Modeling conditional distributions with mixture models: Theory and Inference

John Geweke
University of Iowa, USA

Journal of Applied Econometrics Invited Lecture

Università di Venezia
Italia

June 2, 2005

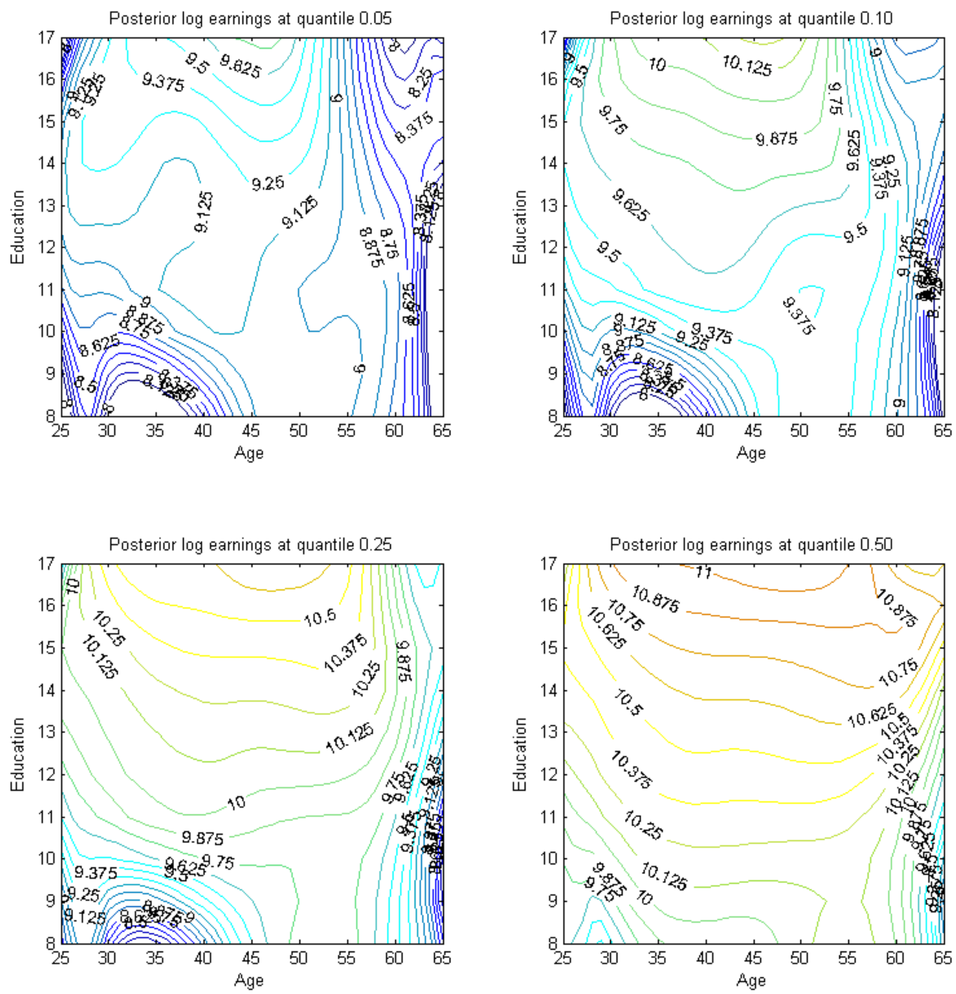
Motivation 1: Conditional distributions

$$\left\{ \underset{r \times 1}{\mathbf{w}_t}, y_t \right\} \quad i.i.d.$$

$$p(y_t \mid \mathbf{w}_t) = ?$$

Example:

w_{t1}	Experience or age of individual t
w_{t2}	Education of individual t
y_t	Earnings or wage of individual t



Quantiles of posterior distribution of log earnings conditional on age and education

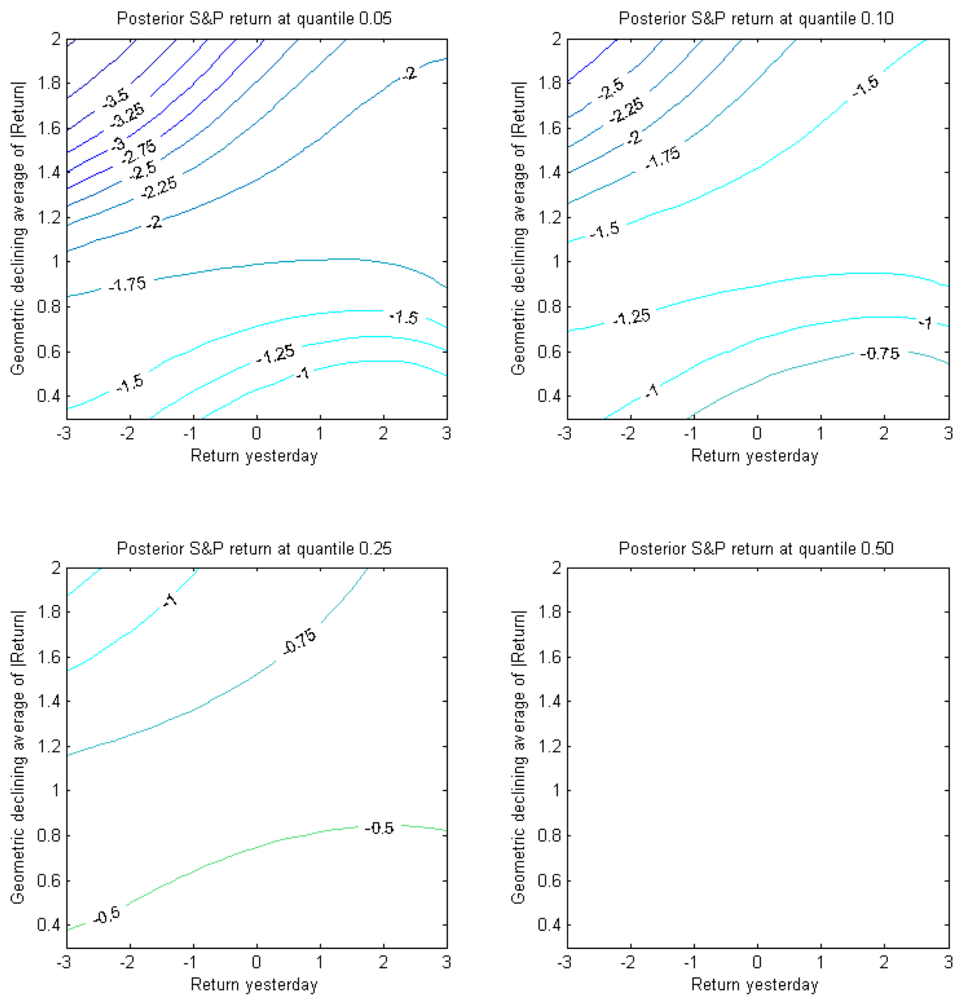
Motivation 2: Financial forecasting and decision making

y_t Return on asset in period t

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

Maximized log-likelihood values, S&P 500 daily returns, 1990-1999

Model	Maximized log-likelihood
$iid\ N\left(\mu, \sigma^2\right)$	-3285.7
GARCH(1,1)	-3028.0
EGARCH(1,1)	-2998.3
$t-$GARCH(1,1)	-2959.0
CMNMM	$>$-2967.5
SMR	$>$-2793.9



Posterior quantiles of S&P 500 predictive distribution, 1990-1999

Common structure of the models

y_t	Variable of interest
$\underset{k\times 1}{\mathbf{x}_t}\ ,\ \underset{p\times 1}{\mathbf{v}_t}$	(Possibly overlapping) sets of covariates
s_t	Latent state, $s_t \in \{1, \ldots, m\}$

$$y_t \mid (\mathbf{x}_t, \mathbf{v}_t, s_t = j) \sim N\left(\boldsymbol{\beta}'\mathbf{x}_t + \boldsymbol{\alpha}_j'\mathbf{v}_t, \sigma_j^2\right)$$

Simple normal mixture model

$$\underset{p\times 1}{\mathbf{x}_t}\ ;\ \ z_t = \mathbf{1}$$

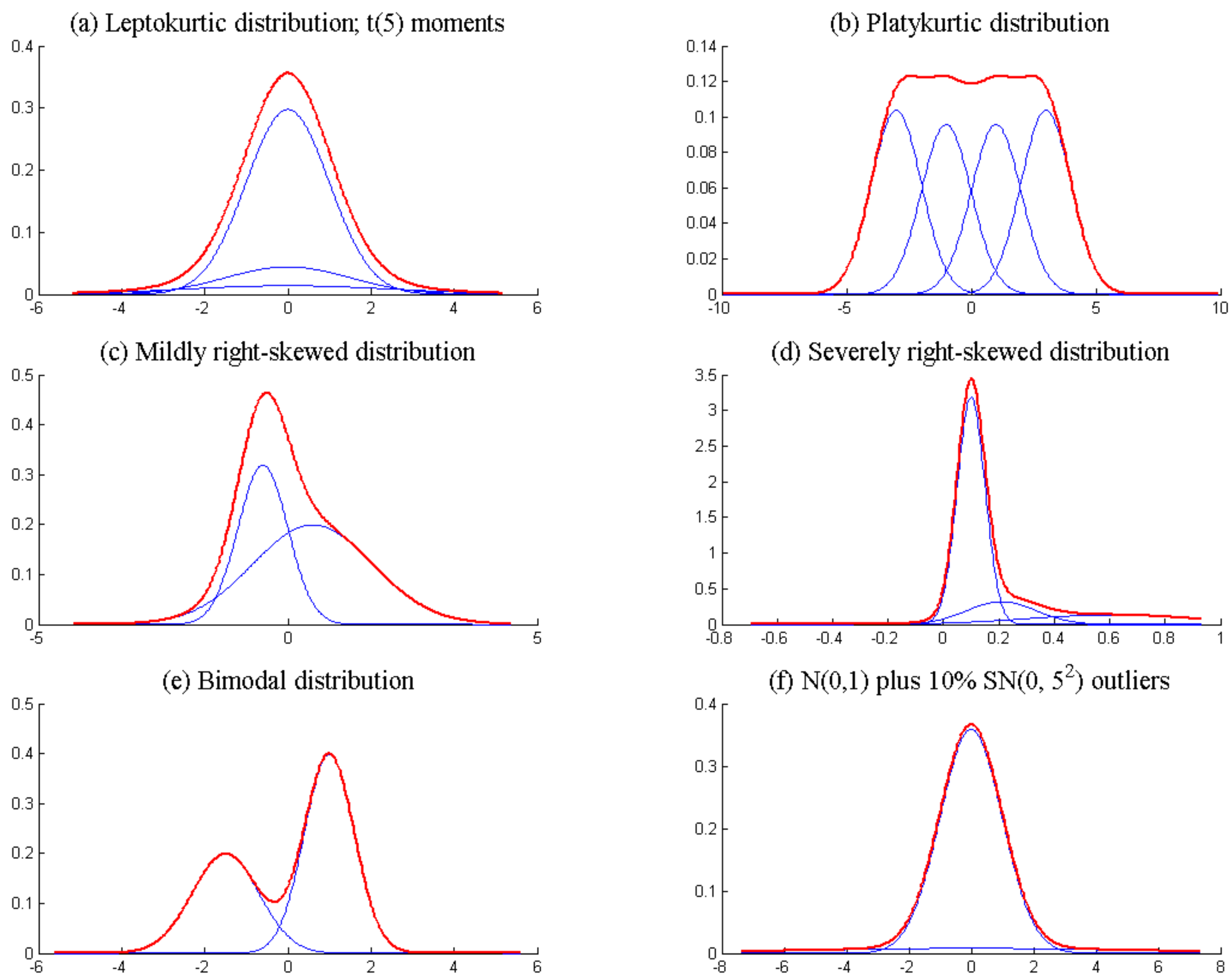
s_t independent of $\mathbf{x}_t$

s_t *i.i.d.*, $P\left(s_t = j\right) = p_j$

$$y_t \mid (\mathbf{x}_t, , s_t = j) \sim N\left(\boldsymbol{\beta}'\mathbf{x}_t + \boldsymbol{\alpha}_j v_t, \sigma_j^2\right)$$

Equivalently

$$y_t = \boldsymbol{\beta}'\mathbf{x}_t + \varepsilon_t,\ \varepsilon_t \sim N\left(\alpha_j, \sigma_j^2\right)$$


(a) Leptokurtic distribution; t(5) moments
(b) Platykurtic distribution
(c) Mildly right-skewed distribution
(d) Severely right-skewed distribution
(e) Bimodal distribution
(f) N(0,1) plus 10% SN(0, 5^2) outliers

Markov normal mixture model

$$\underset{k\times 1}{\mathbf{x}_t}\,;\;\; v_t = 1$$

s_t independent of $\mathbf{x}_t$

$$y_t \mid (\mathbf{x}_t, s_t = j) \sim N\left(\boldsymbol{\beta}'\mathbf{x}_t + \boldsymbol{\alpha}_j, \sigma_j^2\right)$$

$$P\left(s_t = j \mid s_{t-1} = i,\, s_{t-2}, s_{t-3}, \ldots\right) = p_{ij}$$

Parameters

$$\boldsymbol{\beta},\;\; \boldsymbol{\alpha} = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_m \end{pmatrix},\;\; \mathbf{P} = \begin{bmatrix} p_{11} & \cdots & p_{1m} \\ \vdots & & \vdots \\ p_{m1} & \cdots & p_{mm} \end{bmatrix},\;\; \boldsymbol{\sigma} = \begin{pmatrix} \sigma_1^2 \\ \vdots \\ \sigma_m^2 \end{pmatrix};\;\; \boldsymbol{\theta} = \{\boldsymbol{\beta}, \mathbf{P}, \boldsymbol{\alpha}, \boldsymbol{\sigma}\}$$

Inference and forecasting in the Markov normal mixture model

1. A Markov chain Monte Carlo algorithm provides draws $\left\{\boldsymbol{\theta}^{(m)}\right\}$ from
$$p\left(\boldsymbol{\theta}\mid y_1,\ldots,y_T\right).$$

2. Filtered probabilities $P\left(s_T = i \mid \boldsymbol{\theta}, y_1,\ldots,y_T\right)$ and draws from
$$\left(s_1,\ldots s_T\right) \mid \left(\boldsymbol{\theta}, y_1,\ldots,y_T\right)$$
are given by the algorithm of Chib (1995).

3. Then
$$p\left(y_{T+1} \mid \boldsymbol{\theta}, y_1,\ldots,y_T, s_1,\ldots,s_T\right) = p\left(y_{T+1} \mid \boldsymbol{\theta}, s_T = i\right)$$
is a simple mixture of normals disitribution with state probabilities $p_{i1},\ldots,p_{im}$.

Compound Markov normal mixture model

Objectives:

1. Construct a Markov mixture model with flexible components

2. Permit skewed predictive distributions while enforcing absence of serial correlation

3. Parsimonious modeling of Markov mixtures of many components

Serial correlation and skewed distributions

$$y_t \mid (\mathbf{x}_t, , s_t = j) \sim N\left(\boldsymbol{\beta}'\mathbf{x}_t + \boldsymbol{\alpha}_j, \sigma_j^2\right)$$

$$\begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} = \begin{pmatrix} 0 \\ -1 \\ 0.5 \end{pmatrix}$$

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} & 2p_{12} \\ p_{21} & p_{22} & 2p_{22} \\ p_{31} & \frac{1}{2}p_{33} & p_{33} \end{bmatrix}$$

Parameterization of the compound Markov normal mixture model

$$\mathbf{s}_t = \begin{pmatrix} s_{t1} \\ s_{t2} \end{pmatrix}; \;\; s_{t1} \in \{1, \ldots, m_1\}, \; s_{t2} \in \{1, \ldots, m_2\}$$

$$P\left(s_{t1} = j \mid s_{t-1,1} = i, \mathbf{s}_{t-2}, \mathbf{s}_{t-3}, \ldots\right) = p_{ij}$$

$$P\left(s_{t2} = j \mid s_{t1} = i, \mathbf{s}_{t-1}, \mathbf{s}_{t-2}, \ldots\right) = r_{ij}$$

$$y_t \mid (\mathbf{x}_t, s_{t1} = i, s_{t2} = j) \sim N\left(\boldsymbol{\beta}'\mathbf{x}_t + \phi_i + \psi_{ij}, \sigma^2 \cdot \sigma_i^2 \cdot \sigma_{ij}^2\right)$$

$$\boldsymbol{\beta}, \quad \boldsymbol{\phi} = \begin{pmatrix} \phi_1 \\ \vdots \\ \phi_{m_1} \end{pmatrix}, \quad \boldsymbol{\Psi} = \begin{bmatrix} \psi_{11} & \cdots & \psi_{1m_2} \\ \vdots & & \vdots \\ \psi_{m_1 1} & \cdots & \psi_{m_1 m_2} \end{bmatrix},$$

$$\mathbf{P} = \begin{bmatrix} p_{11} & \cdots & p_{1m_1} \\ \vdots & & \vdots \\ p_{m_1 1} & \cdots & p_{m_1 m_1} \end{bmatrix}, \quad \mathbf{R} = \begin{bmatrix} r_{11} & \cdots & r_{1m_2} \\ \vdots & & \vdots \\ r_{m_2 1} & \cdots & r_{m_2 m_2} \end{bmatrix},$$

$$\sigma^2, \quad \boldsymbol{\sigma} = \begin{pmatrix} \sigma_1^2 \\ \vdots \\ \sigma_{m_1}^2 \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11}^2 & \cdots & \sigma_{1m_2}^2 \\ \vdots & & \vdots \\ \sigma_{m_1 1}^2 & \cdots & \sigma_{m_1 m_2}^2 \end{bmatrix};$$

$$\boldsymbol{\theta} = \left\{ \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\Psi}, \mathbf{P}, \mathbf{R}, \boldsymbol{\sigma}^2, \boldsymbol{\sigma}, \boldsymbol{\Sigma} \right\}$$

Example for the case $m_1 = 3$, $m_2 = 2$

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{bmatrix}, \quad \mathbf{R} = \begin{bmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \\ r_{31} & r_{32} \end{bmatrix}$$

is equivalent to a 6-state Markov normal mixture model with the transition matrix

$$\begin{bmatrix} p_{11}r_{11} & p_{11}r_{12} & p_{12}r_{21} & p_{12}r_{22} & p_{13}r_{31} & p_{13}r_{32} \\ p_{11}r_{11} & p_{11}r_{12} & p_{12}r_{21} & p_{12}r_{22} & p_{13}r_{31} & p_{13}r_{32} \\ p_{21}r_{11} & p_{21}r_{12} & p_{22}r_{21} & p_{22}r_{22} & p_{23}r_{31} & p_{23}r_{32} \\ p_{21}r_{11} & p_{21}r_{12} & p_{22}r_{21} & p_{22}r_{22} & p_{23}r_{31} & p_{23}r_{32} \\ p_{31}r_{11} & p_{31}r_{12} & p_{32}r_{21} & p_{32}r_{22} & p_{33}r_{31} & p_{33}r_{32} \\ p_{31}r_{11} & p_{31}r_{12} & p_{32}r_{21} & p_{32}r_{22} & p_{33}r_{31} & p_{33}r_{32} \end{bmatrix}$$

Restrictions on parameters

Let

$$\boldsymbol{\pi} : \boldsymbol{\pi}'\mathbf{P} = \boldsymbol{\pi}'$$

$$E\left(y_t \mid \mathbf{x}_t, s_{t1} = i\right) = \boldsymbol{\beta}'\mathbf{x}_t + \phi_i + \sum_{j=1}^{m_2} r_{ij}\psi_{ij} = \boldsymbol{\beta}'\mathbf{x}_t + \phi_i$$

$$E\left(y_t\right) = \boldsymbol{\beta}'\mathbf{x}_t + \sum_{i=1}^{m_1} \pi_i\phi_i + \sum_{i=1}^{m_1} \pi_i \sum_{j=1}^{m_2} r_{ij}\psi_{ij} = \boldsymbol{\beta}'\mathbf{x}_t$$

Number of "identified" parameters in the model is $m_1\left(m_1 + 3m_2 - 3\right) + k - 1$.

Absence of serial correlation

$$y_t \mid (\mathbf{x}_t, s_{t1} = i, s_{t2} = j) \sim N\left(\boldsymbol{\beta}'\mathbf{x}_t + \phi_i + \psi_{ij}, \sigma^2 \cdot \sigma_i^2 \cdot \sigma_{ij}^2\right)$$

$$P\left(s_{t1} = j \mid s_{t-1,1} = i, \mathbf{s}_{t-2}, \mathbf{s}_{t-3}, \ldots\right) = p_{ij}$$

$$P\left(s_{t2} = j \mid s_{t1} = i, \mathbf{s}_{t-1}, \mathbf{s}_{t-2}, \ldots\right) = r_{ij}$$

Conditional on $\mathbf{x}_t$ $(t = 1, \ldots, T)$, the observables y_t $(t = 1, \ldots, T)$ are serially uncorrelated if $\boldsymbol{\phi} = \mathbf{0}$.

Suppose further that $\mathbf{P}$ is irreducible and aperiodic and its eigenvalues are distinct.

Then the observables $\mathbf{y}_t$ are serially uncorrelated if and only if $\boldsymbol{\phi} = \mathbf{0}$.

Conditionally conjugate prior distributions

$$y_t \mid (\mathbf{x}_t, s_{t1} = i, s_{t2} = j) \sim N\left(\boldsymbol{\beta}'\mathbf{x}_t + \phi_i + \psi_{ij}, \sigma^2 \cdot \sigma_i^2 \cdot \sigma_{ij}^2\right)$$

$$P\left(s_{t1} = j \mid s_{t-1,1} = i, \mathbf{s}_{t-2}, \mathbf{s}_{t-3}, \ldots\right) = p_{ij}$$

$$P\left(s_{t2} = j \mid s_{t1} = i, \mathbf{s}_{t-1}, \mathbf{s}_{t-2}, \ldots\right) = r_{ij}$$

Gaussian:	$\boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\Psi}$
Gaussian conditional on σ^2:	$\boldsymbol{\Phi}$
Inverse Gamma:	$\sigma^2, \boldsymbol{\sigma}, \boldsymbol{\Sigma}$
Gaussian conditional on σ^2, $\boldsymbol{\sigma}$:	$\boldsymbol{\Phi}$
Dirichlet:	Each row of $\mathbf{P}$, each row of $\mathbf{R}$

Smoothly mixing regression models (SMRs)

Begin with the same normal mixture model

$$y_t \mid (\mathbf{x}_t, \mathbf{v}_t, s_t = j) \sim N\left(\underset{k\times 1}{\boldsymbol{\beta}}{}'\mathbf{x}_t + \underset{p\times 1}{\boldsymbol{\alpha}'_j}\mathbf{v}_t, \sigma_j^2\right)$$

Determination of latent states s_t:

$$\underset{m\times 1}{\widetilde{\mathbf{w}}_t} = \boldsymbol{\Gamma}\underset{q\times 1}{\mathbf{z}_t} + \boldsymbol{\zeta}_t; \;\; \boldsymbol{\zeta}_t \overset{iid}{\sim} N(\mathbf{0}, \mathbf{I}_m)$$

$$\widetilde{s}_t = j \;\text{ iff }\; \widetilde{w}_{tj} \geq \widetilde{w}_{ti} \;\forall\; i = 1, \ldots, m$$

$$y_t \sim N\left(\underset{k\times 1}{\boldsymbol{\beta}}{}'\mathbf{x}_t + \underset{p\times 1}{\boldsymbol{\alpha}_j'}\mathbf{v}_t, \sigma_j^2\right), \quad \underset{m\times 1}{\widetilde{\mathbf{w}}_t} = \boldsymbol{\Gamma}\underset{q\times 1}{\mathbf{z}_t} + \boldsymbol{\zeta}_t, \quad \widetilde{s}_t = j \text{ iff } \widetilde{w}_{tj} \geq \widetilde{w}_{ti}\ \forall\ i$$

When $q = 1$, then $z_t = 1$ and probabilities of mixture components are fixed.

(A) If $k > 1$ and $p = 1$:

Simple normal mixture model of disturbances in linear regression

(B) If $k = 1$ and $p > 1$:

Mixture of linear regressions with fixed component probabilities

($k > 1$, $p > 1$: Facilitates hierarchical prior)

When $q > 1$, then probabilities of mixture components depend on $\mathbf{z}_t$

(C) $k = 1$ and $p = 1$:

Mixture of fixed normals, with $\mathbf{w}_t$-dependent probabilities

(D) $k > 1$ and $p = 1$:

Normal mixture model of disturbances in linear regression,

but with covariate-dependent state probabilities

(E) $k = 1$ and $p > 1$:

Mixture of linear regressions with $\mathbf{w}_t$-dependent probabilities

Parameterization issues

$$y_t \sim N\left(\underset{k\times 1}{\boldsymbol{\beta}}{}'\mathbf{x}_t + \underset{p\times 1}{\boldsymbol{\alpha}'_j}\mathbf{v}_t, \sigma^2 \cdot \sigma_j^2\right), \quad \underset{m\times 1}{\widetilde{\mathbf{w}}_t} = \boldsymbol{\Gamma}\underset{q\times 1}{\mathbf{z}_t} + \boldsymbol{\zeta}_t, \quad \widetilde{s}_t = j \text{ iff } \widetilde{w}_{tj} \geq \widetilde{w}_{ti} \; \forall \; i$$

Because $\boldsymbol{\zeta}_t \overset{iid}{\sim} N(\mathbf{0}, \mathbf{I}_m)$ translation but not scaling isues arise in

$$\widetilde{\mathbf{w}}_t = \boldsymbol{\Gamma}\mathbf{z}_t + \boldsymbol{\zeta}_t.$$

Impose $\boldsymbol{\iota}'_m\boldsymbol{\Gamma} = \mathbf{0}$ through

$$\boldsymbol{\Gamma} = \mathbf{P}\cdot\begin{bmatrix} \mathbf{0}'_p \\ \boldsymbol{\Gamma}^* \end{bmatrix}, \text{ where } \mathbf{P} = \underset{m\times m}{\mathbf{P}} := \begin{bmatrix} \boldsymbol{\iota}_m \cdot m^{-1/2} & \mathbf{P}_2 \end{bmatrix}, \; \mathbf{P}'\mathbf{P} = \mathbf{I}_m.$$

$$\boldsymbol{\alpha}' = \left(\boldsymbol{\alpha}'_1, \ldots, \boldsymbol{\alpha}'_m\right), \; \boldsymbol{\sigma}' = \left(\sigma_1^2, \ldots, \sigma_m^2\right)$$

Conditionally conjugate prior distributions

$$y_t \sim N\left(\underset{k\times 1}{\boldsymbol{\beta}}{}'\mathbf{x}_t + \underset{p\times 1}{\boldsymbol{\alpha}'_j}\mathbf{v}_t, \sigma^2 \cdot \sigma_j^2\right), \quad \underset{m\times 1}{\widetilde{\mathbf{w}}_t} = \boldsymbol{\Gamma}\underset{q\times 1}{\mathbf{z}_t} + \boldsymbol{\zeta}_t, \quad \widetilde{s}_t = \text{iff } \widetilde{w}_{tj} \geq \widetilde{w}_{ti} \; \forall \; i$$

Gaussian: $\boldsymbol{\beta}, \boldsymbol{\Gamma}^*$

Gaussian conditional on σ^2: $\boldsymbol{\alpha}$

Inverse gamma: $\sigma^2, \boldsymbol{\sigma}$

Blocking for Gibbs sampling

$$y_t \sim N\left(\underset{k\times 1}{\boldsymbol{\beta}}{}'\mathbf{x}_t + \underset{p\times 1}{\boldsymbol{\alpha}'_j}\mathbf{v}_t, \sigma^2 \cdot \sigma_j^2\right), \quad \underset{m\times 1}{\widetilde{\mathbf{w}}_t} = \boldsymbol{\Gamma}\underset{q\times 1}{\mathbf{z}_t} + \boldsymbol{\zeta}_t, \quad \widetilde{s}_t = j \text{ iff } \widetilde{w}_{tj} \geq \widetilde{w}_{ti}\ \forall\ i$$

$\sigma^2, \sigma_1^2, \ldots, \sigma_m^2$	Separately conditionally independent inverse gamma
$\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$	Jointly conditionally Gaussian
$vec\left(\boldsymbol{\Gamma}^*\right)$	Conditionally Gaussian
$\widetilde{\mathbf{w}}_t$	Orthant-truncated Gaussian times orthant-specific likelihood factors

Covariates in applications to date

Substantively distinct covariates: a_t, b_t

First example:

$$a_t = \text{Age of individual } t, \;\; b_t = \text{Education of individual } t$$

Second example:

$$\begin{aligned} a_t &= \text{Return on asset in period } t-1,\, a_t = y_{t-1} \\ b_t &= g \cdot b_{t-1} + (1-g)\left|a_{t-1}\right|^{\kappa} = \sum_{s=0}^{\infty} g^s \left|y_{t-2-s}\right|^{\kappa} \end{aligned}$$

Covariates of the form:

$$x_{tj} = a_t^{\ell_1} b_t^{\ell_2},\, \ell_1 \in \{0,\ldots,L_1\},\, \ell_2 \in \{0,\ldots,L_2\}$$

Functions of interest

CDFs: $P\left(y_t \le c \mid a_t, b_t\right) = P\left(y_t \le c \mid \mathbf{x}_t, \mathbf{v}_t, \mathbf{z}_t\right)$

Quantiles: $c\left(q\right) = \left\{c : P\left(y_t \le c \mid a_t, b_t\right) = P\left(y_t \le c \mid \mathbf{x}_t, \mathbf{v}_t, \mathbf{z}_t\right) = q\right\}$

Note $P\left(\widetilde{s}_t = j \mid \boldsymbol{\Gamma}, \mathbf{z}_t\right)$

$$
\begin{aligned}
&= P\left[\widetilde{w}_{tj} \ge \widetilde{w}_{ti} \;\; (i = \mathbf{1}, \ldots, m) \mid \boldsymbol{\Gamma}, \mathbf{z}_t\right] \\
&= \int_{-\infty}^{\infty} p\left(\widetilde{w}_{tj} = y \mid \boldsymbol{\Gamma}, \mathbf{z}_t\right) \cdot P\left[\widetilde{w}_{ti} \le y \;\; (i = \mathbf{1}, \ldots, m) \mid \boldsymbol{\Gamma}, \mathbf{z}_t\right] dy \\
&= \int_{-\infty}^{\infty} \phi\left(y - \boldsymbol{\Gamma}\mathbf{z}_t\right) \prod_{i \ne j} \boldsymbol{\Phi}\left(y - \boldsymbol{\gamma}_i'\mathbf{z}_t\right) dy.
\end{aligned}
$$

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.

Detail of prior distributons for $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$, $\boldsymbol{\Gamma}^*$

$$a_t \in [a_1^*, a_2^*] = A, \;\; b_t \in [b_1^*, b_2^*] = B$$

Grid of points

$$\begin{aligned} G \;\; = \;\; & \left\{ \left(a_i, b_j \right) : a_i = a_1^* + i \cdot \boldsymbol{\Delta}_a, b_j = b_1^* + j \cdot \boldsymbol{\Delta}_b \right\}, \\ & (i = 0, \ldots, N_a;\; j = 0, \ldots, N_b; \\ & \boldsymbol{\Delta}_a = (a_2^* - a_1^*) \,/\, (N_a + 1)\,,\; \boldsymbol{\Delta}_b = (b_2^* - b_1^*) \,/\, (N_b + 1)) \end{aligned}$$

Let $\mathbf{x}_{ij}^*$ be the vector corresponding to $\left(a_i, b_j\right)$. Then the prior has $(N_a + 1)(N_b + 1)$ independent components,

$$\boldsymbol{\beta}' \mathbf{x}_{ij}^* \sim N\left[\mu, \tau^2 (N_a + 1)(N_b + 1)\right].$$