba, Shiller, Smith & Kohn)
- Non-Bayesian nonparametric regression (e.g. Härdle & Tsybakov)
- Quantile regression (e.g. Koenker & Basset, Yu & Jones)
- Dirichlet process priors (e.g. Griffin & Steel, Dunson & Pilai)
- Mixture of Experts Models (e.g. Jacobs & Jordan, Jian and Tanner)
Identification issues: Multinomial probit

\[ \tilde{w}_t = \Gamma x_t + \zeta_t , \quad j = \arg \max_{i=1,\ldots,m} (\tilde{w}_{ti}), \quad y_t \sim N \left( \left( \beta + \alpha_j \right)' x_t, (h \cdot h_j)^{-1} \right) \]

Because \( \zeta_t \sim iid \ N(0, I_m) \) translation but not scaling issues arise in

\[ \tilde{w}_t = \Gamma x_t + \zeta_t \]

Impose \( \iota'_m \Gamma = 0 \) through

\[
\begin{align*}
\Gamma_{m \times q} &= P \cdot \begin{bmatrix} 0'_q \\ \Gamma^* \\ (m-1) \times q \end{bmatrix}, & P_{m \times m} &= \begin{bmatrix} \iota_m \cdot m^{-1/2} & P_2 \\ m \times (m-1) \end{bmatrix}, & P'P &= I_n
\end{align*}
\]
Identification issues: Labeling

\[ \tilde{w}_t = \Gamma x_t + \zeta_t, \quad j = \arg \max_{i=1, \ldots, m} (\tilde{w}_t^i), \quad y_t \sim N \left( (\beta + \alpha_j)' x_t, (h \cdot h_j)^{-1} \right) \]

- States have no substantive interpretation and are exchangeable
  - implications for prior distribution
- \( m! \) state permutations
  - \( \implies m! \) reflections of the posterior distribution
- Not to worry: \( p(y_t \mid x_t) \) invariant to permutations of the states
- Geweke (2006) CSDA forthcoming
Conditionally conjugate prior distributions

\[ \tilde{w}_t = \Gamma x_t + \zeta_t, \quad j = \arg \max_{i=1,...,m} (\tilde{w}_{ti}), \quad y_t \sim N \left( (\beta + \alpha_j) \right)^t x_t, \frac{1}{(h \cdot h_j)^{-1}} \]

<table>
<thead>
<tr>
<th>Distribution type</th>
<th>Parameters</th>
<th>Hyperparameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaussian:</td>
<td>( \beta, \Gamma^* )</td>
<td>( \mu, \tau^2_{\beta}, \tau^2_{\gamma} )</td>
</tr>
<tr>
<td>Gaussian conditional on ( h ):</td>
<td>( \alpha_j (j = 1, \ldots, m) )</td>
<td>( \tau^2_{\alpha} )</td>
</tr>
<tr>
<td>Inverse gamma:</td>
<td>( h; h_j (j = 1, \ldots, m) )</td>
<td>( s^2, \nu, \nu^* )</td>
</tr>
</tbody>
</table>
Some theory: Conditions

1. Distribution of $\mathbf{x}$
   - $\mathbf{x} \in \Omega \subseteq \mathbb{R}^n$, $\Omega$ compact
   - $p(\mathbf{x}) > 0$ is continuous w.r.t. Lebesgue measure.

2. Conditional distribution of $\mathbf{y} \in A$
   - For some function $h(\mathbf{x})$, $p(\mathbf{y} | \mathbf{x})$ is an exponential distribution,
     
     $$p(\mathbf{y} | \mathbf{x}) = \exp \{ a[h(\mathbf{x})] \mathbf{y} + b[h(\mathbf{x})] + c(\mathbf{y}) \} \quad \forall \mathbf{x} \in \Omega, \mathbf{y} \in A;$$

     $a(\cdot)$ and $b(\cdot)$ are analytic, $a'(\cdot) \neq 0$, $b'(\cdot) \neq 0$.

   - $h$ is in a ball of finite values in a Sobolev space with sup-norm and second-order differentiability:
     - $\sup_{\mathbf{x} \in \Omega} |h(\mathbf{x})| = c_0 < \infty$
     - $\sup_{i=1,\ldots,n} (\sup_{\mathbf{x} \in \Omega} |\partial h(\mathbf{x}) / \partial x_i|) = c_1 < \infty$
     - $\sup_{i=1,\ldots,n} \left[ \sup_{j=1,\ldots,n} (\sup_{\mathbf{x} \in \Omega} |\partial^2 h(\mathbf{x}) / \partial x_i \partial x_j|) \right] = c_2 < \infty$
Some theory: A definition

- **Definition.** For any two pdf’s \( f ( \cdot ) \) and \( g ( \cdot ) \) w.r.t. Lebesgue measure defined on \( Z \), the Kullback-Leibler directed distance from \( f \) to \( g \) is

\[
KL ( f , g ) = \int_Z f ( z ) \log \left[ \frac{f ( z )}{g ( z )} \right] dz.
\]

- **Remark.** In a model

\[
p ( y_t \mid \theta_A , A ) ( t = 1 , \ldots , T )
\]

let \( T \to \infty \). The posterior density of \( \theta_A \) will collapse about \( \theta_A = \theta_A^* \), where

\[
\theta_A^* = \arg \min_{\theta_A} KL [ p ( y ) , p ( y \mid \theta_A , A ) ]
\]

(Geweke, 2005, Theorem 3.4.2).
Some theory: A result

**Theorem.** For all \( x \in \Omega \),

\[
\sup_{p(y|x)} \inf_{p(y|x,A_m)} KL \left[ p(y \mid x) , p(y \mid x,A_m) \right] < \frac{c}{m^4/n}
\]

where \( c \) depends on \( p(x) \), \( c_0 \), \( c_1 \) and \( c_2 \), but does not depend on \( m \) or \( n \).

**Remarks**

- The result is a consequence of Jiang and Tanner (1999) Theorem 2 for the class of gating functions defined by the multinomial probit model with scalar variance matrix.
- Condition 1 of Jiang and Tanner (1999) for gating functions is important but not given here. They conjecture that it holds for multinomial logit gating functions. The proof for multinomial probit gating functions is not difficult.
Blocking for Gibbs sampling

\[ \tilde{w}_t = \Gamma x_t + \zeta_t, \quad j = \arg \max_{i=1, \ldots, m} (\tilde{w}_{ti}), \quad y_t \sim N \left[ (\beta + \alpha_j)' x_t, (h \cdot h_j)^{-1} \right] \]

\[ h, h_1, \ldots, h_m \quad \text{Separately conditionally independent gamma} \]

\[ \beta, \alpha_1, \ldots, \alpha_m \quad \text{Jointly conditionally Gaussian} \]

\[ \text{vec} (\Gamma^*) \quad \text{Conditionally Gaussian} \]

\[ \tilde{w}_t \quad \text{Gaussian times orthant-specific likelihood factors} \]

Conditional posteriors and code have been tested (Geweke 2004).
Quantile functions of interest

- **Conditional CDFs:** \( P (y_t \leq c \mid x_t) \)
- **Quantiles:** \( c (q) = \{ c : P (y_t \leq c \mid x_t) = q \} \)
- **Note** \( P (\tilde{s}_t = j \mid \Gamma , x_t) \)
  - \( = P [\tilde{w}_{tj} \geq \tilde{w}_{ti} \ (i = 1, \ldots, m) \mid \Gamma , x_t] \)
  - \( = \int_{-\infty}^{\infty} p (\tilde{w}_{tj} = w^* \mid \Gamma , x_t) \cdot P [\tilde{w}_{ti} \leq w^* \ (i = 1, \ldots, m) \mid \Gamma , x_t] \ dw^* \)
  - \( = \int_{-\infty}^{\infty} \phi \left( w^* - \gamma'_j z_t \right) \ \prod_{i \neq j} \Phi \left( y - \gamma'_i z_t \right) \ dw^*. \)

- **Given** \( M \) Markov chain Monte Carlo replications of a mixture model with \( m \) components, the posterior distribution is a mixture of normals with \( M \cdot m \) components.
Earnings example

Data

- 2698 men from PSID, interviewed in 1994, data pertain to 1993
- $25 \leq Age \leq 65$
- Not black
- Earnings exceed $1,000
Prior distribution hyperparameters

- Gaussian priors:
  \[ \beta: \mu = 10, \quad \tau^2_\beta = 1 \]
  \[ \alpha: \mu = 0, \quad \tau^2_\alpha = 4 \]
  \[ \Gamma^* \mu = 0, \quad \tau^2_\gamma = 16 \]

- Inverse gamma priors:
  \[ 2h \sim \chi^2(2) \]
  \[ 2h_j \sim \chi^2(2) \]

- Information contribution from prior relative to data is minute.
Comparison of alternative model specifications: Modified cross-validated log scores

- Randomly permute all 2698 observations (just to be safe; works for time series too)
- Select first $T_1 = 2153$ for inference (full MCMC, gives $\theta^{(m)}$ ($m = 1, \ldots, M = 10^4$)
- Modified cross-validated log scoring rule (Draper and Krnjajic (2005)):

$$
\sum_{t=T_1+1}^{T} \log \left[ p \left( y_t \mid x_t, Y_{T_1}, SMR \right) \right],
$$

$$
p \left( y_t \mid x_t, Y_{T_1}, SMR \right) \approx M^{-1} \sum_{m=1}^{M} p \left( y_t \mid \theta^{(m)}, x_t, SMR \right)
$$
Earnings example

Earnings quantiles: Model with $m = 8$ mixture components
Bayesian Modeling of Conditional Distributions

Earnings example

Median of conditional posterior earnings distribution
Bayesian Modeling of Conditional Distributions

Earnings example

Median of conditional posterior earnings distribution
Bayesian Modeling of Conditional Distributions

Earnings example

Median of conditional posterior earnings distribution

Model with 7 mixture components

Model with 8 mixture components

Model with 9 mixture components

Model with 10 mixture components
Bayesian Modeling of Conditional Distributions

Earnings example

- Earnings quantiles: Model with \( m = 8 \) mixture components

**Posterior earnings at quantile 0.05**

**Posterior earnings at quantile 0.10**

**Posterior earnings at quantile 0.25**

**Posterior earnings at quantile 0.50**
Bayesian Modeling of Conditional Distributions

Earnings example

Earnings quantiles: Model with $m = 8$ mixture components

![Graphs showing posterior earnings at quantiles 0.50, 0.75, 0.90, and 0.95 for education and age.](image-url)
Bayesian Modeling of Conditional Distributions

Earnings example

Earnings quantiles: Model with m = 8 mixture components
Asset Returns Example

The data

- Standard and Poors 500 daily open ($O_t$) and close ($C_t$), 1990-1999
  
  \[(C_{t-1} = O_t)\]

- Dependent variable: Percent log return,
  \[y_t = 100 \cdot \log \left( \frac{C_t}{O_t} \right)\]

- Covariates:
  
  \[
  \begin{align*}
  x_{1t} &= y_{t-1} \\
  x_{2t} &= g \cdot x_{2,t-1} + (1 - g) |a_{t-1}|^\kappa = \sum_{s=0}^{\infty} g^s |y_{t-2-s}|^\kappa
  \end{align*}
  \]

- The evidence favors $g = 0.95$ and $\kappa = 1$. 
The model: Same as in Geweke and Keane (2006)

\[ \tilde{w}_t = \Gamma x_t + \zeta_t, \quad j = \arg \max_{i=1,...,m} (\tilde{w}_{ti}), \quad y_t \sim N \left[ \left( \beta + \alpha_j \right)' x_t, \left( h_h j \right)^{-1} \right] \]

- In Geweke and Keane (2006) linear functions of \( x_t \) are replaced by polynomial functions of \( x_t \).
- Based on the evidence (Modified cross-validated log scores)
  - \( \Gamma x_t \) become second order polynomials in \( x_t \)
  - \( \left( \beta + \alpha_j \right)' x_t \) become zero-order polynomials in \( x_t \).
- Mixture of \( m = 3 \) components
- Work with linear functions of \( x_t \) is currently proceeding.
Model comparison: Predictive likelihoods (Bayesian) and Recursive maximum likelihood (non-Bayesian)


<table>
<thead>
<tr>
<th>Model</th>
<th>Log Predictive likelihood</th>
<th>Log Recursive ML</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal iid</td>
<td>-1848.5</td>
<td>-1848.5</td>
</tr>
<tr>
<td>GARCH(1,1)</td>
<td>-1660.5</td>
<td>-1660.5</td>
</tr>
<tr>
<td>EGARCH(1,1)</td>
<td>-1637.5</td>
<td>-1637.5</td>
</tr>
<tr>
<td>t-GARCH(1,1)</td>
<td>-1625.5</td>
<td>-1624.7</td>
</tr>
<tr>
<td>Stochastic volatility</td>
<td>-1625.4</td>
<td>-1625.4</td>
</tr>
<tr>
<td>SMR</td>
<td>-1622.0</td>
<td>-1622.0</td>
</tr>
</tbody>
</table>

- Dynamic predictive properties of t-GARCH and stochastic volatility are similar.
- Dynamic predictive properties of SMR are quite different from t-GARCH and stochastic volatility.
Bayesian Modeling of Conditional Distributions

Asset Returns Example

Posterior means of population moments
Bayesian Modeling of Conditional Distributions

Asset Returns Example

Posterior means of population moments

Posterior expectation of standard deviation of return distribution

Return yesterday
Bayesian Modeling of Conditional Distributions

Asset Returns Example

Posterior means of population moments

Posterior expectation of skewness of return distribution
Bayesian Modeling of Conditional Distributions

Asset Returns Example

Posterior means of population moments
Bayesian Modeling of Conditional Distributions

Asset Returns Example

Posterior quantiles

Posterior S&P return at quantile 0.05
Bayesian Modeling of Conditional Distributions

- Asset Returns Example
- Posterior quantiles

Posterior S&P return at quantile 0.10

Graph showing the geometric declining average of Return against Return yesterday, with various quantiles indicated as lines on the graph.
Bayesian Modeling of Conditional Distributions

Asset Returns Example

Posterior quantiles

Posterior S&P return at quantile 0.25
Bayesian Modeling of Conditional Distributions

- Asset Returns Example
- Posterior quantiles

Posterior S&P return at quantile 0.95

Graph showing geometric declining average of return versus return yesterday.
Summary

The smoothly mixing regressions model

- is easy to apply
- is competitive with other models
- produces interesting results
- provides fully Bayesian answer to questions of interest
Current research

- Convergence of SMR
- Calibration properties of SMR
- Contrast with parsimoniously parameterized models in time series applications
- Extensions into difficult territory
  - More covariates: $x_t \ (n \times 1)$
  - Multivariate models

\[ p \left( y_t^\ell \mid \begin{array}{c} x_t \\ n \times 1 \end{array} \right) \]