Bayesian Modeling of Conditional Distributions

John Geweke
University of Iowa, USA

Bayesian Econometric Methodology Workshop
Sveriges Riksbank, Stockholm

September 8, 2006

Extension of work described in

Smoothly Mixing Regressions

John Geweke - University of Iowa, USA

Michael Keane - Sydney University of Technology

Forthcoming, Journal of Econometrics

Outline

- Motivation
- Model description (briefly)
- Methods of inference (briefly)
- An application to asset returns
- Model evaluation

Motivation – Conditional distributions

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

$$p\left(y_t \mid \mathbf{x}_t\right) = ?$$

Example:

x_{1t}	Experience or age of individual t
x_{2t}	Education of individual t
y_t	Earnings or wage of individual t

Motivation – Financial forecasting and decision making

y_t Return on asset in period t

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

x_{1t} y_{t-1}

x_{2t} $f\left(|y_{t-2}|, |y_{t-3}|, |y_{t-4}| \ldots\right)$

Motivation – Out of sample model comparisons

Model	Log Predictive likelihood	Log Recursive ML
SMR	-1602.0	
Stochastic volatility	-1625.3	
t-GARCH(1,1)	-1625.5	-1624.7
EGARCH(1,1)		-1637.5
GARCH(1,1)		-1660.5
Normal iid		-1848.5

Motivation – Relevant literature

Models:

- Mixture models (e.g. Titterington et al., McLachlan & Peel)
- Quantile regression (e.g. Koenker)
- Neural network models
  - Machine learning (e.g. Jordan & coauthors)
  - Hierarchical mixtures of experts (e.g. Tanner)

(Beyond paper)

Motivation – Relevant literature

Methods:

Bayesian

Non-Bayesian

Ad hoc

Asset return applications: Parsimoniously parameterized models

*ARCH

Stochastic volatility

Model description – Common structure of the models

y_t Variable of interest

$\underset{n\times 1}{\mathbf{x}_t}$ Vector of covariates

$\underset{k\times 1}{\mathbf{u}_t}, \underset{p\times 1}{\mathbf{v}_t}, \underset{q\times 1}{\mathbf{z}_t}$ Created from $\mathbf{x}_t$ by transformation;

$u_{t1} = v_{t1} = z_{t1}$ always

$\widetilde{s}_t$ Latent state, $\widetilde{s}_t \in \{1, \ldots, m\}$

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

Model description – Simple normal mixture model

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

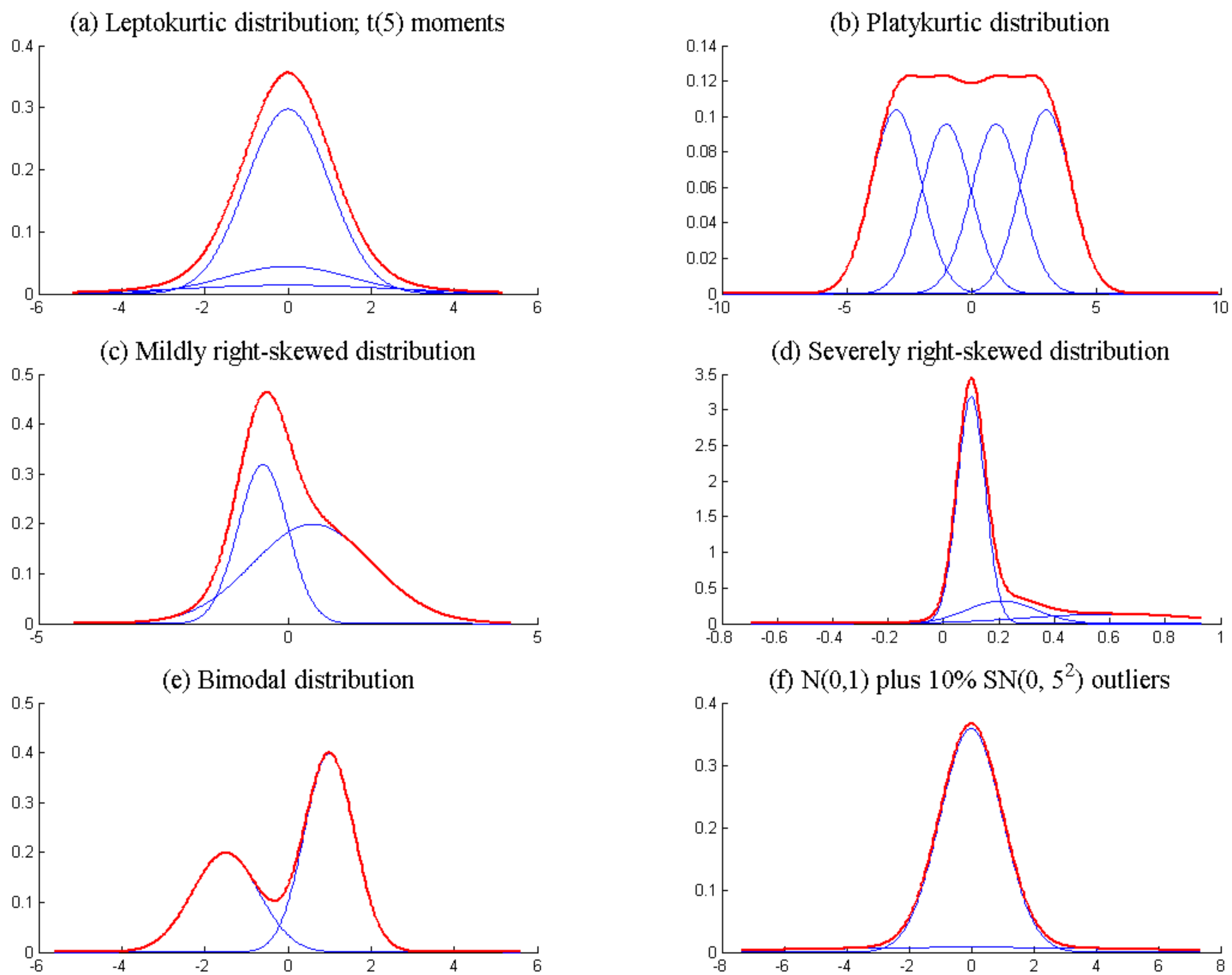
$$\widetilde{s}_t \text{ independent of } \mathbf{x}_t$$

$$\widetilde{s}_t \ i.i.d.,\ P\left(\widetilde{s}_t = j\right) = p_j$$

$$y_t \mid \left(\mathbf{u}_t, \widetilde{s}_t = j\right) \sim N\left(\boldsymbol{\beta}'\mathbf{u}_t + \alpha_j v_t, \sigma_j^2\right)$$

Equivalently, if $\widetilde{s}_t = j$ then

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


(a) Leptokurtic distribution; t(5) moments
0.4
0.3
0.2
0.1
0
-6
-4
-2
0
2
4
6
(b) Platykurtic distribution
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
-10
-5
0
5
10
(c) Mildly right-skewed distribution
0.5
0.4
0.3
0.2
0.1
0
-5
0
5
(d) Severely right-skewed distribution
3.5
3
2.5
2
1.5
1
0.5
0
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1
(e) Bimodal distribution
0.5
0.4
0.3
0.2
0.1
0
-6
-4
-2
0
2
4
6
(f) N(0,1) plus 10% SN(0, 5²) outliers
0.4
0.3
0.2
0.1
0
-8
-6
-4
-2
0
2
4
6
8

Model description – What is new

Begin with the same normal mixture model

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

Determination of latent states $\widetilde{s}_t$:

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

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

Model description – Special cases of the model

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

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

($k = p > 1$, $\mathbf{v}_t = \mathbf{u}_t$: Facilitates hierarchical prior)

Model description – More special cases of the model

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

When $q > 1$, state probabilities are $\mathbf{z}_t$-dependent

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

Mixture of fixed normals with $\mathbf{z}_t$-dependent state probabilities

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

Mixture of regression disturbances, $\mathbf{z}_t$-dependent state probabilities

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

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

Model description – Parameterization issues

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

$\boldsymbol{\zeta}_t \overset{iid}{\sim} N(\mathbf{0}, \mathbf{I}_m) \implies$ translation but not scaling issues in $\widetilde{\mathbf{w}}_t = \boldsymbol{\Gamma}\mathbf{z}_t + \boldsymbol{\zeta}_t$..

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

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

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

Model description – Covariates

Substantively distinct covariates: x_{1t}, x_{2t}

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

Covariates of the form:

$$
u_{tj} = x_{1t}^{\ell_1} x_{2t}^{\ell_2},\ \ell_1 \in \{0, \ldots, L_1\},\ \ell_2 \in \{0, \ldots, L_2\}
$$

Model description – Conditionally conjugate prior distributions

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

Distribution type	Parameters	Hyperparameters
Gaussian:	$\boldsymbol{\beta}, \boldsymbol{\Gamma}^*$	$\underline{\mu}$, $\underline{\tau}_\beta^2$, $\underline{\tau}_\gamma^2$
Gaussian conditional on σ^2:	$\boldsymbol{\alpha}$	$\underline{\tau}_\alpha^2$
Inverse gamma:	$\sigma^2, \boldsymbol{\sigma}$	$\underline{s}^2$, $\underline{\nu}$, $\underline{\nu}^*$

Model description – Detail of prior distributions for $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$, $\boldsymbol{\Gamma}^*$

$$x_{1t} \in \left[x_1^a, x_1^b\right], \;\; x_{2t} \in \left[x_2^a, x_2^b\right]$$

Choose N_1, N_2, define $\boldsymbol{\Delta}_1 = \left(x_1^b - x_1^a\right)/N_1$, $\boldsymbol{\Delta}_2 = \left(x_2^b - x_2^a\right)/N_2$ and then

$$G = \Big\{(x_{1i}, x_{2i}) : x_{1i} = x_1^a, x_1^a + \boldsymbol{\Delta}_1, \ldots, x_{x1}^a + (N_1 - 1)\,\boldsymbol{\Delta}_1, x_1^b; \\ x_{2i} = x_2^a, x_2^a + \boldsymbol{\Delta}_2, \ldots, x_{x2}^a + (N_2 - 1)\,\boldsymbol{\Delta}_2, x_2^b\Big\}$$

Let $\mathbf{c}_{ij}$ be the vector corresponding to $\left(x_{1i}, x_{2j}\right)$. Then

$$\boldsymbol{\beta}'\mathbf{c}_{ij} \overset{iid}{\sim} N\left[\underline{\mu}, \underline{\tau}_\beta^2 (N_a + 1)(N_b + 1)\right] \quad (i = 0, \ldots, N_1; j = 0, \ldots, N_2).$$

Methods of inference – Blocking for Gibbs sampling

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

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

Methods of inference – Functions of interest

CDFs: $P\left(y_t \leq c \mid x_{1t}, x_{2t}\right) = P\left(y_t \leq c \mid \mathbf{u}_t, \mathbf{v}_t, \mathbf{z}_t\right)$

Quantiles: $c\left(q\right) = \{c : P\left(y_t \leq c \mid x_{1t}, x_{2t}\right) = P\left(y_t \leq c \mid \mathbf{u}_t, \mathbf{v}_t, \mathbf{z}_t\right) = q\}$

$$P\left(\widetilde{s}_t = j \mid \mathbf{\Gamma}, \mathbf{z}_t\right) = P\left[\widetilde{w}_{tj} \geq \widetilde{w}_{ti} \ \ (i = \mathbf{1}, \ldots, m) \mid \mathbf{\Gamma}, \mathbf{z}_t\right]$$
$$= \int_{-\infty}^{\infty} p\left(\widetilde{w}_{tj} = y \mid \mathbf{\Gamma}, \mathbf{z}_t\right) \cdot P\left[\widetilde{w}_{ti} \leq y \ \ (i = \mathbf{1}, \ldots, m) \mid \mathbf{\Gamma}, \mathbf{z}_t\right] dy$$
$$= \int_{-\infty}^{\infty} \phi\left(y - \boldsymbol{\gamma}_j' \mathbf{z}_t\right) \prod_{i \neq j} \mathbf{\Phi}\left(y - \boldsymbol{\gamma}_i' \mathbf{z}_t\right) dy.$$

The posterior distribution is a mixture of normals with $M \cdot m$ components.

An application to asset returns – Data

Variable of interest:

$$y_t = 100 \log \left(p_t / p_{t-1} \right): \quad \text{Daily S\&P 500 returns, 1990 - 1999}$$

The covariate vectors $\mathbf{u}_t$, $\mathbf{v}_t$ and $\mathbf{z}_t$ are interactive polynomials in the variables

$$\begin{aligned} x_{1t} &= \text{Return in period } t-1,\ x_{2t} = y_{t-1} \\ x_{2t} &= (1-g) \sum_{s=0}^{\infty} g^s \left| y_{t-2-s} \right|^{\kappa} \end{aligned}$$

An application to asset returns – Model

$$x_{1t} = y_{t-1}, \;\; x_{2t} = .05 \sum_{s=0}^{\infty} .95^s \left| y_{t-2-s} \right|$$

$$\begin{aligned} \widetilde{w}_{tj} &= \gamma_{1j} + \gamma_{2j} x_{1t} + \gamma_3 x_{2t} \\ &\quad + \gamma_4 x_{4j} x_{1t}^2 + \gamma_5 x_{1t} x_{2t} + \gamma_6 x_{2t}^2 + \zeta_{tj} \;\; (j = 1,2,3) \end{aligned}$$

$$P\left(\widetilde{s}_t = j \mid x_{1t}, x_{2t}\right) = P\left[\widetilde{w}_{tj} \geq \widetilde{w}_{ti} \;\; (i = 1,2,3) \mid \mathbf{x}_{1t}, x_{2t}\right]$$

$$y_t \mid \left(\widetilde{s}_t = j, x_{1t}, x_{2t}\right) \sim N\left(\mu_j, \sigma^2 \cdot \sigma_j^2\right) \quad (j = 1,2,3)$$

An application to asset returns – Priors

Gaussian priors

$$\begin{array}{ll} \boldsymbol{\beta}: & \mu = 0,\ \tau^2 = 1 \\ \boldsymbol{\alpha}: & \mu = 0,\ \tau^2 = 9 \\ \boldsymbol{\Gamma}^* & \mu = 0,\ \tau^2 = 16 \end{array}$$

Grid G: -10 to 10 for $x_{1t} = y_{t-1}$; 0 to 10 for $x_{2t} = (1-g)\sum_{s=0}^{\infty} g^s \left|y_{t-2-s}\right|^{\kappa}$

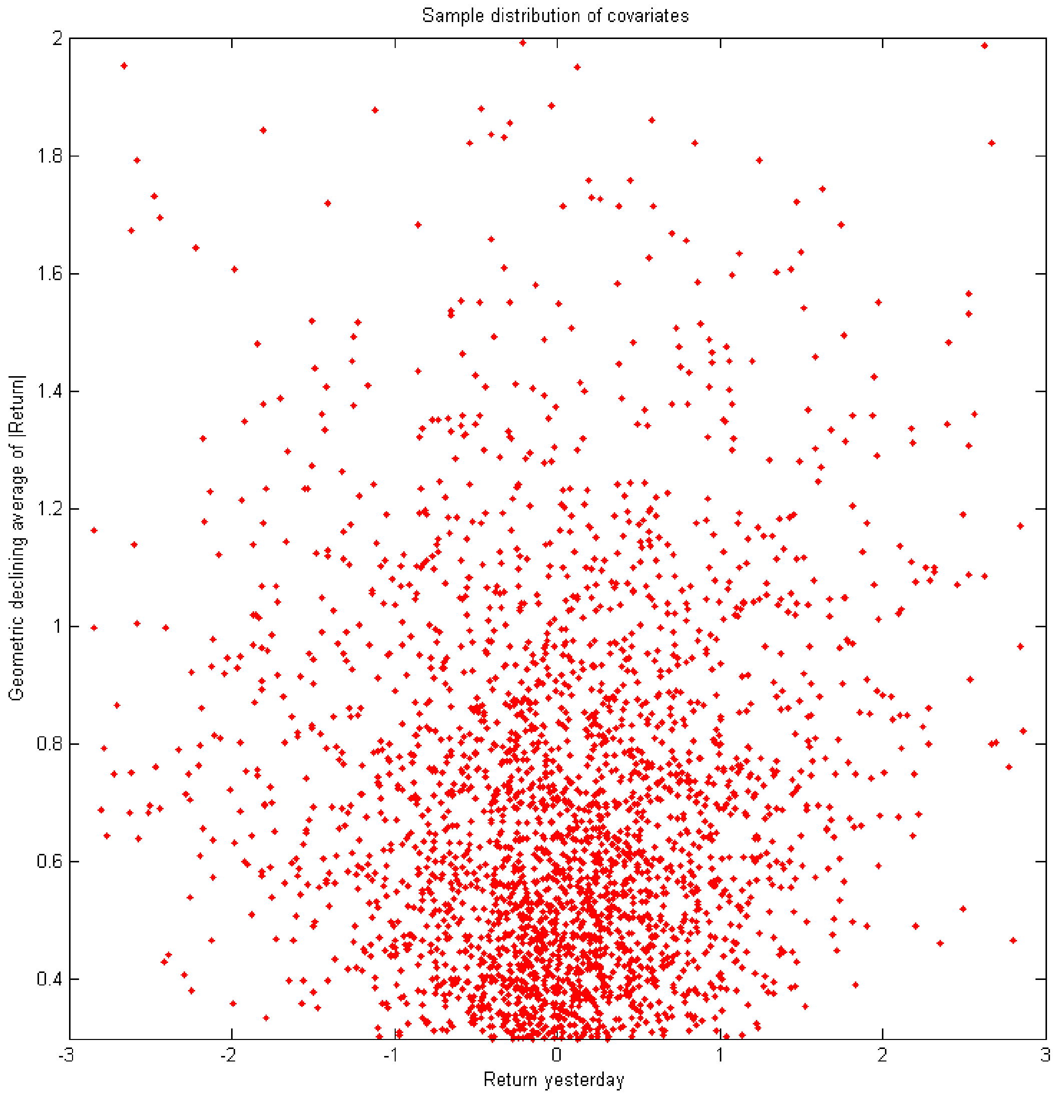
Inverse gamma priors:

$$\begin{aligned} 2/\sigma^2 &\sim \chi^2(2) \\ 2/\sigma_j^2 &\sim \chi^2(2) \end{aligned}$$

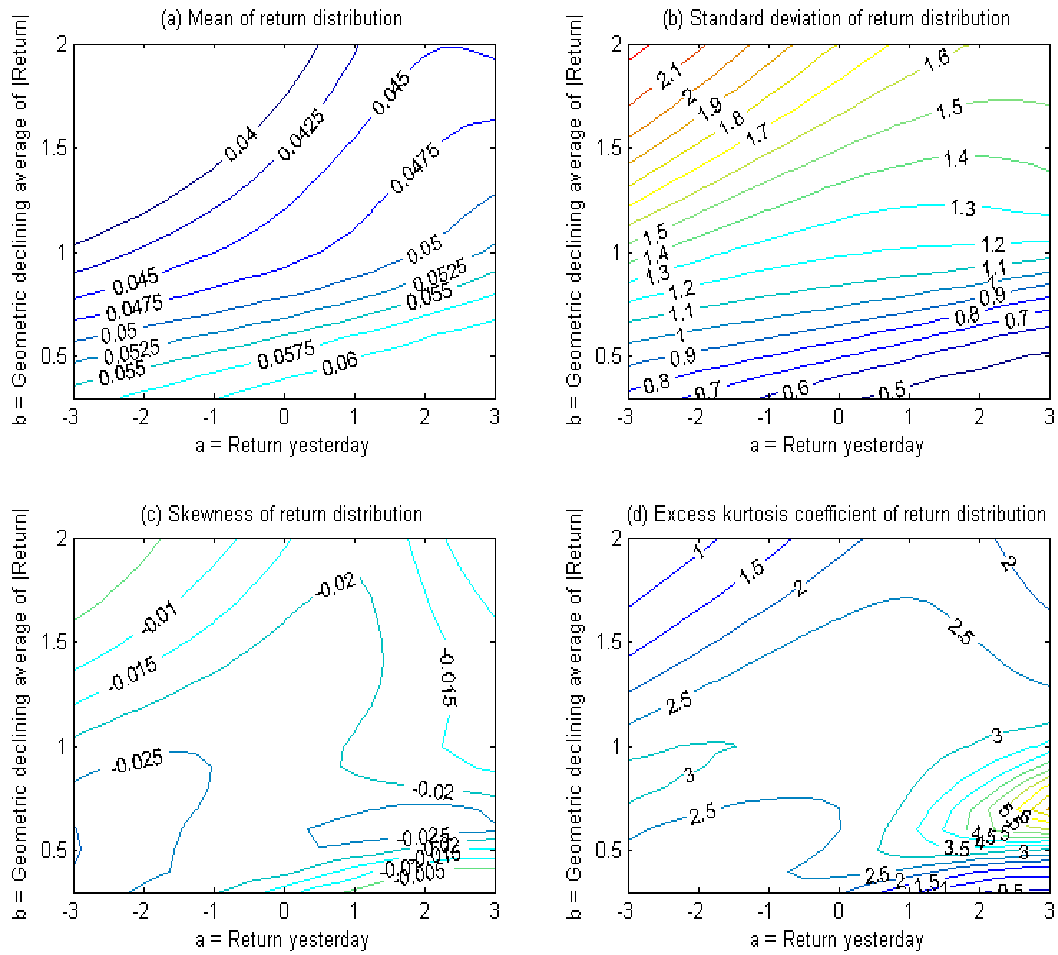
An application to asset returns – Data and a few technical details

All results for S&P 500 daily returns, Jan. 2, 1990 - Dec. 31, 1999

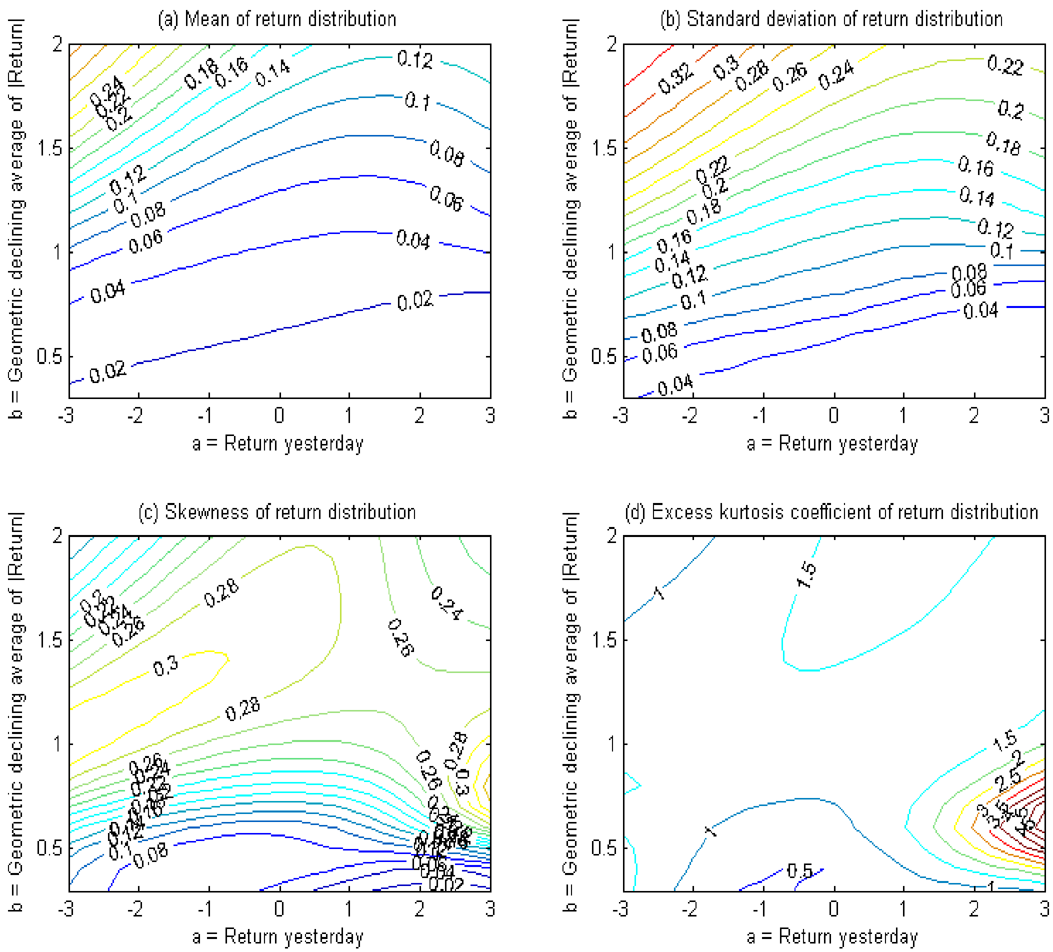
1. 12,000 MCMC iterations
2. Every 100'th iteration recorded
3. Of the 120 iterations recorded
   (a) First 20 discarded
   (b) Remaining 100 used for analysis



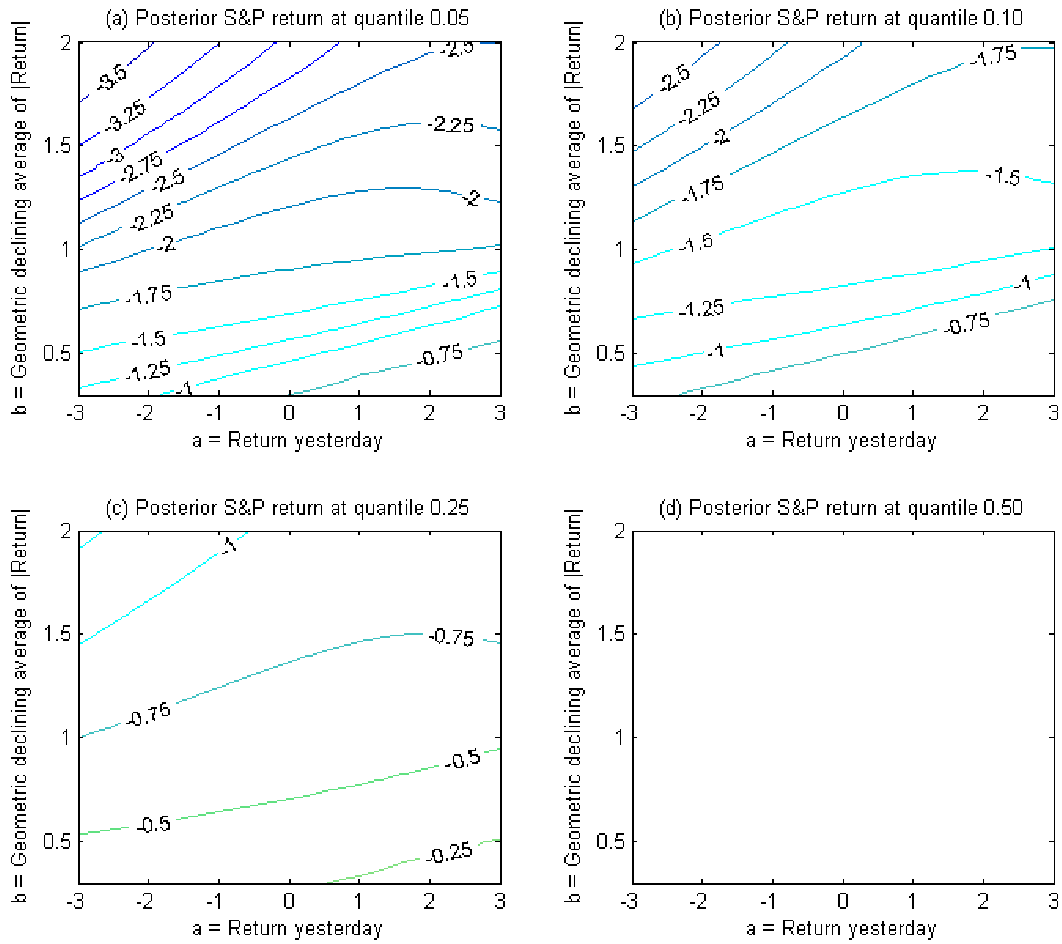
Sample distribution of x_{1t} and x_{2t}, S&P 500 returns illustration



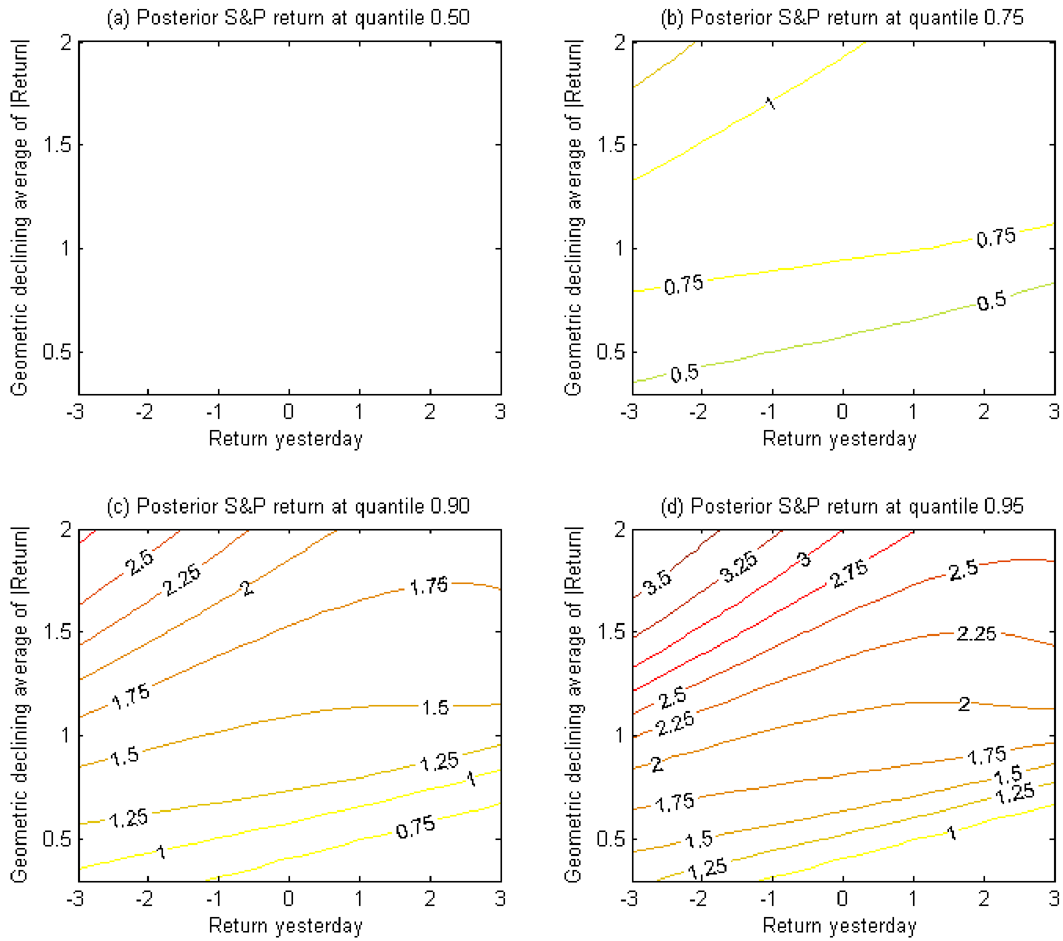
Posterior means of four population conditional moments



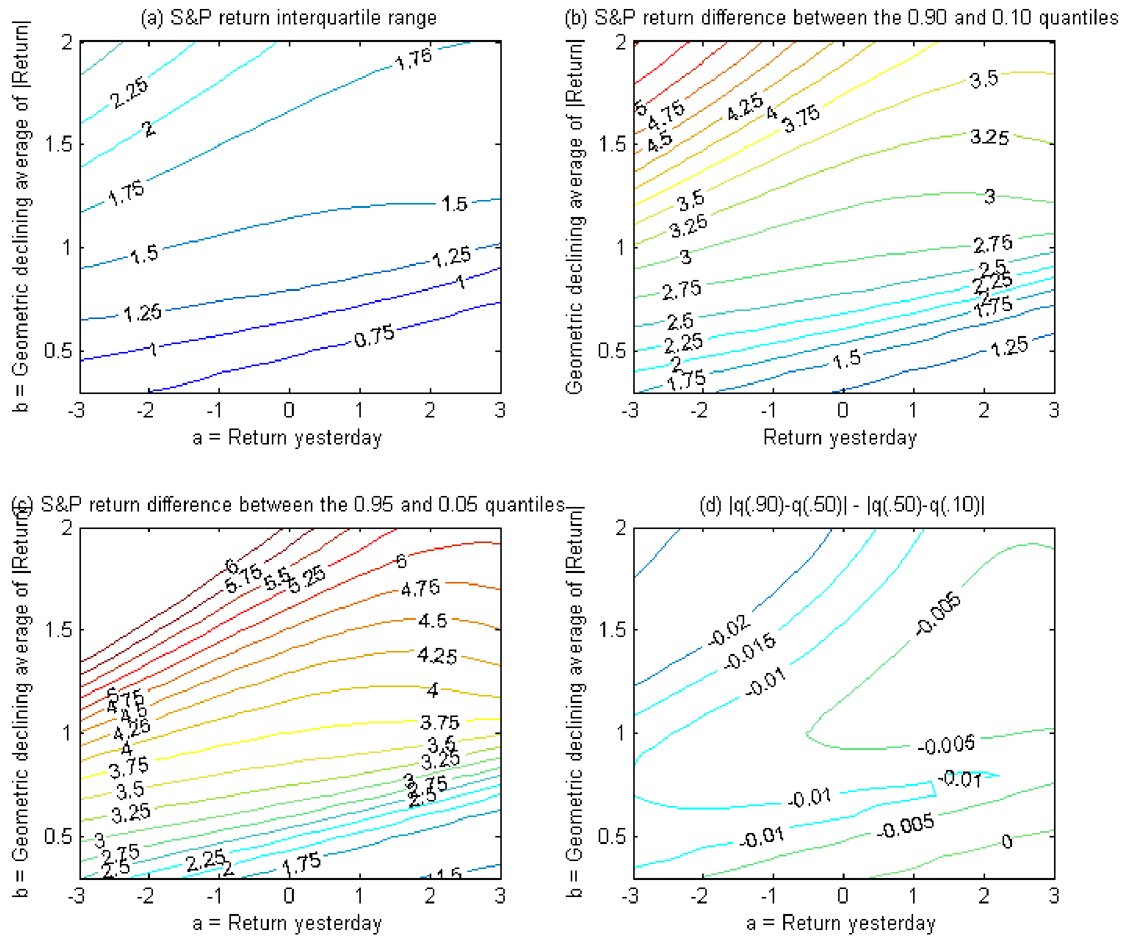
Posterior standard deviations of four population conditional moments



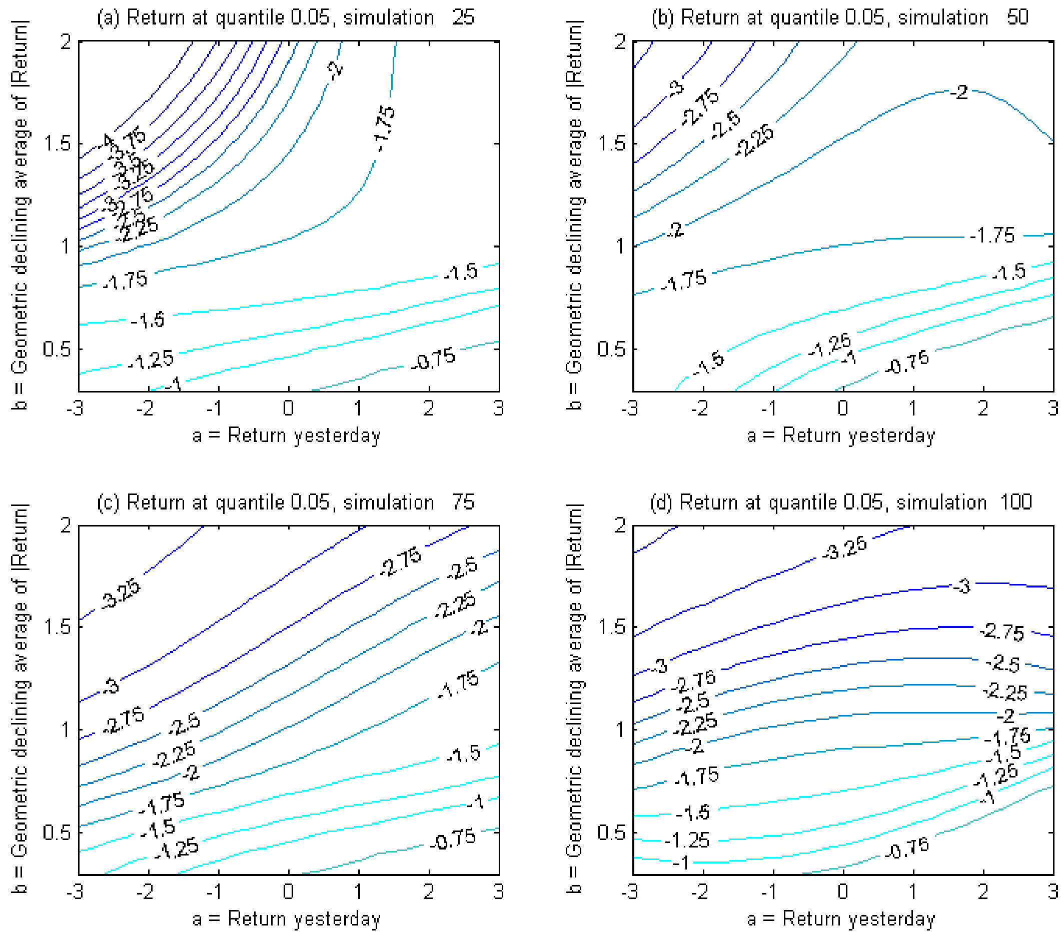
Quantiles of the posterior conditional distribution of returns



Quantiles of the posterior conditional distribution



Aspects of the dispersion of the posterior conditional distribution



5% quantile of the conditional distribution of returns

Model evaluation – Analytical Model Comparison

$$\frac{P(A \mid \mathbf{y}^o)}{P(B \mid \mathbf{y}^o)} = \frac{P(A)}{P(B)} \cdot \frac{p(\mathbf{y}^o \mid A)}{p(\mathbf{y}^o \mid B)}$$

$$p(\mathbf{y}^o \mid A) = \prod_{t=1}^{T} p\left(\mathbf{y}_t^o \mid \mathbf{Y}_{t-1}^o, A\right) \iff \log p(\mathbf{y}^o \mid A) = \sum_{t=1}^{T} \log p\left(\mathbf{y}_t^o \mid \mathbf{Y}_{t-1}^o, A\right)$$

$$\begin{aligned} p\left(\mathbf{y}_t^o \mid \mathbf{Y}_{t-1}^o, A\right) &= \int_{\Theta_A} p\left(\mathbf{y}_t^o \mid \boldsymbol{\theta}_A, \mathbf{Y}_{t-1}^o, A\right) p\left(\boldsymbol{\theta}_A \mid \mathbf{Y}_{t-1}^o, A\right) d\boldsymbol{\theta}_A \\ &\approxeq M^{-1} \sum_{m=1}^{M} p\left(\mathbf{y}_t^o \mid \boldsymbol{\theta}_A^{(m)}, \mathbf{Y}_{t-1}^o, A\right). \end{aligned}$$

Model evaluation – Log-predictive likelihood

$$\log p\left(\mathbf{y}^o_{T_1+1},\dots\mathbf{y}^o_T \mid \mathbf{Y}^o_{T_1}, A\right) = \sum_{t=1}^{T} \log p\left(\mathbf{y}^o_t \mid \mathbf{Y}^o_{t-1}, A\right)$$

Sample $(1,\dots,T_1)$: 1990-1994   Prediction $(T_1+1,\dots,T)$: 1995-1999

Model	Log Predictive likelihood	Log Recursive ML
SMR	-1602.0	
Stochastic volatility	-1625.3	
t-GARCH(1,1)	-1625.5	-1624.7
EGARCH(1,1)		-1637.5
GARCH(1,1)		-1660.5
Normal iid		-1848.5

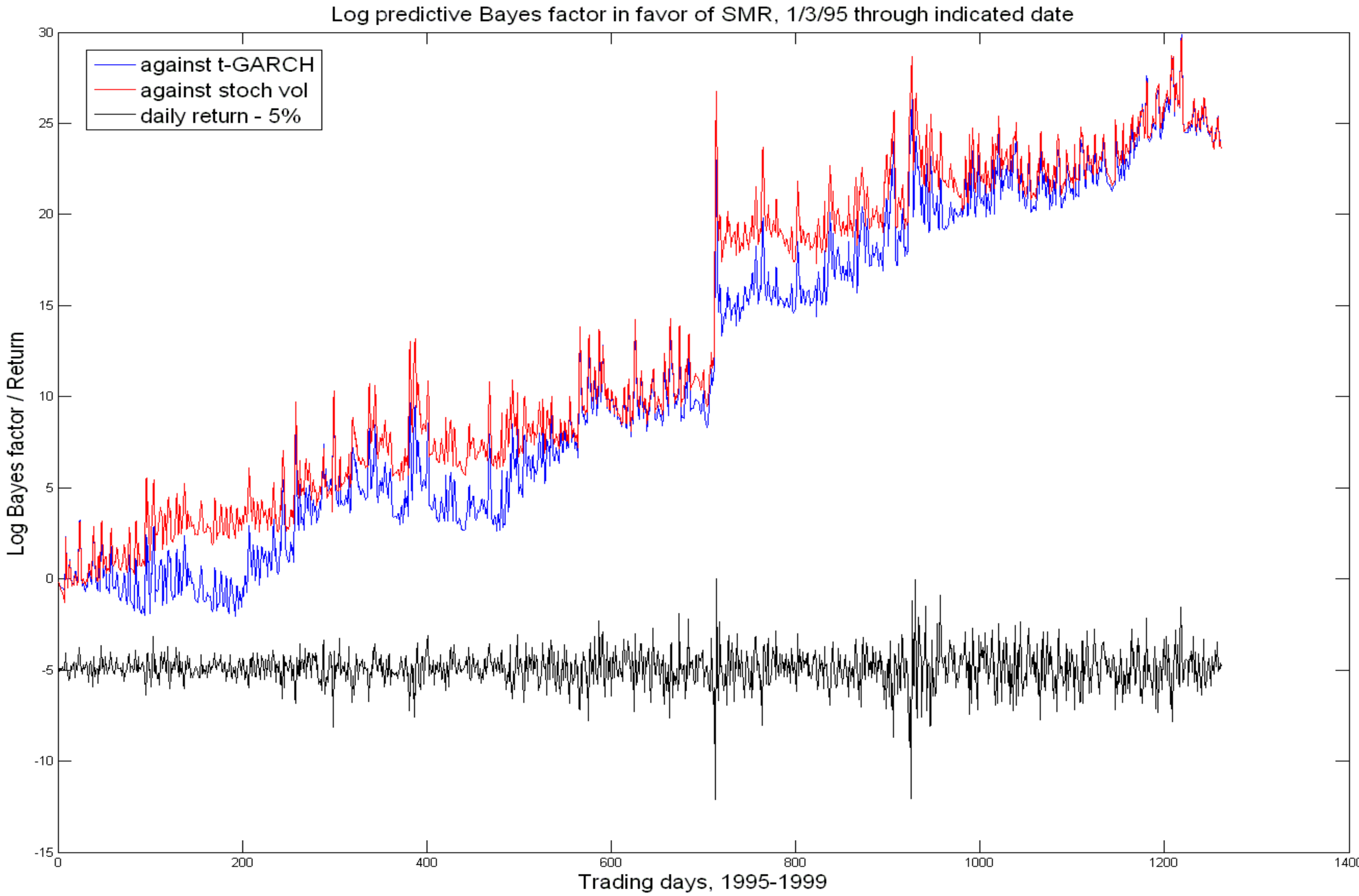
(Beyond paper)

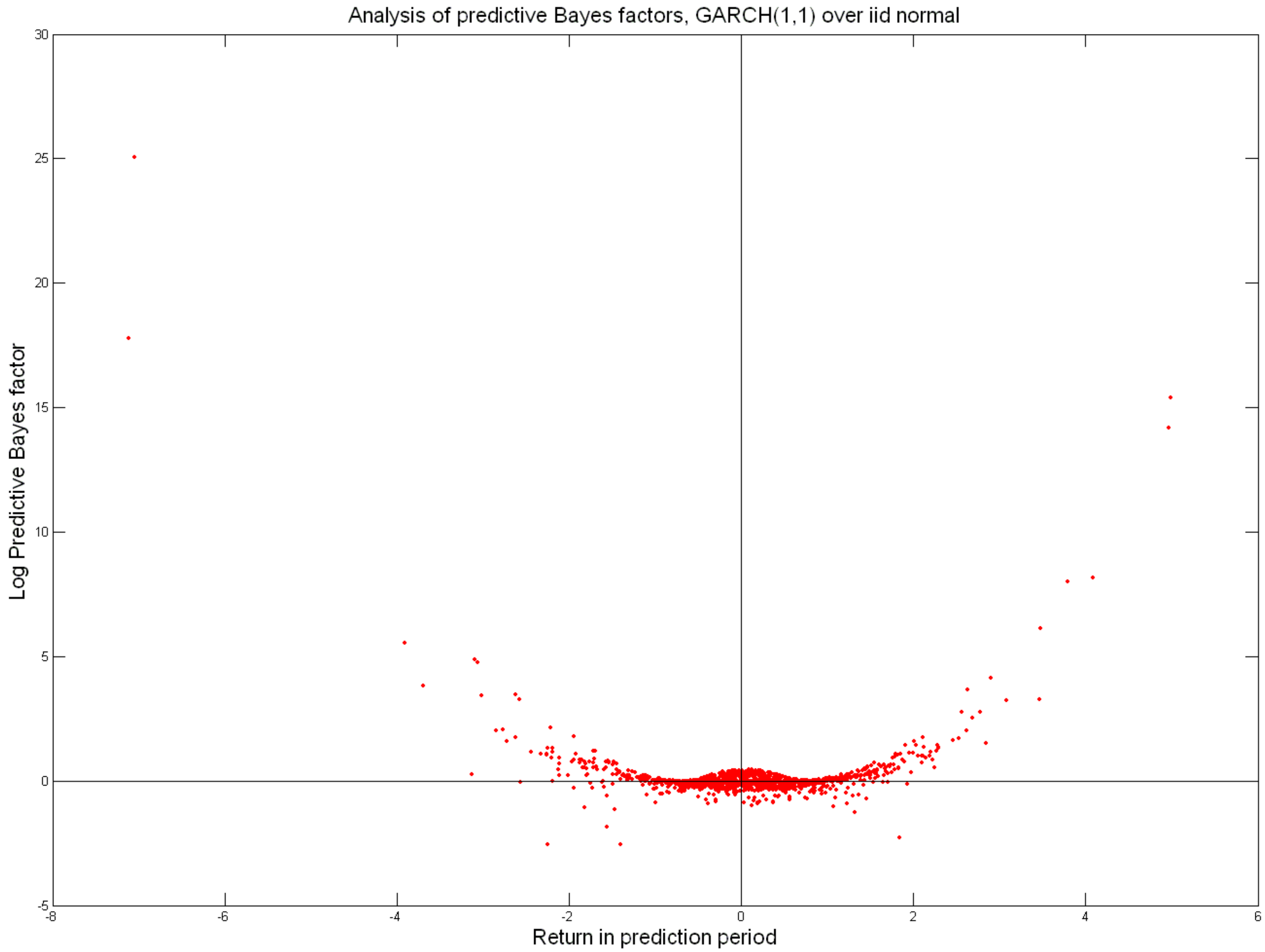
Model evaluation – Log predictive Bayes factors

$$\log p\left(\mathbf{y}^o_{T_1+1}, \ldots \mathbf{y}^o_T \mid \mathbf{Y}^o_{T_1}, A\right) = \sum_{t=T_1+1}^{T} \log p\left(\mathbf{y}^o_t \mid \mathbf{Y}^o_{t-1}, A\right)$$

$$\log \frac{p\left(\mathbf{y}^o_{T_1+1}, \ldots \mathbf{y}^o_T \mid \mathbf{Y}^o_{T_1}, A\right)}{p\left(\mathbf{y}^o_{T_1+1}, \ldots \mathbf{y}^o_T \mid \mathbf{Y}^o_{T_1}, B\right)} = \sum_{t=T_1+1}^{T} \log \frac{p\left(\mathbf{y}^o_t \mid \mathbf{Y}^o_{t-1}, A\right)}{p\left(\mathbf{y}^o_t \mid \mathbf{Y}^o_{t-1}, B\right)}$$

(Beyond paper)




Analysis of predictive Bayes factors, GARCH(1,1) over iid normal
Log Predictive Bayes factor
Return in prediction period
30
25
20
15
10
5
0
-5
-8
-6
-4
-2
0
2
4
6

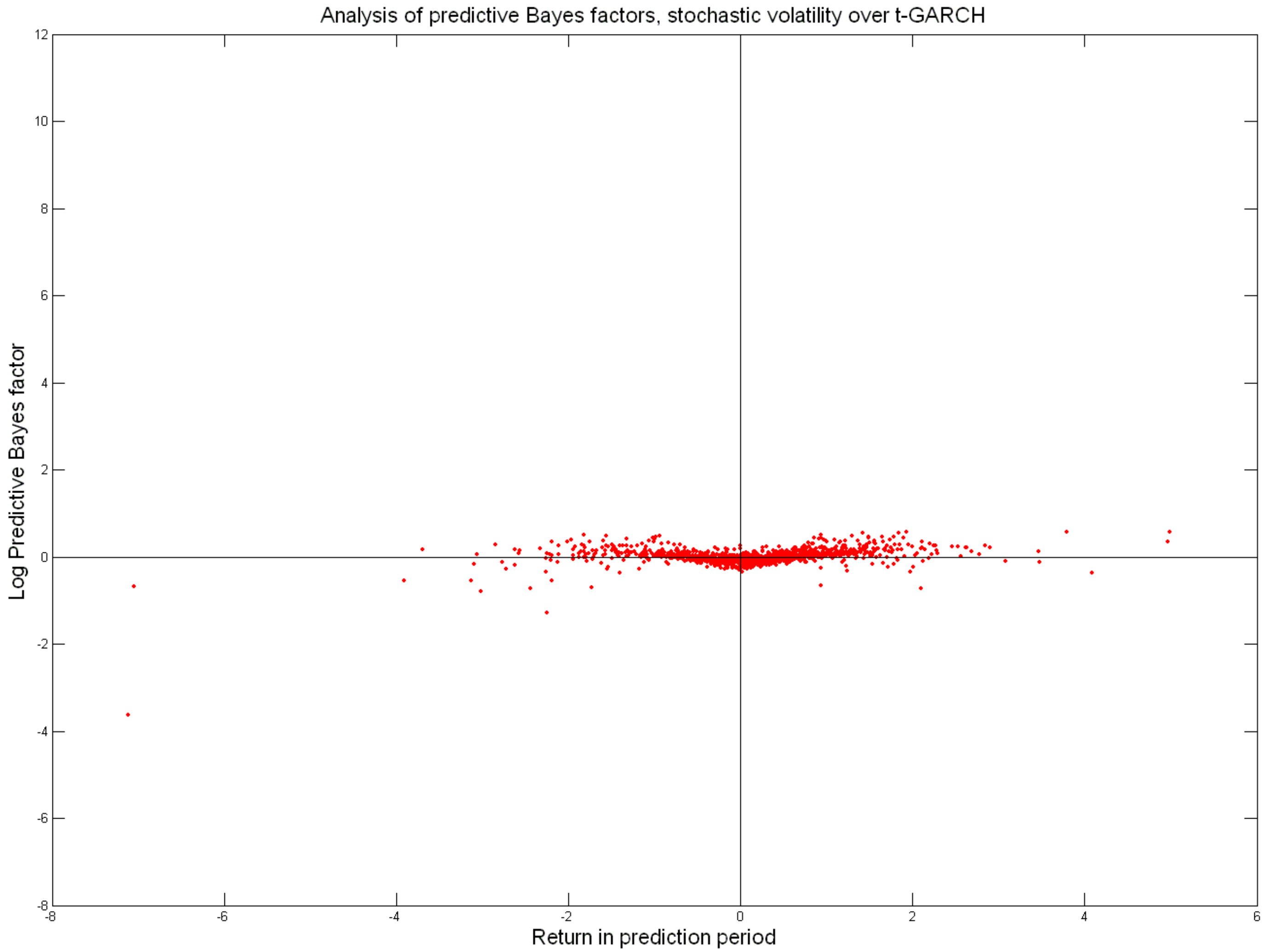
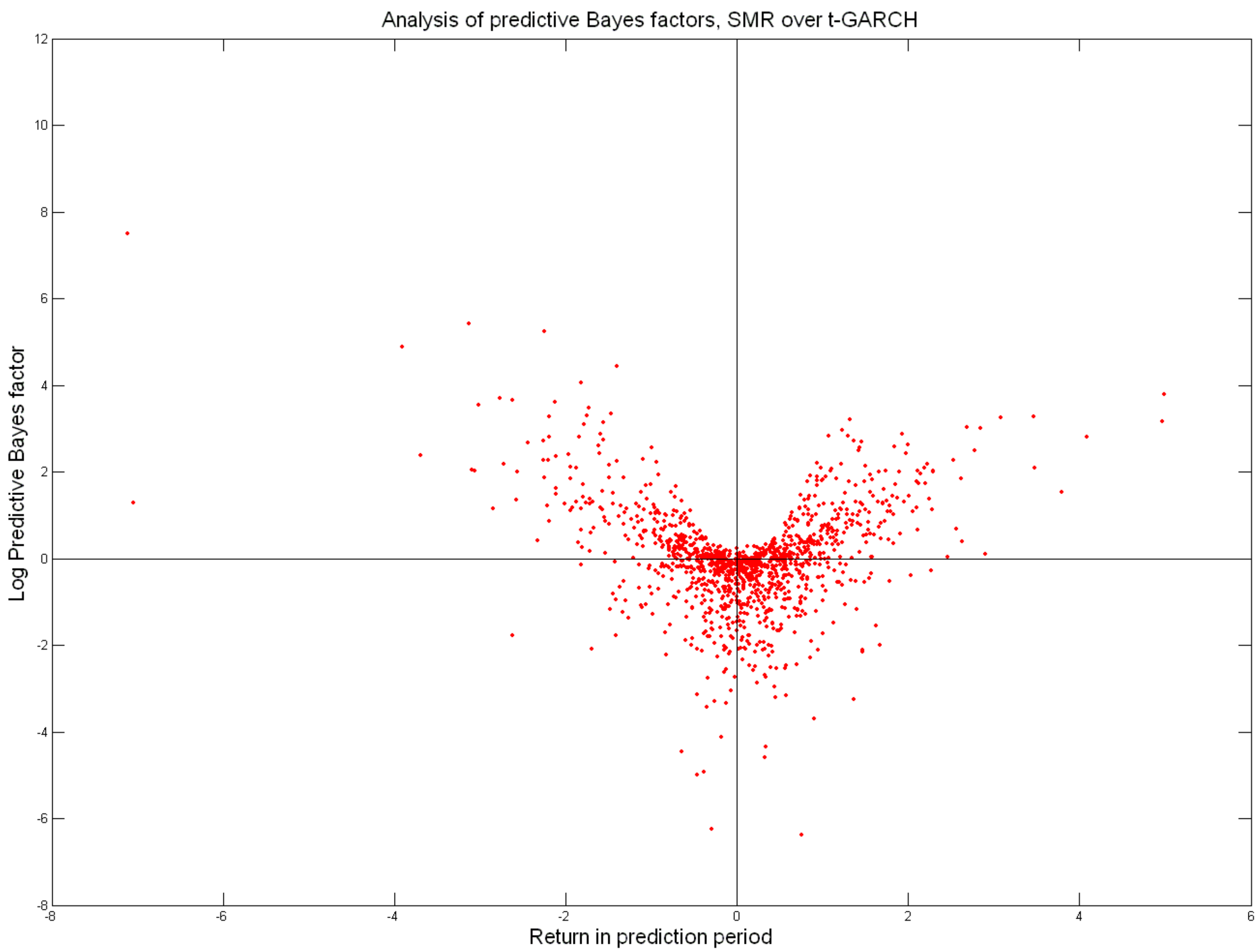
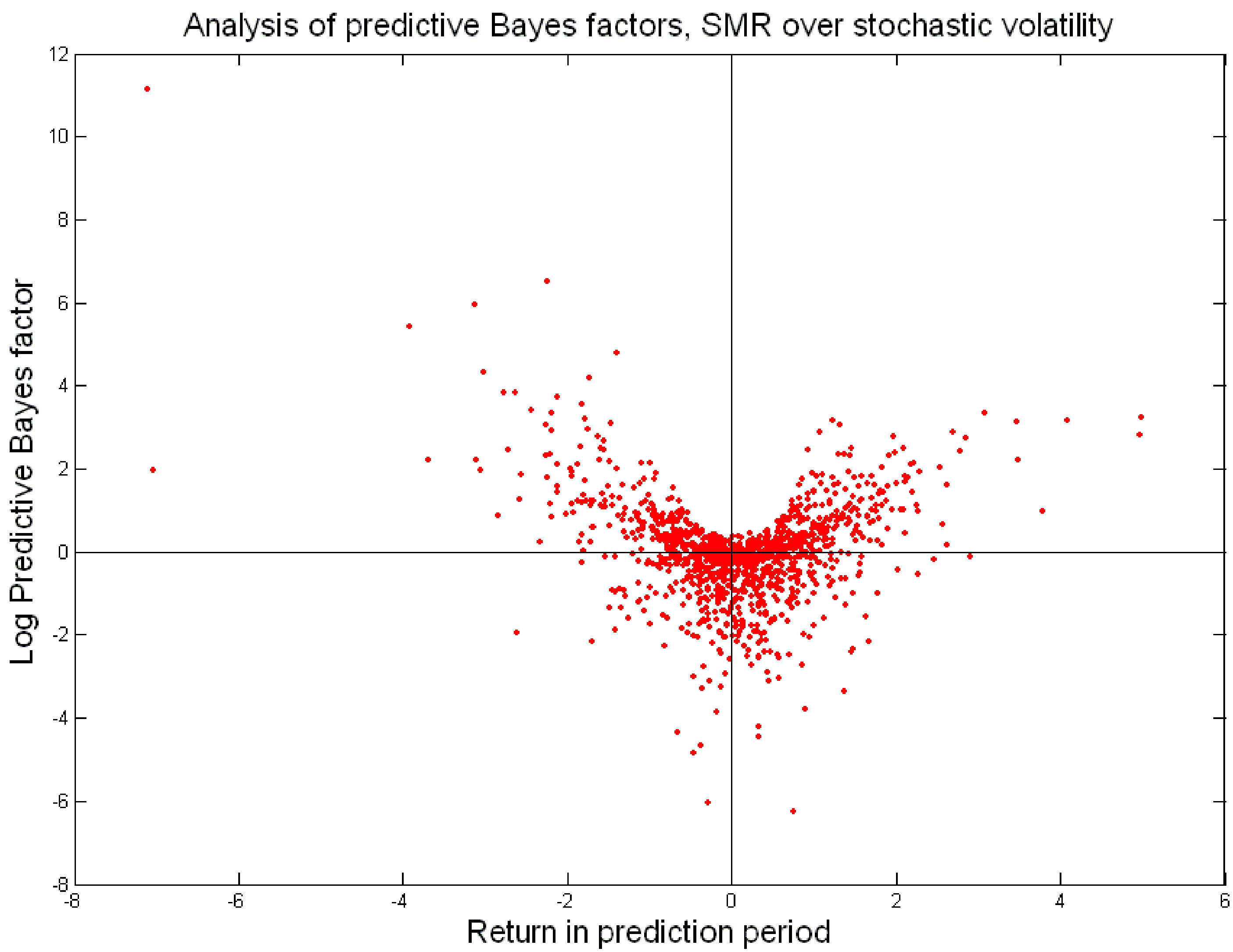

Analysis of predictive Bayes factors, stochastic volatility over t-GARCH
Log Predictive Bayes factor
Return in prediction period
12
10
8
6
4
2
0
-2
-4
-6
-8
-8
-6
-4
-2
0
2
4
6

Figure 1:

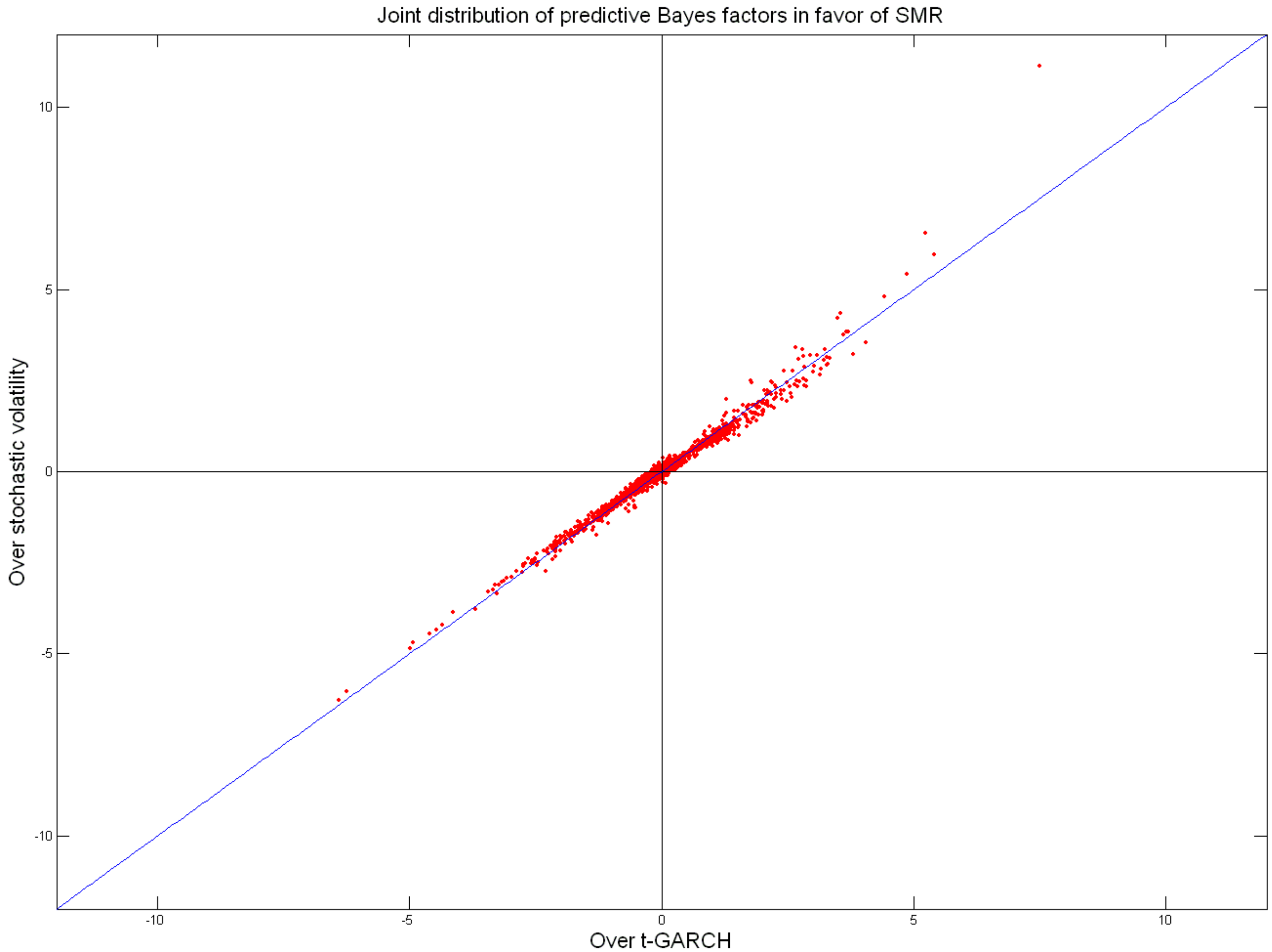
(Beyond paper)



(Beyond paper)



(Beyond paper)



(Beyond paper)

Model evaluation – Posterior Predictive Analysis

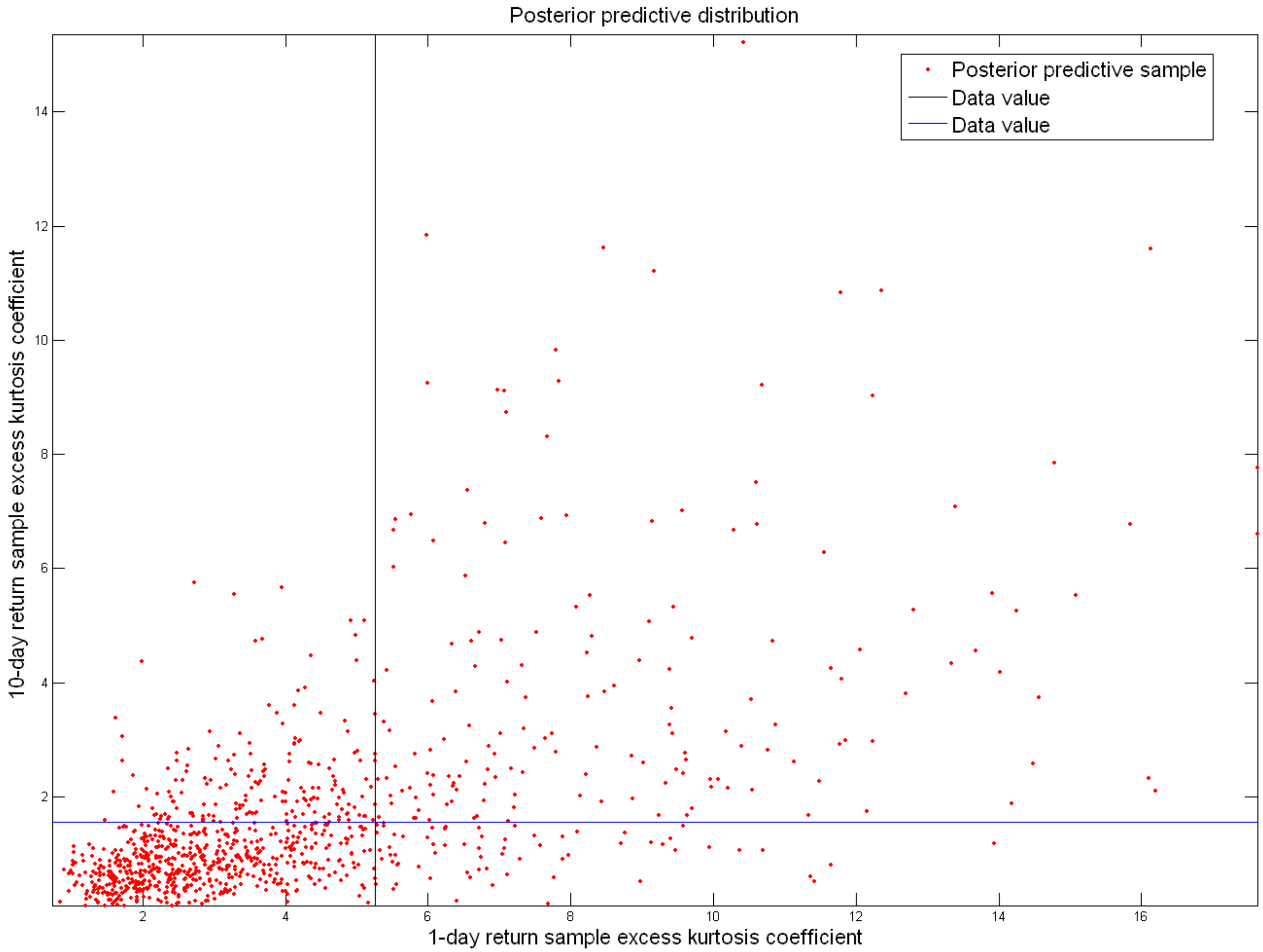
Simulation exercise:

$$\begin{aligned} \boldsymbol{\theta}_A^{(m)} &\sim p\left(\boldsymbol{\theta}_A \mid \mathbf{y}^o, A\right) \\ \mathbf{y}^{(m)} &\sim p\left(\mathbf{y} \mid \boldsymbol{\theta}_A^{(m)}, A\right) \\ h^{(m)} &= h\left(\mathbf{y}^{(m)}\right) \end{aligned}$$

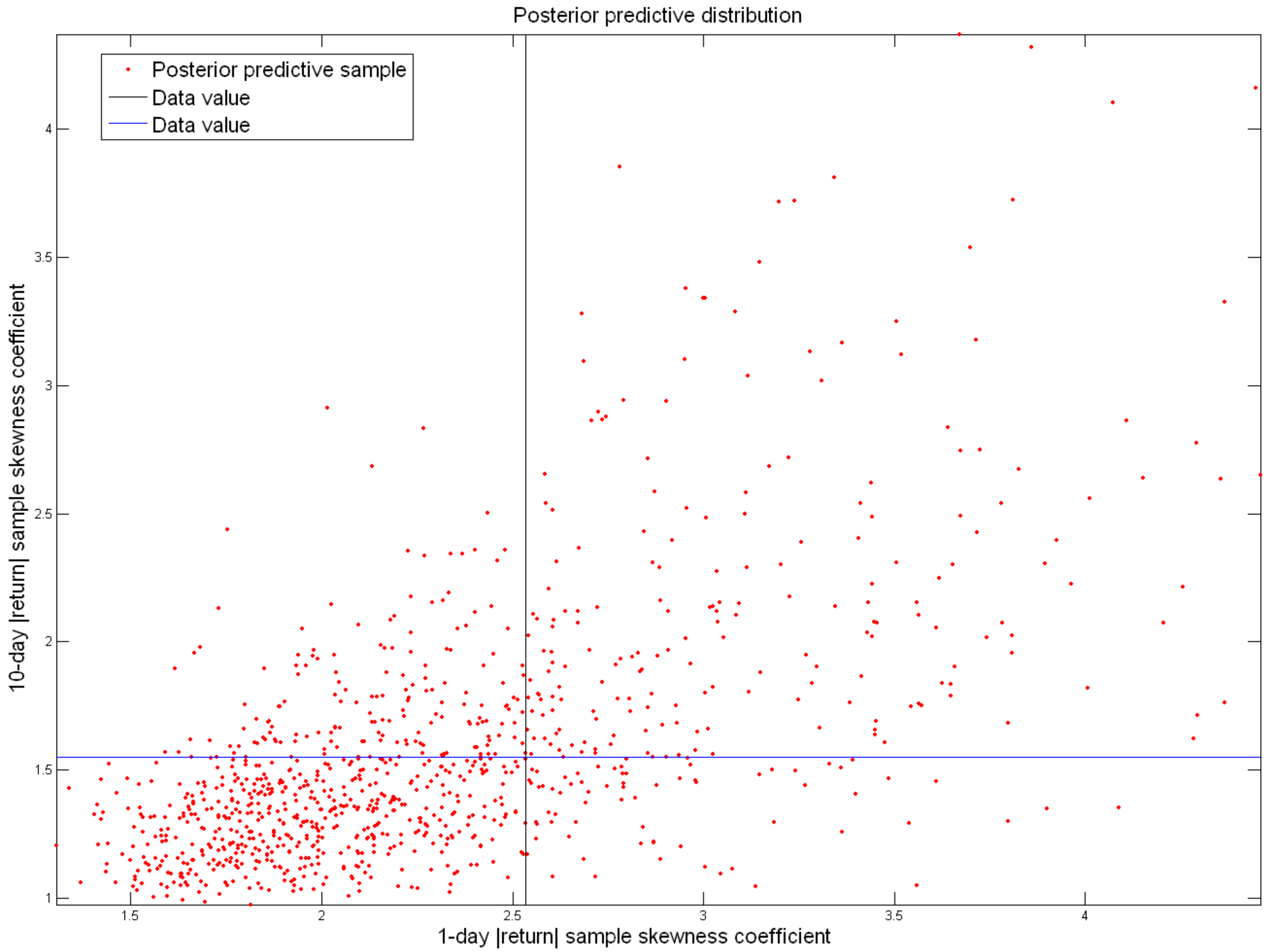
(Beyond paper)

Function of interest (for a 10-year sample)	1990-1999	Inverse cdf
1-day |return| GPH d estimate	0.8609	0.646
1-day return sample variance of return	0.7882	0.594
1-day return sample skewness coefficient	-0.3424	0.849
1-day return sample excess kurtosis coefficient	5.2630	0.241
10-day return sample variance of return	6.7745	0.772
10-day return sample skewness coefficient	-0.2201	0.564
10-day return sample excess kurtosis coefficient	1.5528	0.341
1-day |return| sample variance of return	0.3871	0.532
1-day |return| sample skewness coefficient	2.5343	0.282
1-day |return| sample excess kurtosis coefficient	13.2883	0.247
1-day return .01 quantile	-2.3236	0.482
1-day return .05 quantile	-1.3952	0.429
1-day return .10 quantile	-0.9264	0.346
10-day return .01 quantile	-6.1556	0.255
10-day return .05 quantile	-3.8353	0.315
10-day return .10 quantile	-2.5709	0.261
Correlation 1-day return(t-1) and |return(t)|	-0.1030	0.585

(Beyond paper)



(Beyond paper)



(Beyond paper)