Nonparametric Dynamic Conditional Beta

Maheu, John M and Shamsi, Azam

McMaster University, McMaster University

16 September 2016

Nonparametric Dynamic Conditional Beta*

John M. Maheu † Azam Shamsi ‡

August 25, 2016

Abstract

This paper derives a dynamic conditional beta representation using a Bayesian semiparametric multivariate GARCH model. The conditional joint distribution of excess stock returns and market excess returns are modeled as a countably infinite mixture of normals. This allows for deviations from the elliptic family of distributions. Empirically we find the time-varying beta of a stock nonlinearly depends on the contemporaneous value of excess market returns. In highly volatile markets, beta is almost constant, while in stable markets, the beta coefficient can depend asymmetrically on the market excess return. The model is extended to allow nonlinear dependence in Fama-French factors.

*Maheu thanks the SSHRC of Canada for financial support. We are grateful for comments from Ron Balvers, Martin Burda and Jia Liu. We thank participants of the CFE 2015 conference, Melbourne Bayesian Econometric Workshop 2015 and the Doctoral Workshop in Applied Econometrics, University of Toronto 2016 for helpful comments.

†DeGroote School of Business, McMaster University, Hamilton, ON, Canada. E-mail address: maheujm@mcmaster.ca.

‡DeGroote School of Business, McMaster University, Hamilton, ON, Canada. E-mail address: shamsa3@mcmaster.ca.

1 Introduction

This paper nonparametrically estimates the dynamic conditional beta of a stock using a Bayesian semiparametric multivariate GARCH model. This extends Engle's (2015) parametric version of dynamic conditional beta to the case of an unknown general continuous distribution. In this setting the whole distribution can affect the compensation for risk.

Researchers have long studied the beta coefficient of a stock which represents the nondiversifiable risk arising from exposure to market movements. Traditional approaches estimate the beta coefficient by regressing excess stock returns on the excess market return as in the one-factor Capital Asset Pricing Model (CAPM, Sharpe (1964) and Lintner (1965)), or exploiting more empirically supported asset pricing models, such as Fama-French three-factor model, which incorporate additional explanatory variables (Fama & French (1993)). Our multivariate model nests both cases, but allows for time variation in the conditional second moments. There is a large literature based on multivariate GARCH (MGARCH) models that link a time varying beta to the conditional second moments. Some examples include Bollerslev et al. (1988),Giannopoulos (1995), McCurdy & Morgan (1992) and Choudhry (2002).

Recently Engle (2015) proposes a multivariate normal GARCH model from which the conditional distribution defines the dynamic beta coefficient. This directly links time-varying second moments to the time-varying beta in a consistent fashion. The parametric pricing relationship holds more generally for the elliptic family of distributions. This is an attractive approach but may be limiting if the parametric distributional assumptions are not valid.

A key insight of our approach is that if the joint distribution of excess stock returns and market returns are correctly specified then it follows that their contemporaneous pricing relationship is completely determined by the associated conditional distribution. Therefore, we semiparametrically model the conditional distribution as a countably infinite mixture of normals. Each normal component in the mixture has a conditional covariance directed by a MGARCH process. Our model nests the Gaussian and Student-t distribution as special cases but importantly allows for deviations from the elliptic family of distributions. This includes asymmetric distributions which the elliptic family omit being only symmetric. The mixing is over both the mean vector and covariance matrix.

We follow Jensen and Maheu (2013) to implement a Bayesian semi-parametric MGARCH model and extend it to allow for asymmetric shocks in volatility. The data strongly support the semiparametric MGARCH specification over Gaussian and Student-t distributional alternatives.

In this framework, the conditional distribution of stock returns given the market excess return (and possibly other factors) can be represented as an infinite mixture with weights written as functions of the value of the market excess return. Consequently, the beta coefficient of a security at each time will depend nonparametrically on the contemporaneous value of market return, as opposed to the beta derived from existing models which is insensitive to the contemporaneous value of the market return.

Although the time series of the realized conditional betas from the semiparametric model are similar to the benchmark model we find significant dependence in beta as a function of the contemporaneous value of the market excess return. In the parametric models, beta is constant as a function of the market excess return.

When the market is highly volatile, beta is not affected by unexpected shocks in the market return. While in a calm market, beta can change dramatically from unexpected shocks. For stocks which are highly correlated with the market, an unexpected shock during calm periods increases the beta coefficient. The effect is the reverse for the stocks with low correlation with the market. In other words, when an asset is highly correlated with the market, large moves in a stable market increase the conditional covariance between the market and the asset more than they increase the conditional variance of the market, resulting in a significant increase in the beta coefficient. When an asset has low conditional correlation with the market, large moves in a stable market increase the conditional variance of the market more than they increase the conditional covariance between the market and the asset, leading to a drop in conditional beta. These are important contemporaneous dynamics that are absent in other models.

The remainder of the paper is structured as follows. We begin by reviewing the benchmark model which is an MGARCH model with Student-t innovations. Section 3 provides a general theoretical setting of the multivariate model used in this study. Section 4 summarizes key features of the semiparametric MGARCH model and the use of the Dirichlet process prior. Posterior sampling is detailed in Section 5. The derivation of the nonparametric dynamic conditional beta is presented in Section 6. Data is introduced in Section 7, and Section 8 assesses the performance of the proposed model and compares it to the benchmark model. Applications of the MGARCH-DPM model are found in Section 9 and Section 10 extends the application to the Fama-French three-factor model. Section 11 concludes and an Appendix defines distributions and collects the detailed derivations.

2 Benchmark Model

Our benchmark model is a straightforward extension of Engle (2015). Engle (2015) defines dynamic conditional beta using a multivariate GARCH (MGARCH) model assuming a multivariate normal distribution as the joint density of stock returns and factors. We replace the normal distribution with a Student-t to accommodate the fat-tails in the data. Let the excess stock return on asset i be $r_{i,t}$ and a vector of regressors (factors) including the excess market return be $r_{f,t} = (r_{f_1,t}, r_{f_2,t}, ..., r_{f_q,t})'$. $r_t = (r'_{i,t}, r'_{f,t})'$ is assumed to follow the MGARCH-t

$$r_t|r_{1:t-1} \sim t(\mu, H_t, \nu), \tag{2.1}$$
$$H_t = \Gamma_0 + \Gamma_1 \odot (r_{t-1} - \eta)(r_{t-1} - \eta)' + \Gamma_2 \odot H_{t-1}, \tag{2.2}$$

where $t(\mu, \Sigma, \nu)$ denotes a t-distribution (see appendix) with mean vector μ, scale matrix Σ and degree of freedom ν and $r_{1:t-1} = \{r_1, \ldots, r_{t-1}\}$ is the information set available at time $t-1$. The scale matrix, H_t, is based on the vector-diagonal multivariate GARCH model of Ding & Engle (2001) but other MGARCH formulations could be used. The symbol $\odot$ denotes the Hadamard product. The parameter is $\Gamma = \{\Gamma_0, \Gamma_1, \Gamma_2, \eta\}$, with the symmetric positive definite matrices parameterized as $\Gamma_0 = \Gamma_0^{1/2}(\Gamma_0^{1/2})'$, $\Gamma_1 = \gamma_1(\gamma_1)'$, and $\Gamma_2 = \gamma_2(\gamma_2)'$ where Γ_0 is a lower triangular $(q+1) \times (q+1)$ matrix and γ_1, γ_2 and η are $(q+1)$-vectors. η permits a nonlinear asymmetric response to shocks and can be considered a multivariate version of the asymmetric GARCH model (Engle & Ng 1993).

Partition $r_t = (r'_{1,t}, r'_{2,t})'$ into a k_1 and k_2 $(k_1 + k_2 = q + 1)$ vector and similarly

$\mu = (\mu_1', \mu_2')'$ and

$$H_t = \begin{bmatrix} H_{11,t} & H_{12,t} \\ H_{12,t} & H_{22,t} \end{bmatrix}.$$

Applying the properties of the Student-t distribution (Roth 2013) the conditional distribution of $r_{1,t}$ given $r_{2,t}$ is

$$\begin{aligned} r_{1,t}|r_{2,t} &\sim t(\mu_{1|2}, H_{t,1|2}, \nu_{1|2}), && (2.3)\\ \mu_{1|2} &= \mu_1 + H_{12,t}H_{22,t}^{-1}(r_{2,t} - \mu_2), && (2.4)\\ H_{t,1|2} &= \frac{\nu + (r_{2,t} - \mu_2)' H_{22,t}^{-1}(r_{2,t} - \mu_2)}{\nu + k_2}(H_{11,t} - H_{12,t}H_{22,t}^{-1}H_{12,t}'), && (2.5)\\ \nu_{1|2} &= \nu + k_2, && (2.6) \end{aligned}$$

where the conditional mean is $\mu_{1|2}$, the conditional scale matrix is $H_{t,1|2}$ and the degree of freedom $\nu_{1|2}$.

This is a useful result in that it tells us how the conditional distribution of $r_{1,t}$ reacts to any value of $r_{2,t}$. For instance, if $r_{1,t} \equiv r_{i,t}$ and conditioning on one factor, the excess market return, $r_{2,t} \equiv r_{m,t}$, substituting into (2.4) directly gives a dynamic risk premium for asset i as

$$E[r_{i,t}|r_{m,t}, H_t] = \mu_i + H_{12,t}H_{22,t}^{-1}(r_{m,t} - \mu_m). \qquad (2.7)$$

This tells how the expected excess return of asset i reacts to any value of the market. If the market shock is zero ($r_{m,t} = \mu_m$) then the expected value is μ_i but for all other realizations the market shock impacts the expected return of the asset. Engle identifies the dynamic conditional beta that arises from the joint relationship as

$$\beta_t = H_{22,t}^{-1}H_{12,t}. \qquad (2.8)$$

This is the derivative of (2.7) with respect to $r_{m,t}$. A conditional pricing relationship is obtained by setting $r_{2,t} \equiv E[r_{m,t}|r_{1:t-1}]$ and substituting into (2.7).

There are several advantage to modeling excess returns in this way. First, it confronts the simultaneous nature of the asset return and the factors that price the risk premium. Rather than specifying a single equation partial equilibrium relationship the model begins with the full joint dynamics. Second, the joint distribution of the asset and the factors directly pin down the conditional distribution and the implications for the risk premium. The dynamic beta is a function of the conditional covariance matrix. This is a general result that holds for the elliptic family of distributions.

The model is estimated from a Bayesian perspective. The posterior density has the non-standard form

$$p(\mu, \Gamma, \nu|r_{1:T}) \propto p(\nu)p(\mu)p(\Gamma) \times \prod_{t=1}^{T} t(r_t|\mu, H_t, \nu), \qquad (2.9)$$

where $t(r_t|\mu, H_t, \nu)$ is the density of the Student-t distribution, and $p(\nu)p(\mu)p(\Gamma)$ is the prior density for μ, Γ, ν. Posterior draws of the parameters vector are simulated with a Metropolis-Hastings sampler.

Although attractive, the conditional distribution in (2.3) has some drawbacks. The conditional beta derived from MGARCH-t model, at each time, is constant with respect to the contemporaneous value of market return (Equation 2.8), and consequently, the conditional expected return of the stock is a linear function of the factor returns. This pricing relationship will not hold for more general distributions not belonging to the elliptic family. The elliptic family of distributions are symmetric about their mean and do not account for asymmetry observed in financial returns.

This model imposes a strong assumption on the functional form of the joint distribution of the data. In this paper, we remove this restrictive assumption by employing a Dirichlet process mixture (DPM) to model the unknown joint distribution of returns. This results in a potentially non-constant conditional beta and a nonlinear conditional expected value of the stock as a function of the contemporaneous value of the market return.

3 MGARCH-DPM Model

Unlike the benchmark model that assumes a specific parametric joint distribution for the individual asset returns and the factors, we model this joint distribution nonparametrically by an infinite mixture of normal distributions which can approximate any continuous multivariate distribution. Recall that $r_t = (r_{i,t}, r_{f_1,t}, ..., r_{f_q,t})'$ represents the excess return vector of an individual stock and q factors at time t. The infinite mixture representation can be written as

$$r_t | H_t, \mu, B, W \sim \sum_{j=1}^{\infty} \omega_j N(\mu_j, (H_t^{1/2}) B_j (H_t^{1/2})'). \tag{3.1}$$

where $H_t^{1/2}$ is the Cholesky decomposition of H_t, $\mu = \{\mu_1, \mu_2, \dots\}$, $B = \{B_1, B_2, \dots\}$ and $W = \{\omega_1, \omega_2, \dots\}$ is the vector of the weights, such that $\omega_j \geq 0$ for all j and $\sum_{j=1}^{\infty} \omega_j = 1$. The mixing is over the mean vector and the component B_j of the covariance matrix. The second component, H_t of the covariance matrix captures volatility clustering through time but is not a function of j.

The conditional mean can be derived in exactly the same way as in the benchmark model except it will follow an infinite mixture of conditional normal distributions. If $r_{ft} = (r_{f_1,t}, ..., r_{f_q,t})'$ then the conditional density of $r_{i,t}$ given $r_{f,t}$ is a mixture distribution as well and the conditional expectation can be written as the following weighted mixture

$$E(r_{i,t} | r_{f,t}, H_t) = \sum_{j=1}^{\infty} q_j(r_{f,t}) E(r_{i,t} | r_{f,t}, \mu_j, B_j, H_t). \tag{3.2}$$

The weights, $q_j(r_{f,t})$ are a function of the factors and affect how much each conditional expectation, $E(r_{i,t} | r_{f,t}, \mu_j, B_j, H_t)$, in the mixture contributes. The details on the derivations will be explained later but for now it is important to see that unlike the parametric model the conditional expectation is not a linear function of the factors. To obtain the nonparametric conditional beta, we take the derivative of (3.2) with respect to the desired factor. The conditional beta is not constant in general but it changes as the contemporaneous value of the corresponding factor changes. The next section introduces the Dirichlet process prior to estimate this model. In Section 6 we derive the nonparametric conditional beta.

4 A Bayesian Model

In Bayesian inference the Dirichlet process (DP) prior (Ferguson 1973) is a standard prior used for infinite dimensional objects such as (3.1). A draw from a DP, $G \sim DP(\alpha, G_0)$, is almost surely a discrete distribution and is governed by two parameters. The concentration parameter α, a positive scalar and a base distribution G_0. The nonparametric distribution G is centered on the base distribution G_0, which can be considered as the prior guess; $E(G) = G_0$. The concentration parameter measures the strength of belief in G_0. The larger α, the stronger belief in G_0 and the more distinct elements we have with appreciable mass. Lo (1984) introduces Dirichlet process mixture (DPM) model in which G is the mixing measure over a continuous kernel. This has become a standard Bayesian approach to nonparametric estimation of an unknown continuous distribution. In this paper, G is the unknown distribution that governs the mixing over the mean vector and covariance matrix of the normal kernel in our mixture model.

The model (MGARCH-DPM) is an extension of Jensen & Maheu (2013) and allows for asymmetry in the MGARCH process from shocks to volatility and fat tails without making any restrictive assumption. The hierarchical form of the model is,

$$r_t|\phi_t, H_t \sim N(\xi_t, H_t^{1/2}\Lambda_t(H_t^{1/2})'), \; t = 1, ..., T \tag{4.1}$$
$$\phi_t \equiv \{\xi_t, \Lambda_t\}|G \sim G, \tag{4.2}$$
$$G|\alpha, G_0 \sim DP(\alpha, G_0), \tag{4.3}$$
$$G_0 \equiv N(\mu_0, D) \times \mathcal{W}^{-1}(B_0, \nu_0), \tag{4.4}$$
$$H_t = \Gamma_0 + \Gamma_1 \odot (r_{t-1} - \eta)(r_{t-1} - \eta)' + \Gamma_2 \odot H_{t-1}. \tag{4.5}$$

In this model ξ_t is a $(q+1)$-vector and Λ_t is a symmetric positive definite matrix and H_t follows the same MGARCH specification as the benchmark parametric model. $\mathcal{W}^{-1}(B_0, \nu_0)$ represents an inverse Wishart distribution (see appendix) with scale matrix B_0 and degree of freedom ν_0.

Sethuraman (1994) characterizes a stick-breaking representation of the DP. Combining this with the normal kernel gives the associated stick breaking representation of the MGARCH-DPM density as

$$p(r_t|\mu, B, W, H_t) = \sum_{j=1}^{\infty} \omega_j N(r_t|\mu_j, H_t^{1/2}B_j(H_t^{1/2})'), \tag{4.6}$$
$$\omega_1 = v_1, \; \omega_j = v_j \prod_{l=1}^{j-1}(1 - v_l), \; j > 1, \tag{4.7}$$
$$v_j \stackrel{iid}{\sim} \text{Beta}(1, \alpha), \tag{4.8}$$
$$\mu_j \stackrel{iid}{\sim} N(\mu_0, D), \quad B_j \stackrel{iid}{\sim} \mathcal{W}^{-1}(B_0, \nu_0), \tag{4.9}$$

where $N(r_t|\mu_j, H_t^{1/2}B_j(H_t^{1/2})')$ denotes the multivariate normal density with mean μ_j and covariance $H_t^{1/2}B_j(H_t^{1/2})'$ evaluated at r_t. Note that μ and B are the set of unique points of support in the discrete distribution G while ξ_t and Λ_t denote draws from G in (4.2), with the possibility of repeated draws of μ_j and B_j.

The model nests several special cases. First, the Gaussian model is obtained when $\alpha \to 0$ as $\omega_1 = 1$, $\omega_j = 0, \forall j > 1$ and $B_1 = I$. The Student-t model results from μ_j being constant for all j and $\alpha \to \infty$, since $G \to G_0$, the inverse Wishart distribution.

5 Posterior Sampling

To estimate the unknown parameters in (4.1)-(4.5), we apply an MCMC sampler along with the slice sampler of Walker (2007) and Kalli et al. (2011). Slice sampling introduces a latent variable, $u_t \in (0,1)$, to elegantly convert an infinite sum to a finite mixture model which makes the sampling feasible. Estimating the joint posterior density of u_t and other model parameters and then integrating out the slice variable u_t recovers the desired posterior density. In practice, this means jointly sampling all parameters including the slice variable but then discarding u_t. Define u_t such that the joint density of (r_t, u_t) given $(W, \Theta \equiv (\mu, B))$ is given by

$$f(r_t, u_t|W, \Theta) = \sum_{j=1}^{\infty} \mathbf{1}(u_t < \omega_j) N(r_t|\mu_j, (H_t^{1/2})' B_j H_t^{1/2}). \tag{5.1}$$

Let $s_{1:T} = \{s_1, ..., s_T\}$ be the configuration set that partitions the data $r_{1:T}$ into c distinct clusters such that observation r_t is assigned parameter $\theta_{s_t} = (\mu_{s_t}, B_{s_t})$. Let $n_j = \{\#t|s_t = j\}$ be the number of observations allocated to state j. The full likelihood is

$$p(r_{1:T}, u_{1:T}, s_{1:T}|W, \Theta) = \Pi_{t=1}^T \mathbf{1}(u_t < \omega_{s_t}) N(r_t|\mu_{s_t}, (H_t^{1/2}) B_{s_t} (H_t^{1/2})'), \tag{5.2}$$

and the joint posterior is proportional to

$$p(W_{1:K}) \Pi_{j=1}^K p(\mu_j, B_j) \Pi_{t=1}^T \mathbf{1}(u_t < \omega_{s_t}) N(r_t|\mu_{s_t}, (H_t^{1/2}) B_{s_t} (H_t^{1/2})') \tag{5.3}$$

where K is the smallest natural number that satisfies the condition $\sum_{j=1}^K \omega_j > 1 - \min\{u_t\}_{t=1}^T$ and $W_{1:K}$ denotes the finite set of W and similarly for other parameters $\mu_{1:K}$ and $B_{1:K}$. Having defined the notation, the steps of the MCMC algorithm are discussed next.

Steps of MCMC algorithm for MGARCH-DPM

1. The posterior distribution of $\theta_j = (\mu_j, B_j)$, $j = 1, ..., K$: Using the transformation $z_t = H_t^{-1/2} r_t$, and applying the results of conditionally conjugate priors for the linear regression model we have

$$B_j|r_{1:T}, s_{1:T}, \mu_j, \Gamma \sim \mathcal{W}^{-1}\left(n_j + \nu_0, B_0 + \sum_{s_t=j} (z_t - H_t^{-1/2}\mu_j)(z_t - H_t^{-1/2}\mu_j)'\right) \tag{5.4}$$

$$\mu_j|r_{1:T}, s_{1:T}, B_j, \Gamma \sim N(\bar{\mu}, \bar{D}) \tag{5.5}$$

in which

$$\bar{D}^{-1} = D^{-1} + \sum_{t|s_t=j} H_t^{-1/2'} B_j^{-1} H_t^{-1/2}, \ \bar{\mu} = \bar{D}\left(\sum_{t|s_t=j} H_t^{-1/2'} B_j^{-1} z_t + D^{-1}\mu_0\right). \tag{5.6}$$

2. Updating $v_j,\ j = 1, ..., K$.

$$v_j|S \sim \text{Beta}\left(1 + \sum_{t=1}^{T} \mathbf{1}(s_t = j), \alpha + \sum_{t=1}^{T} \mathbf{1}(s_t > j)\right). \tag{5.7}$$

   Then we update $W_{1:K}$ based on (4.7).

3. Updating $u_t,\ t = 1, ..., T$. $u_t|s_{1:T} \sim \mathcal{U}(0, \omega_{s_t})$. Then update K such that $\sum_{j=1}^{K} \omega_j > 1 - \min\{u_t\}_{t=1}^{T}$. Additional ω_j and θ_j will need to be generated from the priors if K is incremented.

4. Updating $s_{1:T}$. For each $t = 1, ..., T$,

$$p(s_t = j|r_{1:T}) \propto \mathbf{1}(\omega_j > u_t) N(r_t|\mu_j, H_t^{1/2} B_j (H_t^{1/2})'), j = 1, ..., K. \tag{5.8}$$

5. Updating α: Assuming a gamma prior $\alpha \sim \mathcal{G}(a_0, b_0)$ (see appendix) α can be sampled following the two steps below (Escobar & West 1995). Recall that c is the number of alive clusters defined as the number of clusters in which at least one observation is allocated. Note that $c \leq K$. Then the sampling steps are as follows.

   (a) $(\tau|\alpha, c) \sim \text{Beta}(\alpha + 1, T)$.

   (b) Sample α from

$$\alpha|\tau \sim \pi_\tau \mathcal{G}(a_0 + c, b_0 - \log(\tau)) + (1 - \pi_\tau)\mathcal{G}(a_0 + c - 1, b_0 - \log(\tau)),$$

   where π_τ is defined by $\frac{\pi_\tau}{1-\pi_\tau} = \frac{a_0+c-1}{T(b_0-\log(\tau))}$.

6. Updating GARCH parameters $\Gamma = (\Gamma_0^{1/2}, \gamma_1, \gamma_2, \eta)$. The conditional posterior is

$$p(\Gamma|\mu, B, S, r_{1:T}, H_{1:T}) \propto p(\Gamma) \times \prod_{t=1}^{T} N(r_t|\mu_{s_t}, H_t^{1/2} B_{s_t} (H_t^{1/2})') \tag{5.9}$$

   which is not of standard form, and we apply a Metropolis-Hastings sampler. Given the current value Γ of the chain, the proposal Γ' is sampled $\Gamma' \sim N(\Gamma, \widehat{V})$. The draw is accepted with probability

$$\min\{p(\Gamma'|\mu, B, S, r_{1:T}, H_{1:T})/p(\Gamma|\mu, B, S, r_{1:T}, H_{1:T}), 1\},$$

   and otherwise rejected. $\widehat{V}$ is proportional to the inverse Hessian matrix of $\ell = \log[p(\Gamma|\mu, B, S, r_{1:T}, H_{1:T})]$ evaluated at its posterior mode, $\widehat{\Gamma}$, which is computed once at the start of estimation. $\widehat{V}$ is scaled to achieve an acceptance rate between 0.2 and 0.5. In this paper we apply Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm to approximate the posterior mode of ℓ.

6 Nonparametric Dynamic Conditional Beta

To study the behaviour of the conditional beta of an individual stock, we first consider a special case of our model, $r_t = (r_{i,t}, r_{m,t})$ where $r_{i,t}$ and $r_{m,t}$ represent an individual stock's excess return and the market excess return, respectively. Applying the posterior sampling algorithm, we sample model parameters for many iterations and after dropping a set of burn-in draws we have the following set of sampled parameters:

$$\{(\mu_j^{(g)}, B_j^{(g)}), v_j^{(g)}, j = 1, ..., K^{(g)}\}, \{s_t^{(g)}, u_t^{(g)}, t = 1, ..., T\}, H_{1:T}^{(g)} = \{H_1^{(g)}, ..., H_T^{(g)}\}, \quad (6.1)$$

for $g = 1, ..., M$ where M is the number of MCMC iterations. At each iteration $g = 1, ..., M$ of the algorithm, a draw of $G|r_{1:T}$, can be written as

$$G^{(g)} = \sum_{j=1}^{K^{(g)}} \omega_j^{(g)} \delta_{\theta_j^{(g)}} + \left(1 - \sum_{j=1}^{K^{(g)}} \omega_j^{(g)}\right) G_0(\theta), \quad (6.2)$$

where $\theta_j^{(g)} = (\mu_j^{(g)}, B_j^{(g)})$ and $\delta_{\theta_j^{(g)}}$ is a mass point at $\theta_j^{(g)}$.

Combining this with the normal kernel gives the predictive density for the generic return $(\mathbf{r}_{i,t}, \mathbf{r}_{m,t})$ conditional on $G^{(g)}$ as

$$p(\mathbf{r}_{i,t}, \mathbf{r}_{m,t}|r_{1:T}, G^{(g)}) = \sum_{j=1}^{K^{(g)}} \omega_j^{(g)} f(\mathbf{r}_{i,t}, \mathbf{r}_{m,t}|\theta_j^{(g)}) + \left(1 - \sum_{j=1}^{K^{(g)}} \omega_j^{(g)}\right) \int f(\mathbf{r}_{i,t}, \mathbf{r}_{m,t}|\theta) G_0(\theta) d\theta, \quad (6.3)$$

where $f(\mathbf{r}_{i,t}, \mathbf{r}_{m,t}|\theta)$ is the multivariate normal density.

To assess the nonlinear regression function $E(\mathbf{r}_{i,t}|\mathbf{r}_{m,t}, r_{1:T})$, or the conditional beta of the individual stock i, we require the conditional density derived from this predictive (joint) density of $(\mathbf{r}_{i,t}, \mathbf{r}_{m,t})$. Therefore,

$$\begin{aligned} p(\mathbf{r}_{i,t}|\mathbf{r}_{m,t}, r_{1:T}, G^{(g)}) &= \frac{p(\mathbf{r}_{i,t}, \mathbf{r}_{m,t}|r_{1:T}, G^{(g)})}{p(\mathbf{r}_{m,t}|r_{1:T}, G^{(g)})} \quad (6.4) \\ &= \frac{p(\mathbf{r}_{i,t}, \mathbf{r}_{m,t}|r_{1:T}, G^{(g)})}{\sum_{j=1}^{K^{(g)}} \omega_j^{(g)} f_2(\mathbf{r}_{m,t}|\theta_j^{(g)}) + \left(1 - \sum_{j=1}^{K^{(g)}} \omega_j^{(g)}\right) \int f_2(\mathbf{r}_{m,t}|\theta) G_0(\theta) d\theta} \\ &= \sum_{j=1}^{K^{(g)}} q_j^{(g)}(\mathbf{r}_{m,t}) f(\mathbf{r}_{i,t}|\mathbf{r}_{m,t}, \theta_j^{(g)}) + \left(1 - \sum_{j=1}^{K^{(g)}} q_j^{(g)}(\mathbf{r}_{m,t})\right) f(\mathbf{r}_{i,t}|\mathbf{r}_{m,t}, G_0), \quad (6.5) \end{aligned}$$

where

$$q_j^{(g)}(\mathbf{r}_{m,t}) = \frac{\omega_j^{(g)} f_2(\mathbf{r}_{m,t}|\theta_j^{(g)})}{\sum_{j=1}^{K^{(g)}} \omega_j^{(g)} f_2(\mathbf{r}_{m,t}|\theta_j^{(g)}) + \left(1 - \sum_{j=1}^{K^{(g)}} \omega_j^{(g)}\right) \int f_2(\mathbf{r}_{m,t}|\theta) G_0(\theta) d\theta} \quad (6.6)$$

and $f_2(\mathbf{r}_{m,t}|\theta_j^{(g)})$ is the marginal (normal) density of $\mathbf{r}_{m,t}$ and $f(\mathbf{r}_{i,t}|\mathbf{r}_{m,t}, G_0)$ is the conditional distribution using the base measure. The terms $q_j^{(g)}(\mathbf{r}_{m,t})$ determine which components in the mixture receive more weight. Clusters that have a marginal density $f_2(\mathbf{r}_{m,t}|\theta_j^{(g)})$ that has a higher likelihood value for $\mathbf{r}_{m,t}$ will receive larger weights. The marginal density, and hence relative weight of clusters, will change with $\mathbf{r}_{m,t}$ as well as over time through the MGARCH component, H_t. These features will determine the relative weights on the cluster specific conditional expectations which we derive next.

Our focus is on the conditional mean of $\mathbf{r}_{i,t}$ given $\mathbf{r}_{m,t}$. Using the properties of the normal distribution the conditional mean directly comes from (6.5) and is

$$E(\mathbf{r}_{i,t}|\mathbf{r}_{i,t}, r_{1:T}, G^{(g)}) = \sum_{j=1}^{K^{(g)}} q_j^{(g)}(\mathbf{r}_{m,t})[\mu_{j,1}^{(g)} + \beta_{jt}^{(g)}(\mathbf{r}_{m,t} - \mu_{j,2}^{(g)})] + \tag{6.7}$$

$$\left(1 - \sum_{j=1}^{K^{(g)}} q_j^{(g)}(\mathbf{r}_{m,t})\right) \frac{\int [\mu_1 + \beta_t(\mathbf{r}_{m,t} - \mu_2)] N(\mathbf{r}_{m,t}|\mu_2, (H_t^{(g)^{1/2}} B H_t^{(g)^{1/2'}})_{22}) p(\mu, B) d\mu dB}{\int N(\mathbf{r}_{m,t}|\mu_2, (H_t^{(g)^{1/2}} B H_t^{(g)^{1/2'}})_{22}) p(\mu, B) d\mu dB}.$$

The cluster specific beta is defined as

$$\beta_{jt}^{(g)} = \frac{(H_t^{(g)^{1/2}} B_j H_t^{(g)^{1/2'}})_{12}}{(H_t^{(g)^{1/2}} B_j H_t^{(g)^{1/2'}})_{22}} \tag{6.8}$$

where the subscript (i, j) on $()_{ij}$ denotes element (i, j) of the matrix and β_t in the second line of (6.7) is defined as $\beta_{jt}^{(g)}$ except B_j is replaced with B. The numerator and denominator in the last term of (6.7) can be approximated by simulation.

Integrating all parameter and distributional uncertainty results in an estimate of the predictive conditional mean as

$$E(\mathbf{r}_{i,t}|\mathbf{r}_{m,t}, r_{1:T}) \approx \frac{1}{M} \sum_{g=1}^{M} E(\mathbf{r}_{i,t}|\mathbf{r}_{m,t}, r_{1:T}, G^{(g)}). \tag{6.9}$$

The predictive mean of the conditional beta is the derivative of this conditional expectation of $\mathbf{r}_{i,t}$ given $\mathbf{r}_{m,t}$, (6.9) with respect to $\mathbf{r}_{m,t}$. This is,

$$b_{m,t}(\mathbf{r}_{m,t}) = \frac{\partial E(r_{i,t}|r_{m,t}, r_{1:T})}{\partial r_{m,t}}\Big|_{r_{m,t}=\mathbf{r}_{m,t}}. \tag{6.10}$$

Full details on this derivative and estimate are provided in the appendix.

7 Data

We use the value-weighted index constructed by the Center of Research in Security Prices (CRSP) as a proxy for market returns. Daily market excess returns as well as four individual stock excess returns for IBM, General Electric or GE, Exxon or XOM, and Amgen or AMGN are obtained from 2000/01/03 to 2013/12/31 (3521 daily observations). Excess returns are derived after subtracting the risk-free return approximated by the three-month Treasury bill rate. All returns are scaled by 100. Figure 1 displays the data and Table 1 reports summary statistics. All individual stocks display skewness and excess kurtosis. Figure 1 shows that returns with absolute large (small) value tend to be followed by other large (small) absolute returns reflecting volatility clustering.

8 Model Performance

To compare the MGARCH-DPM model with the parametric MGARCH-t model we compute each model's predictive likelihood. The predictive likelihood for $r_{L:T}$, $L < T$ is expressed in terms of the one-step-ahead predictive likelihoods,

$$m(r_{L:T}|r_{1:L-1},\mathcal{M}) = \Pi_{t=L}^{T} p(r_t|r_{1:t-1},\mathcal{M}) \tag{8.1}$$

where $\mathcal{M}$ denotes the particular model (MGARCH-DPM or MGARCH-t), and $L > 1$ is chosen to eliminate the influence of the priors on model comparison. We can approximate the one-step-ahead predictive likelihoods, $p(r_t|r_{1:t-1},\mathcal{M})$, by averaging the data density over draws of the unknown parameters conditional on the data history $r_{1:t-1}$. This integrates out parameter and distributional uncertainty as

$$\begin{aligned} p(r_t|r_{1:t-1},\mathcal{M}) &= \int p(r_t|\theta, r_{1:t-1},\mathcal{M})p(\theta|r_{1:t-1},\mathcal{M})d\theta \qquad (8.2) \\ &\approx \frac{1}{M}\sum_{g=1}^{M} p(r_t|\theta^{(g)}, r_{1:t-1},\mathcal{M}) \end{aligned}$$

where $\theta^{(g)}$ is a posterior draw from $p(\theta|r_{1:t-1},\mathcal{M})$ and $p(r_t|\theta^{(g)}, r_{1:t-1},\mathcal{M})$ is the data density given $\theta^{(g)}$ and $r_{1:t-1}$ for model $\mathcal{M}$. Note that we are able to compute $H_t^{(g)}$ at each iteration of the MCMC since we have $H_{t-1}^{(g)}$ and GARCH parameters: $H_t^{(g)} = \Gamma_0^{(g)} + \Gamma_1^{(g)} \odot (r_{t-1} - \eta^{(g)})(r - \eta^{(g)})'_{t-1} + \Gamma_2^{(g)} \odot H_{t-1}^{(g)}$.

Based on (8.2), the predictive likelihoods for the two models are estimated as

$$p(r_t|r_{1:t-1}, \text{MGARCH-t}) \approx \frac{1}{M}\sum_{g=1}^{M} t(r_t|\mu^{(g)}, H_t^{(g)}, \nu^{(g)}), \tag{8.3}$$

$$p(r_t|r_{1:t-1}, \text{MGARCH-DPM}) \approx \frac{1}{M}\sum_{g=1}^{M} N(r_t|\mu^{(g)}_{s_t^{(g)}}, H_t^{(g)^{1/2}} B^{(g)}_{s_t^{(g)}} H_t^{(g)^{1/2'}}). \tag{8.4}$$

In MGARCH-DPM model, at each iteration g, $s_t^{(g)}$ is drawn from one of the $K^{(g)}+1$ components with weights $\omega_j^{(g)}$ $j = 1,\ldots,K^{(g)}$ and $1 - \sum_{j=1}^{K^{(g)}} \omega_j^{(g)}$. When $s_t^{(g)} = K^{(g)}+1$ a new parameter $\theta \sim G_0$ is drawn.

The following priors are used in estimation. In the MGARCH-t model, $\nu \sim \mathcal{U}(2,100)$, and $\mu \sim N(0, 0.1I)$ for both models. For each of GARCH parameters in both models, we set $\Gamma_{0,ij}^{1/2} \sim N(0,100)\mathbf{1}_S$, $\gamma_{1,i} \sim N(0,100)\mathbf{1}_S$ and $\gamma_{1,i} \sim N(0,100)\mathbf{1}_S$, $i = 1,\ldots,q+1$, $j \leq i$ as prior distribution where S denotes the following restriction: $\text{diag}(\Gamma_0^{1/2}) > 0, \gamma_{11} > 0, \gamma_{22} > 0$ to impose identification. For the concentration parameter $\alpha \sim \mathcal{G}(2,8)$. The prior on α controls the number of the distinct components in the mixture model, although with a large number of observations the effect of the prior is diminished. For the hyper-parameters of the base measure G_0, we set $B_0 = (\nu_0 - q - 1)I$ which makes $E(B) = I$ and centers the conditional covariance of r_t at H_t. $\nu_0 = 8$, but other values for ν_0 do not change our conclusions.

Table 2 reports the log-predictive likelihoods for the MGARCH-t and MGARCH-DPM models, and the log Bayes factor, for the last 500 ($L = 3021$) observations, from 2012/01/05 to 2013/12/31. Bivariate models based on daily excess returns on IBM, GE, XOM and AMGN each with excess market returns are considered. The results strongly support our semi-parametric model relative to the benchmark model. For instance, log-Bayes factors are all greater than 370. This is very strong evidence of significant deviations from the Student-t MGARCH model.

Figure 2 displays the time-series of the market and IBM excess returns as well as the difference in the log-predictive likelihood of the two models using

$$\log p(r_t|r_{1:t-1}, \text{MGARCH-DPM}) - \log p(r_t|r_{1:t-1}, \text{MGARCH-t}). \quad (8.5)$$

Positive values favour the MGARCH-DPM specification. This figure shows that the MGARCH-DPM model almost always outperforms MGARCH-t model. There are large differences when the market or IBM returns are extreme.

9 Semiparametric Conditional Beta

This section presents empirical estimates of the nonparametric dynamic conditional beta from the MGARCH-DPM model for several individual stocks and compares them with the corresponding counterpart from the parametric MGARCH-t model. Not only does the beta computed in this way change over time, but also the time-varying conditional beta is sensitive to the contemporaneous value of excess market return. This implies that the value of the systematic risk of an asset at each time depends on the level of the market return.

The model is applied to derive a nonparametric conditional beta (calculated in Section 6) using excess returns on a single stock and on the market return ($q = 1$). This results in a conditional expected return of the individual stock comparable to the conditional CAPM model. Later, additional factors are explored.

The analysis reported here is based on 25000 iterations of the MCMC algorithm. The first 15000 draws were dropped as burn-in and the following 10000 used for inference. The average acceptance rate of GARCH parameters is about 20% and about 30% for parametric and nonparametric models, respectively.

Tables 3-6 report the posterior mean and the 0.95 probability density intervals of the fixed parameters for both models and for different stocks. The estimated MGARCH parameters from the two models are consistent. The tables report c, the number of components in the mixture used to estimate the unknown density. On average, the joint density of IBM, XOM, GE with the market is estimated using about 3.6-5.6 components but the density intervals indicate considerable uncertainty. However, for AMGN and the market, about 15 components are used, showing that this joint density is far more complex than the others. These results are compatible with the small degree of freedom estimated in the benchmark models. Estimates of η_1 and η_2 are consistently positive indicating a larger response to the conditional covariance from negative shocks.

Figures 3 - 6 compares the posterior mean of the *realized* beta over time derived from both models for each of the stocks. For MGARCH-t model, the posterior mean of (2.8) is reported while for the MGARCH-DPM model the posterior mean of (6.10) is evaluated at the realized excess market return value for time t. As seen in the figures, both models result in very similar time series for the conditional beta.

Figure 7 - 10 illustrates posterior mean of each stock's conditional beta as a function of the contemporaneous market excess return using (6.10) at several dates. These figures show that beta is changing over time and, more importantly, at each time the value of beta is sensitive to the contemporaneous value of the market excess return. For each stock there are dates that beta is a constant function of the market return which would be consistent with the MGARCH-t model. However, each stock has dates in which beta is nonlinearly dependent on the market return. Moreover, often beta is asymmetrically

related to the market; when the market excess return increases (large positive values), conditional beta drops more significantly (Figure 7 - 9).

The nonlinear relationship between beta and the market transfers directly into the conditional expected excess return. For example, Figure 11 displays the posterior mean of the conditional expected excess return of IBM given different values of the contemporaneous market excess return, derived from (6.9), for dates for which the conditional betas are illustrated in Figure 7. This figure clearly shows how the nonlinear conditional beta results in the nonlinear conditional expected return.

To investigate the significance of this nonlinear relationship Figures 12 - 15 display the posterior mean of the nonparametric conditional beta as a function of the market excess return as well as the 0.90 density intervals for selected dates for each stock. Beta derived from the MGARCH-t model is included and is a constant function at each time. It is clear from these figures that there are significant departures in beta from the constant beta from the MGARCH-t model.

Finally, Figures 16 - 19 provide a three dimensional version of Figures 7-10 for each firm. In some periods beta is essentially flat and consistent with the MGARCH-t model while in other times beta is very sensitive to the market return.

9.1 Summary of Empirical Results

As the empirical results illustrate, the conditional beta is time-varying and at each time depends on the contemporaneous market excess return, as opposed to the constant beta of the benchmark model.

The previous results show some periods in which the conditional beta is insensitive to the value of $r_{m,t}$ (beta is almost constant with respect to $r_{m,t}$) while in other time periods beta changes significantly with $r_{m,t}$. To measure the sensitivity of $b_{m,t}(\mathbf{r}_{m,t})$ to $\mathbf{r}_{m,t}$ at each time t consider the following measure

$$d_t = \max_{\mathbf{r}_{m,t}} b_{m,t}(\mathbf{r}_{m,t}) - \min_{\mathbf{r}_{m,t}} b_{m,t}(\mathbf{r}_{m,t}), \tag{9.1}$$

where $b_{m,t}(\mathbf{r}_{m,t})$ is defined in (6.10). Large values of d_t indicate that $b_t(\mathbf{r}_{m,t})$ is strongly sensitive to $\mathbf{r}_{m,t}$, while a $d_t = 0$ indicates no sensitivity. The MGARCH-t model has a $d_t = 0$ for all t. Figure 20 illustrates this d_t over time for all individual stocks. Among these four stocks, the dynamic conditional beta for IBM and XOM have the most and the least sensitivity to $\mathbf{r}_{m,t}$, respectively. What is apparent is that during relatively high volatility periods such as 2002-03, 2009 and 2011:6-2012, d_t attains its smallest values over the sample. In these periods shocks to the market are expected to be large. During lower volatility periods large shocks to the market and firms are unexpected and the conditional beta adjusts accordingly.

To investigate how $b_{m,t}(\mathbf{r}_{m,t})$ changes with different market conditions Figures 21-24 show the broad trends that we find in all stocks. When the market is highly volatile, an individual stock's conditional beta is less affected by unexpected shocks in the contemporaneous market return. While in a calm market, the conditional beta changes remarkably from unexpected shocks to the market. However, the changes depend on the stocks correlation with the market.

When the market is calm, an unexpected shock increases the conditional beta for a stock that is highly correlated with the market, while this effect is completely the reverse for stocks with low correlation with the market. In other words, when an asset is highly correlated with the market, a large move in a stable market increases the conditional covariance between the market and the asset more than it increases the conditional variance of the market, resulting in a significant increase in the conditional beta. When an asset is less correlated with the market, a large move in a stable market increases the conditional variance of the market more than it increases the conditional covariance between the market and the asset, leading to a drop in the conditional beta.

It is often the case that the effect on $b_{m,t}(\mathbf{r}_{m,t})$ from $\mathbf{r}_{m,t}$ is asymmetric. Frequently $b_{m,t}(\mathbf{r}_{m,t})$ is more sensitive to large positive values of $\mathbf{r}_{m,t}$ compared to negative values. In addition, when the market is calm, we see both u-shape and inverse u-shape patterns for the conditional beta of all stocks.

10 Extension to Fama-French three-factor model

Fama & French (1993) assert that the common variation in stock returns is largely captured by three portfolios consisting of the market portfolio, a mimicking portfolio for size ($r_{SMB,t}$), and mimicking portfolio for book-to-market value ($r_{HML,t}$). Applying MGARCH-DPM model in four dimensions ($q = 3$), we estimate a dynamic nonparametric version of the static Fama-French three-factor model. As in the previous bivariate application, there will be a dynamic nonparametric beta on each factor conditional on a specified factor value. That is, we estimate the joint model and back out the conditional distribution of the firm return $r_{i,t}$ and the associated betas given values of $\mathbf{r}_{m,t}, \mathbf{r}_{SMB,t}$ and $\mathbf{r}_{HML,t}$. Daily data for $r_{SMB,t}$ and $r_{HML,t}$ are obtained from Kenneth French's website, from Jan 3, 2000 to Dec 31, 2013 (3521 observations). In this case, the number of GARCH parameters increases to 26; $(\frac{(q+1)(q+2)}{2} + 2(q+1) + 2(q+1))$. This model, FF-MGARCH-DPM, replaces (4.1) with

$$r_t \equiv (r_{IBM,t}, r_{m,t}, r_{SMB,t}, r_{HML,t})|\phi_t, H_t \sim N(\xi_t, H_t^{1/2}\Lambda_t(H_t^{1/2})'),\ t = 1, ..., T. \quad (10.1)$$

Applied to IBM data Table 7 reports the posterior mean of the number of distinct clusters is about 3.84 and $\alpha \approx 0.35$. This is a reduction for both parameters compared to the previous one factor model.

The nonparametric conditional beta in FF-MGARCH-DPM model is a vector. It is defined analogously to (6.10) as the partial derivative with respect to the factor. For instance, beta for size factor is defined as

$$b_{SMB,t} = \frac{\partial E(r_{IBM,t}|r_{m,t}, r_{SMB,t}, r_{t,HML}, r_{1:T})}{\partial r_{SMB,t}}\Big|_{\substack{r_{m,t}=\mathbf{r}_{m,t} \\ r_{SMB,t}=\mathbf{r}_{SMB,t} \\ r_{HML,t}=\mathbf{r}_{HML,t}}} \quad (10.2)$$

with a similar expression for the other factor coefficients $b_{m,t}$ and $b_{HML,t}$. Note that each beta is a potentially nonlinear function of $\mathbf{r}_{m,t}, \mathbf{r}_{SMB,t}, \mathbf{r}_{HML,t}$. Figure 25 illustrates the time series patterns of the posterior mean of the realized beta coefficients over time. This is $b_{m,t}$, $b_{HML,t}$ and $b_{SMB,t}$ evaluated at the realized values of $r_{m,t}, r_{SMB,t}$ and $r_{HML,t}$ for our dataset. This figure, consistent with Engle's (2015) results, shows that the beta on $r_{HML,t}$ is the most variable. The market beta from the previous one-factor model for IBM (Section 9) is included in this figure. The pattern of two series is very similar in both models, which supports the one-factor specification.

Figure 26 shows the posterior mean of $b_{m,t}$ as a function of the market excess return at several time periods. To produce the plot, the other two factors are set to their sample mean value. As we see, even after accounting for size and value factors in the model, the conditional beta (market coefficient) at each time is sensitive to the value of the contemporaneous market excess returns.

The three-dimensional Figure 27 illustrates the posterior mean of $b_{m,t}$, $b_{SMB,t}$, and $b_{HML,t}$ as functions of excess market return over time. It is evident that besides market coefficient, coefficients on $r_{SMB,t}$, and $r_{HML,t}$ are nonlinearly dependent on the value of the contemporaneous market excess return.

In summary, our results show that the conditional beta at each time does depend on the contemporaneous value of the market excess return. The market beta from the 3-factor model are consistent with the findings from the previous one-factor model. The empirical results show that besides the market beta, coefficients on the other two factors are also nonlinearly dependent on the value of the contemporaneous market excess returns.

11 Conclusion

This paper derives a dynamic conditional beta representation using a Bayesian semiparametric multivariate GARCH model. We show that predictive Bayes factors strongly support this semiparametric model over a multivariate GARCH with Student-t innovations. Empirically we find the time-varying beta from our model nonlinearly depends on the contemporaneous value of excess market return. In highly volatile markets, beta is almost constant, while in stable markets, the beta coefficient can depend asymmetrically on the contemporaneous value of the market excess return. The results are robust to an extension of Fama-French factors and reveal nonlinear dependence in all beta coefficients of the factors.

12 Appendix

12.1 Distributions

If $\mathbf{r} \sim t(\mu, \Sigma, \nu)$ then the density function of the Student-t (Bauwens et al. 2000) is

$$f(\mathbf{r}|\nu, \mu, \Sigma) = \frac{\Gamma(\frac{\nu+p}{2})}{\Gamma(\frac{\nu}{2})\pi^{p/2}}|\Sigma|^{-1/2}\left[1+\frac{1}{\nu}(\mathbf{r}-\mu)^T\Sigma^{-1}(\mathbf{r}-\mu)\right]^{-(\nu+p)/2}, \nu > 0.$$

The $q \times q$ matrix B follows an inverse Wishart density with a symmetric positive definite scale matrix B_0 and degree of freedom $\nu_0 \geq q+1$, if its pdf can be written as

$$f(B|B_0, \nu_0) = \frac{|B_0|^{\nu_0/2}}{2^{\frac{q\nu_0}{2}}\pi^{\frac{q(q-1)}{4}}\Pi_{i=1}^{q}\Gamma(\frac{\nu_0+1-i}{2})}|B|^{-\frac{\nu_0+q+1}{2}}\exp\left[-\frac{1}{2}tr(B^{-1}B_0)\right],$$

with $E(B) = \frac{1}{\nu_0-q-1}B_0$.

The pdf of the Gamma distribution $\mathcal{G}(a, b)$ with shape parameter a and scale parameter b is written as

$$f(x|a, b) = \frac{1}{\Gamma(a)b^a}x^{a-1}e^{-x/b}, \ x \in (0, \infty), \quad E(x) = ab.$$

12.2 Derivation of the nonparametric conditional beta

$$E(\mathbf{r}_{i,t}|\mathbf{r}_{i,t}, r_{1:T}, G^{(g)}) = \sum_{j=1}^{K^{(g)}} q_j^{(g)}(\mathbf{r}_{m,t})[\mu_{j,1}^{(g)} + \beta_{jt}^{(g)}(\mathbf{r}_{m,t} - \mu_{j,2}^{(g)})] + \tag{12.1}$$

$$\left(1 - \sum_{j=1}^{K^{(g)}} q_j^{(g)}(\mathbf{r}_{m,t})\right) \frac{\int[\mu_1 + \beta_t(\mathbf{r}_{m,t} - \mu_2)]N(\mathbf{r}_{m,t}|\mu_2, (H_t^{(g)^{1/2}}BH_t^{(g)^{1/2'}})_{22})p(\mu, B)d\mu dB}{\int N(\mathbf{r}_{m,t}|\mu_2, (H_t^{(g)^{1/2}}BH_t^{(g)^{1/2'}})_{22})p(\mu, B)d\mu dB}.$$

Let

$$A_1 = \int [\mu_1 + \beta_t(\mathbf{r}_{m,t} - \mu_2)]N(\mathbf{r}_{m,t}|\mu_2, (H_t^{(g)^{1/2}}BH_t^{(g)^{1/2'}})_{22})p(\mu, B)d\mu dB, \tag{12.2}$$

$$A_2 = \int N(\mathbf{r}_{m,t}|\mu_2, (H_t^{(g)^{1/2}}BH_t^{(g)^{1/2'}})_{22})p(\mu, B)d\mu dB. \tag{12.3}$$

A_1 and A_2 can be easily approximated by Monte Carlo simulation as follows

$$A_1 \approx \frac{1}{N}\sum_{n=1}^{N}[\mu_{n,1} + \beta_{n,t}^{(g)}(\mathbf{r}_{m,t} - \mu_{n,2})]N(\mathbf{r}_{m,t}|\mu_{n,2}, (H_t^{(g)^{1/2}}B_nH_t^{(g)^{1/2'}})_{22}) \tag{12.4}$$

$$A_2 \approx \frac{1}{N}\sum_{n=1}^{N} N(\mathbf{r}_{m,t}|\mu_{n,2}, (H_t^{(g)^{1/2}}B_nH_t^{(g)^{1/2'}})_{22}) \tag{12.5}$$

where μ_n and B_n, $n = 1, ..., N$ are i.i.d draws from the prior $p(\mu, B)$ which in our model is $N(\mu|\mu_0, D)$ and $\mathcal{W}^{-1}(B|B_0, \nu_0)$, and

$$\beta_{nt}^{(g)} = \frac{(H_t^{(g)^{1/2}} B_n H_t^{(g)^{1/2'}})_{12}}{(H_t^{(g)^{1/2}} B_n H_t^{(g)^{1/2'}})_{22}}. \tag{12.6}$$

Now we obtain the posterior mean of the nonparametric conditional beta by taking the derivative of 12.1:

$$b_{m,t}(\mathbf{r}_{m,t}) = \frac{1}{M}\sum_{g=1}^{M} b_{m,t}(\mathbf{r}_{m,t}, G^{(g)}) = \frac{1}{M}\sum_{g=1}^{M} \frac{\partial E(r_{i,t}|r_{m,t}, r_{1:T}, G^{(g)})}{\partial r_{m,t}}\Big|_{r_{m,t}=\mathbf{r}_{m,t}}. \tag{12.7}$$

After replacing A_1 and A_2 with their approximations we have

$$\begin{aligned}
\frac{\partial E(r_{i,t}|r_{m,t}, r_{1:T}, G^{(g)})}{\partial r_{m,t}} &\approx \sum_{j=1}^{K^{(g)}} q_j^{(g)}(r_t^m)\beta_{jt}^{(g)} \\
&+ \sum_{j=1}^{K^{(g)}} q_j'^{(g)}(r_t^m)[\mu_{j,1}^{(g)} + \beta_{jt}^{(g)}(r_t^m - \mu_{j,2}^{(g)})] \\
&- \sum_{j=1}^{K^{(g)}} q_j'^{(g)}(r_t^m)\frac{\sum_n[\mu_{n,1} + \beta_{tn}^{(g)}(r_t^m - \mu_{n,2})]N(r_t^m|\mu_{n,2}, (H_t^{(g)^{1/2}} B_n H_t^{(g)^{1/2'}})_{22})}{\sum_n N(r_t^m|\mu_{n,2}, (H_t^{(g)^{1/2}} B_n H_t^{(g)^{1/2'}})_{22})} \\
&+ \left(1 - \sum_{j=1}^{K^{(g)}} q_j^{(g)}(r_t^m)\right)\Big\{\frac{\sum_n \beta_{tn}^{(g)} N(r_t^m|\mu_{n,2}, (H_t^{(g)^{1/2}} B_n H_t^{(g)^{1/2'}})_{22})}{\sum_n N(r_t^m|\mu_{n,2}, (H_t^{(g)^{1/2}} B_n H_t^{(g)^{1/2'}})_{22})} \\
&+ \frac{\sum_n[\mu_{n,1} + \beta_{tn}^{(g)}(r_t^m - \mu_{n,2})]N'(r_t^m|\mu_{n,2}, (H_t^{(g)^{1/2}} B_n H_t^{(g)^{1/2'}})_{22})}{\sum_n N(r_t^m|\mu_{n,2}, (H_t^{(g)^{1/2}} B_n H_t^{(g)^{1/2'}})_{22})} \\
&- \frac{[\sum_n[\mu_{n,1} + \beta_{tn}^{(g)}(r_t^m - \mu_{n,2})]N(r_t^m|\mu_{n,2}, (H_t^{(g)^{1/2}} B_n H_t^{(g)^{1/2'}})_{22})]\sum_n N'(r_t^m|\mu_{n,2}, (H_t^{(g)^{1/2}} B_n H_t^{(g)^{1/2'}})_{22})}{[\sum_n N(r_t^m|\mu_{n,2}, (H_t^{(g)^{1/2}} B_n H_t^{(g)^{1/2'}})_{22})]^2}\Big\}
\end{aligned} \tag{12.8}$$

where $\beta_{jt}^{(g)}$, $\beta_{nt}^{(g)}$, and $q_j^{(g)}(r_t^m)$ are defined in Equations (6.8), (12.6), and (6.6), respectively, and $N'(x|.)$ is the derivative of the pdf of Normal distribution with respect to x. In the 3-factor model the derivations follow similarly but the derivative will be a vector of size 3, each element of which is the coefficient of market excess return, size factor, or value factor.

References

Bauwens, L., Lubrano, M. & Richard, J.-F. (2000), *Bayesian Inference in Dynamic Econometric Models*, Oxford University Press, Oxford.

Bollerslev, T., Engle, R. F. & Wooldridge, J. M. (1988), 'A Capital Asset Pricing Model with Time-varying Covariances', *The Journal of Political Economy.* **96**(1), 116–131.

Choudhry, T. (2002), 'The Stochastic Structure of the Time-Varying Beta: Evidence from UK Companies', *The Manchester School* **70**(6), 768–791.

Ding, Z. & Engle, R. F. (2001), 'Large Scale Conditional Covariance Matrix Modeling, Estimation and Testing', *Academia Economic Papers* **29**(2), 157–184.

Engle, R. F. (2015), Dynamic conditional beta. SSRN: http://ssrn.com/abstract=2404020.

Engle, R. F. & Ng, V. K. (1993), 'Measuring and Testing the Impact of News on Volatility', *The Journal of Finance* **48**(5), 1749–1778.

Escobar, M. D. & West, M. (1995), 'Bayesian Density Estimation and Inference using Mixtures', *Journal of the American Statistical Association* **90**(430), 577–588.

Fama, E. F. & French, K. R. (1993), 'Common Risk Factors in the Returns on Stocks and Bonds', *Journal of Financial Economics* **33**(1), 3–56.

Ferguson, T. S. (1973), 'A Bayesian Analysis of Some Nonparametric Problems', *The Annals of Statistics* **1**, 209–230.

Giannopoulos, K. (1995), 'Estimating the Time-varying Components of International Stock Markets Risk', *European Journal of Finance* **1**(2), 129–164.

Jensen, M. J. & Maheu, J. M. (2013), 'Bayesian Semiparametric Multivariate GARCH Modeling', *Journal of Econometrics* **176**(1), 3–17.

Kalli, M., Griffin, J. E. & Walker, S. G. (2011), 'Slice Sampling Mixture Models', *Statistics and Computing* **21**(1), 93–105.

Lintner, J. (1965), 'The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets', *The Review of Economics and Statistics* **47**(1), 13–37.

Lo, A. Y. (1984), 'On a Class of Bayesian Nonparametric Estimates: I. Density Estimates.', *The Annals of Statistics* **12**(1), 351–357.

McCurdy, T. H. & Morgan, I. (1992), 'Evidence of Risk Premiums in Foreign Currency Futures Markets', *Review of Financial Studies* **5**(1), 65–83.

Roth, M. (2013), On the Multivariate t Distribution. Technical report from Automatic Control at Linkopings universitet, Report no.: LiTH-ISY-R-3059.

Sethuraman, J. (1994), 'A Constructive Definition of Dirichlet Priors.', *Statistica Sinica* **4**, 639–650.

Sharpe, W. F. (1964), 'Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk', *The Journal of Finance* **19**(3), 425–431.

Walker, S. G. (2007), 'Sampling the Dirichlet Mixture Model with Slices', *Communications in Statistics– Simulation and Computation* **36**(1), 46–57.

Stock	Mean	Variance	Skewness	Kurtosis	Max	Min
Market	0.017	1.744	-0.070	7.067	11.350	-8.950
IBM	0.028	3.070	0.230	7.834	13.019	-15.567
GE	-0.003	4.277	0.323	8.397	19.702	-12.797
XOM	0.032	2.672	0.367	11.163	17.180	-13.950
AMGN	0.034	4.758	0.508	5.907	15.090	13.437

Table 1: Summary statistics of the daily excess returns on the market portfolio, IBM, GE, XOM and AMGN, from 2000/01/03 to 2013/12/31 (3521 observations).

	log-predictive likelihood			
Model	IBM	GE	XOM	AMGN
MGARCH-DPM	−983.27	−964.99	−875.473	−1140.12
MGARCH-t	−1353.67	−1369.03	−1300.21	−1571.32
log Bayes factor	370.39	404.031	424.73	431.19

Table 2: This table reports the log-predictive likelihood for the bivariate MGARCH-t and MGARCH-DPM models and the log-Bayes factor, for the last 500 observations, from 2012/03/12 to 2013/12/31. Bivariate data are daily excess market returns coupled with excess returns on IBM, GE, XOM and AMGN from 2000/01/03 to 2013/12/31.

IBM	MGARCH-DPM		MGARCH-t	
Parameter	Post. Mean	95% DI	Post. Mean	95% DI
γ_{01}	0.102	$(0.055, 0.146)$	0.023	$(0.015, 0.037)$
γ_{02}	-0.043	$(-0.081, 0.003)$	-0.042	$(-0.053, -0.034)$
γ_{03}	0.020	$(0.001, 0.053)$	0.020	$(0.002, 0.048)$
γ_{11}	0.247	$(0.199, 0.307)$	0.150	$(0.144, 0.160)$
γ_{12}	0.267	$(0.232, 0.313)$	0.224	$(0.210, 0.233)$
γ_{21}	0.971	$(0.965, 0.977)$	0.975	$(0.971, 0.977)$
γ_{22}	0.953	$(0.945, 0.961)$	0.955	$(0.951, 0.961)$
μ_1			0.025	$(0.016, 0.046)$
μ_2			0.041	$(0.022, 0.074)$
ν			5.37	$(5.01, 5.54)$
c	5.6	$(3.00, 11.0)$		
α	0.571	$(0.070, 1.61)$		
η_1	0.570	$(0.349, 0.714)$	0.807	$(0.776, 0.864)$
η_2	0.533	$(0.434, 0.618)$	0.507	$(0.451, 0.644)$

Table 3: IBM Estimates: This table displays posterior mean and 95% density intervals (DI) for the parameters of MGARCH-DPM and MGARCH-t models. Data is daily excess returns on IBM and excess market returns. Data is from Jan 3, 2000 to Dec 31, 2013 (3521 observations).

XOM	MGARCH-DPM		MGARCH-t	
Parameter	Post. Mean	95% DI	Post. Mean	95% DI
γ_{01}	0.141	$(0.108, 0.182)$	0.110	$(0.012, 0.200)$
γ_{02}	0.014	$(-.003, 0.030)$	0.016	$(-0.058, 0.073)$
γ_{03}	0.014	$(0.001, 0.041)$	0.032	$(0.001, 0.082)$
γ_{11}	0.250	$(0.223, 0.283)$	0.228	$(0.165, 0.310)$
γ_{12}	0.238	$(0.198, 0.287)$	0.228	$(0.175, 0.288)$
γ_{21}	0.956	$(0.947, 0.965)$	0.958	$(0.935, 0.977)$
γ_{22}	0.960	$(0.953, 0.969)$	0.958	$(0.939, 0.974)$
μ_1			0.025	$(-0.076, 0.129)$
μ_2			0.022	$(-0.050, 0.092)$
ν			9.89	$(6.16, 13.90)$
c	3.6	$(2.00, 9.00)$		
α	0.324	$(0.011, 1.15)$		
η_1	0.480	$(0.345, 0.591)$	0.436	$(-0.051, 0.775)$
η_2	0.524	$(0.436, 0.613)$	0.514	$(0.279, 0.708)$

Table 4: XOM Estimates: This table displays posterior mean and 95% density intervals (DI) for the parameters of MGARCH-DPM and MGARCH-t models. Data is daily excess returns on XOM and excess market returns. Data is from Jan 3, 2000 to Dec 31, 2013 (3521 observations).

GE	MGARCH-DPM		MGARCH-t	
Parameter	Post. Mean	95% DI	Post. Mean	95% DI
γ_{01}	0.061	$(0.023, 0.093)$	0.031	$(0.012, 0.056)$
γ_{02}	-0.033	$(-0.054, -0.014)$	-0.029	$(-0.039, -0.008)$
γ_{03}	0.018	$(0.001, 0.052)$	0.036	$(0.022, 0.052)$
γ_{11}	0.196	$(0.174, 0.216)$	0.170	$(0.145, 0.188)$
γ_{12}	0.204	$(0.181, 0.225)$	0.180	$(0.168, 0.192)$
γ_{21}	0.974	$(0.967, 0.981)$	0.974	$(0.970, 0.981)$
γ_{22}	0.964	$(0.957, 0.970)$	0.971	$(0.967, 0.974)$
μ_1			0.004	$(-0.034, 0.029)$
μ_2			0.049	$(0.015, 0.071)$
ν			6.47	$(5.35, 7.05)$
c	5.04	$(3.00, 10.0)$		
α	0.501	$(0.060, 1.42)$		
η_1	0.554	$(0.414, 0.707)$	0.633	$(0.555, 0.785)$
η_2	0.464	$(0.395, 0.539)$	0.463	$(0.416, 0.561)$

Table 5: GE Estimates: This table displays posterior mean and 95% density intervals (DI) for the parameters of MGARCH-DPM and MGARCH-t models. Data is daily excess returns on GE and excess market returns. Data is from Jan 3, 2000 to Dec 31, 2013 (3521 observations).

AMGN	MGARCH-DPM		MGARCH-t	
Parameter	Post. Mean	95% DI	Post. Mean	95% DI
γ_{01}	0.137	$(0.089, 0.171)$	0.084	$(0.065, 0.106)$
γ_{02}	-0.011	$(-0.031, 0.012)$	-0.028	$(-0.044, -0.007)$
γ_{03}	0.016	$(0.001, 0.039)$	0.034	$(0.015, 0.059)$
γ_{11}	0.211	$(0.182, 0.239)$	0.165	$(0.156, 0.175)$
γ_{12}	0.188	$(0.172, 0.211)$	0.228	$(0.195, 0.242)$
γ_{21}	0.965	$(0.945, 0.958)$	0.973	$(0.971, 0.976)$
γ_{22}	0.951	$(0.945, 0.958)$	0.956	$(0.950, 0.965)$
μ_1			0.002	$(-0.014, 0.035)$
μ_2			0.038	$(0.024, 0.070)$
ν			5.81	$(5.56, 6.08)$
c	15	$(7.00, 28.0)$		
α	2.41	$(0.500, 5.21)$		
η_1	0.508	$(0.428, 0.596)$	0.768	$(0.686, 0.876)$
η_2	0.542	$(0.459, 0.630)$	0.479	$(0.443, 0.566)$

Table 6: AMGN Estimates: This table displays posterior mean and 95% density intervals (DI) for the parameters of MGARCH-DPM and MGARCH-t models. Data is daily excess returns on AMGN and excess market returns. Data is from Jan 3, 2000 to Dec 31, 2013 (3521 observations).

# of GARCH parameters	Posterior mean of α	Posterior mean of c
26	0.35	3.84

Table 7: Posterior mean of the precision parameter and the number of distinct components for the FF-MGARCH-DPM model. Data is daily excess returns for IBM, the market, SMB, and HML from 2000/01/03 to 2013/12/31 (3521 observations).

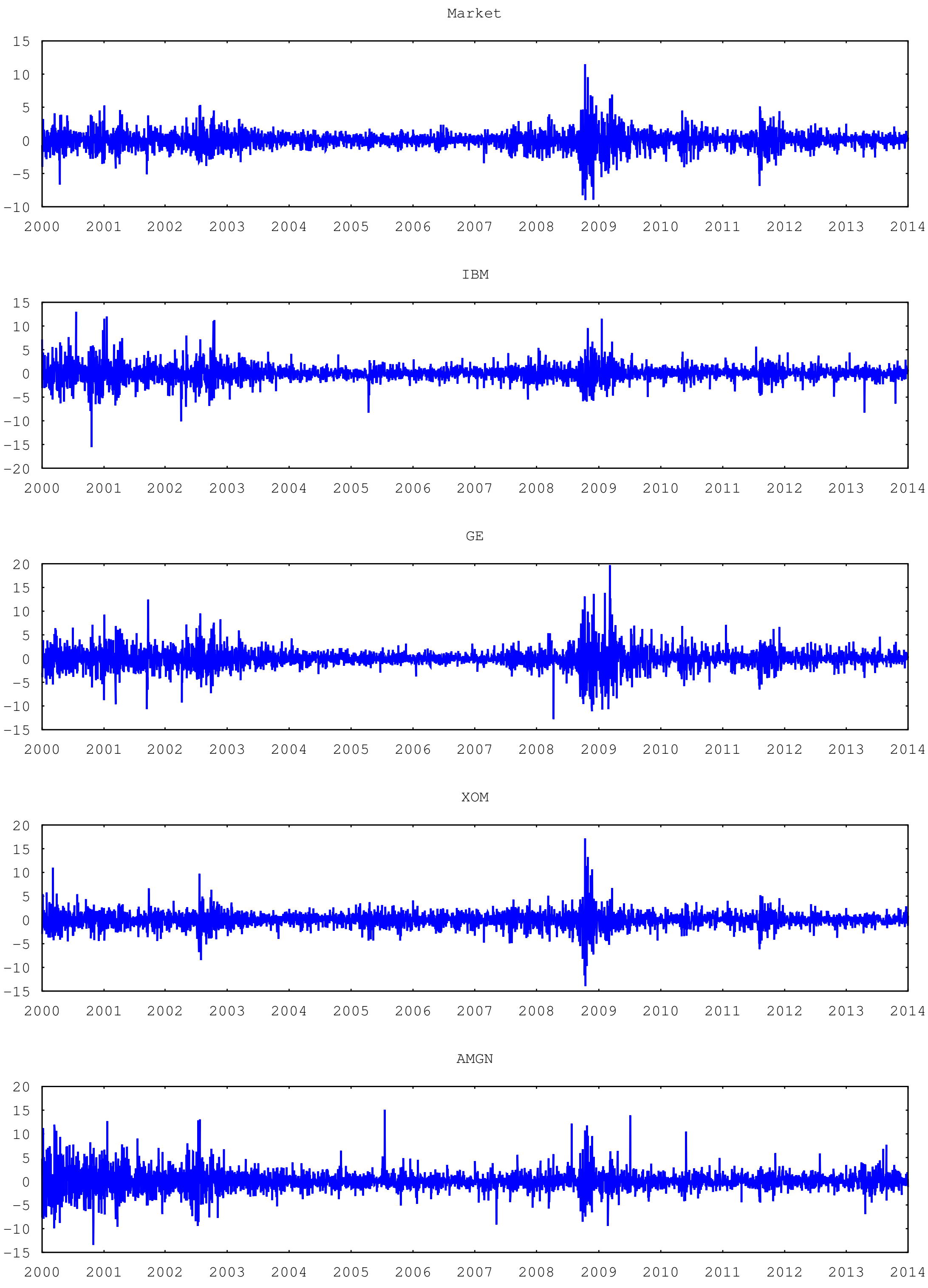

Figure 1: Daily excess returns on the market, IBM, GE, XOM and AMGN.

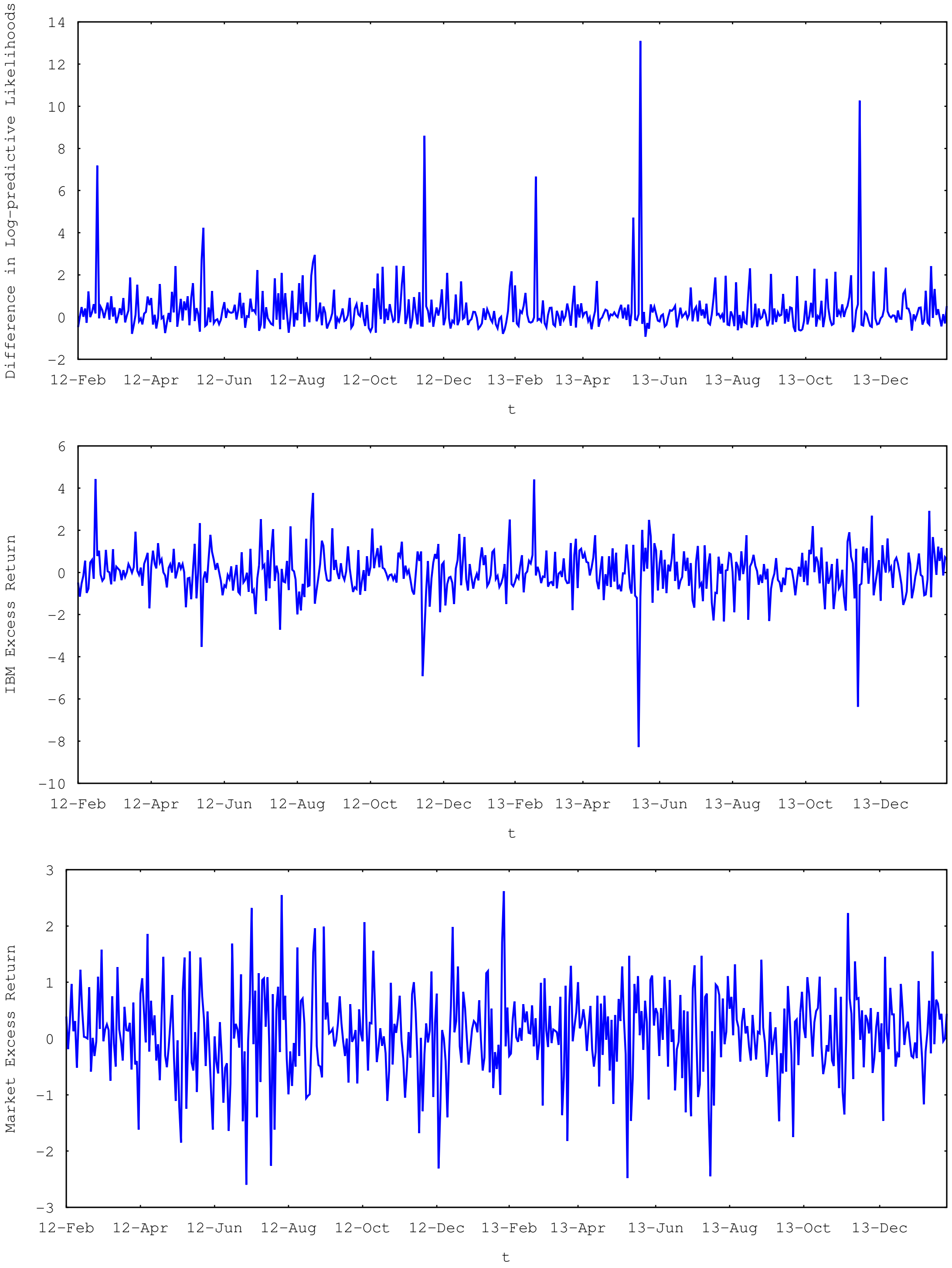

Figure 2: The first panel indicates the difference of log predictive likelihood of the two models corresponding to each of the last 500 observations, from 2012/01/05 to 2013/12/31, for MGARCH-t and MGARCH-DPM. The second and third panel illustrate the time series returns on IBM and the market.

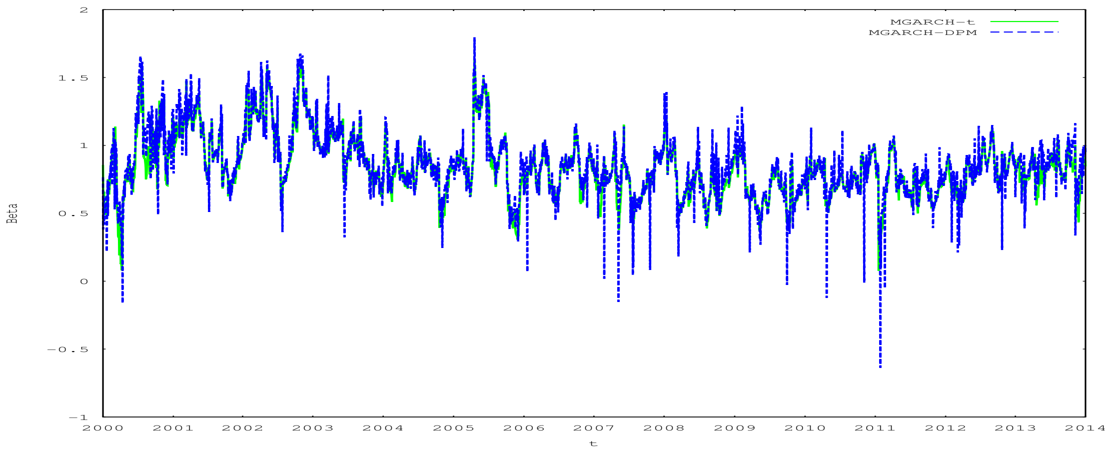

Figure 3: IBM: Realized conditional beta over time from MGARCH-t and MGARCH-DPM models.

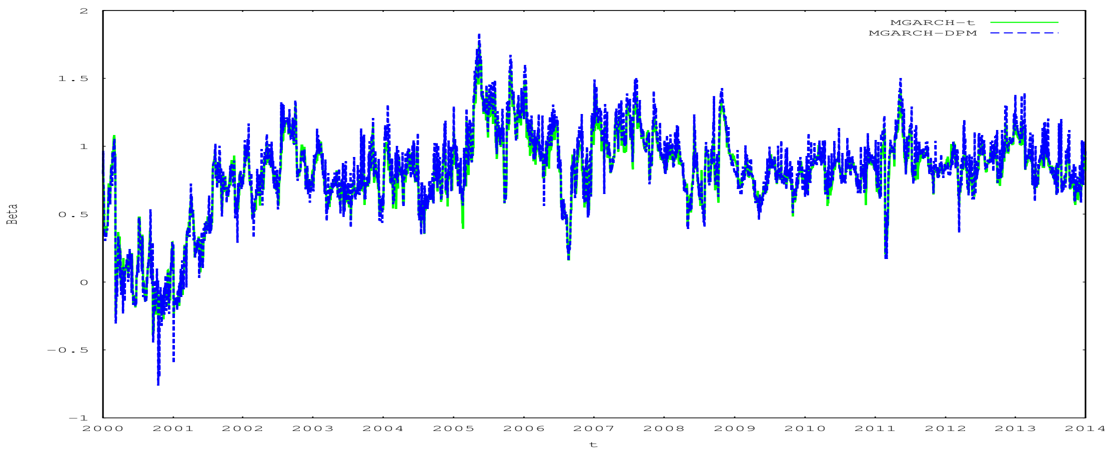

Figure 4: XOM: Realized conditional beta over time from MGARCH-t and MGARCH-DPM models.

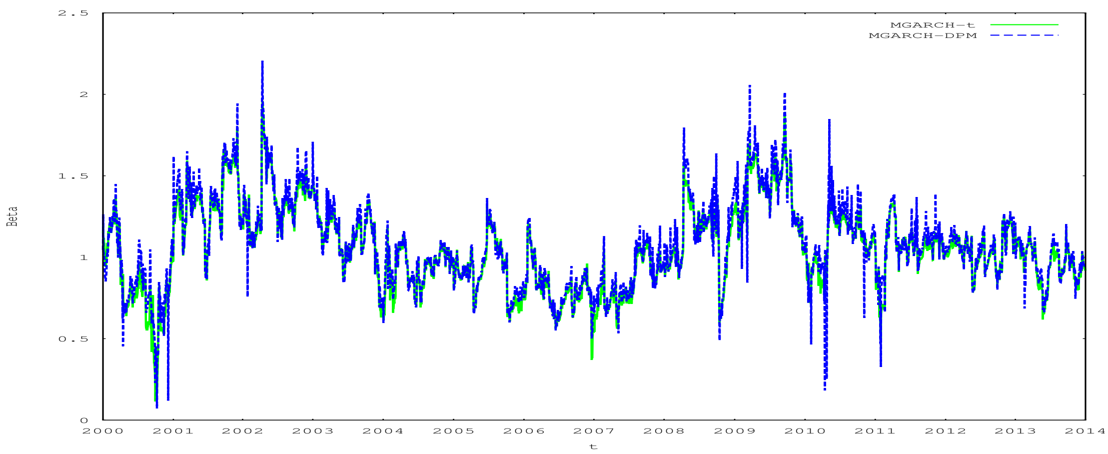

Figure 5: GE: Realized conditional beta over time from MGARCH-t and MGARCH-DPM models.

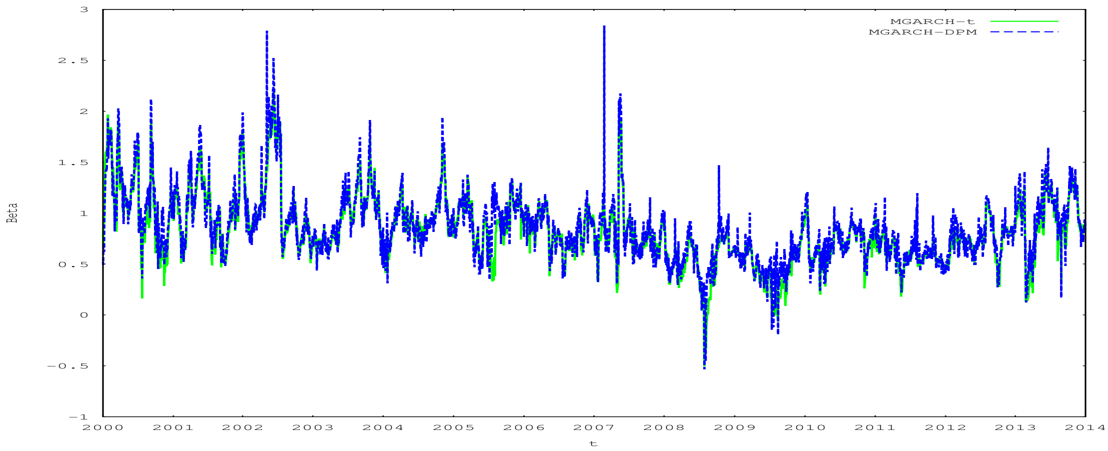

Figure 6: AMGN: Realized conditional beta over time from MGARCH-t and MGARCH-DPM models.

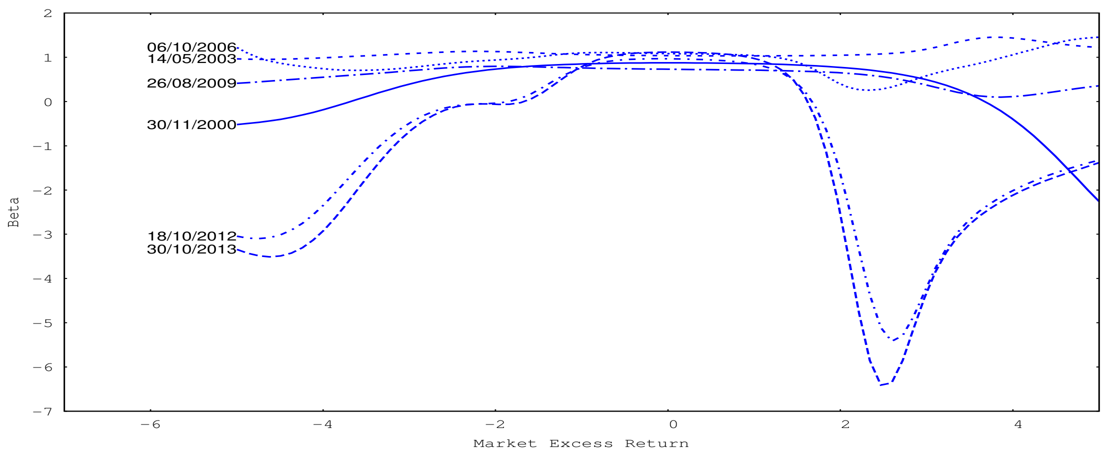

Figure 7: IBM: posterior mean of conditional beta as a function of the market excess return for different dates.

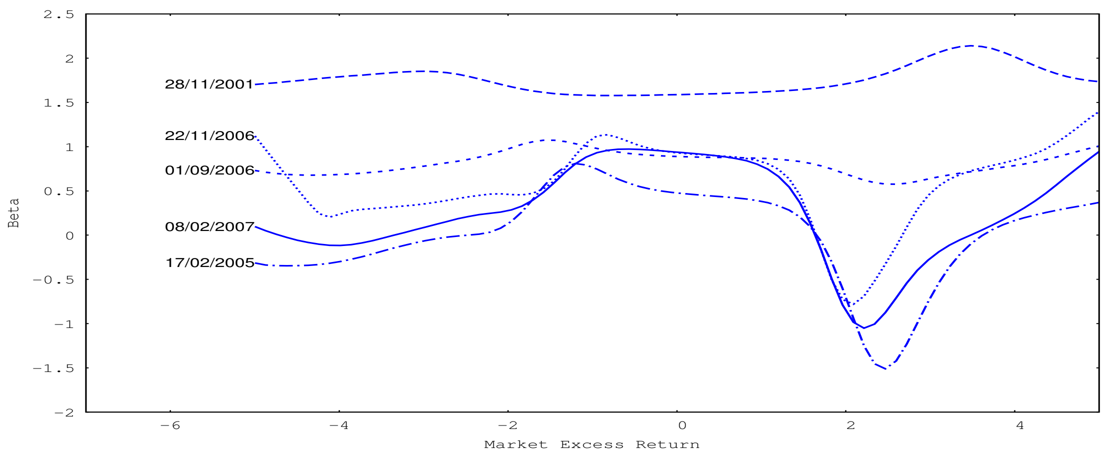

Figure 8: XOM: posterior mean of conditional beta as a function of the market excess return for different dates.

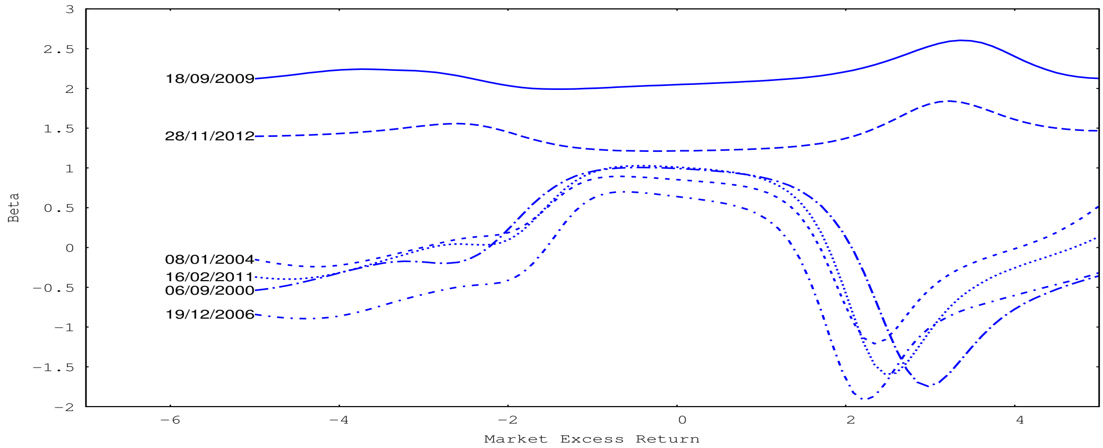

Figure 9: GE: posterior mean of conditional beta as a function of the market excess return for different dates.

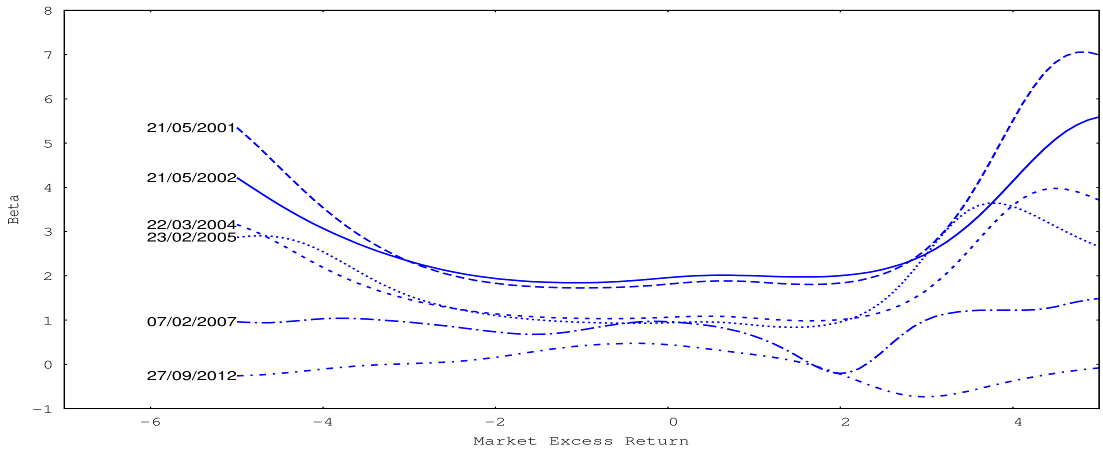

Figure 10: AMGN: posterior mean of conditional beta as a function of the market excess return for different dates.

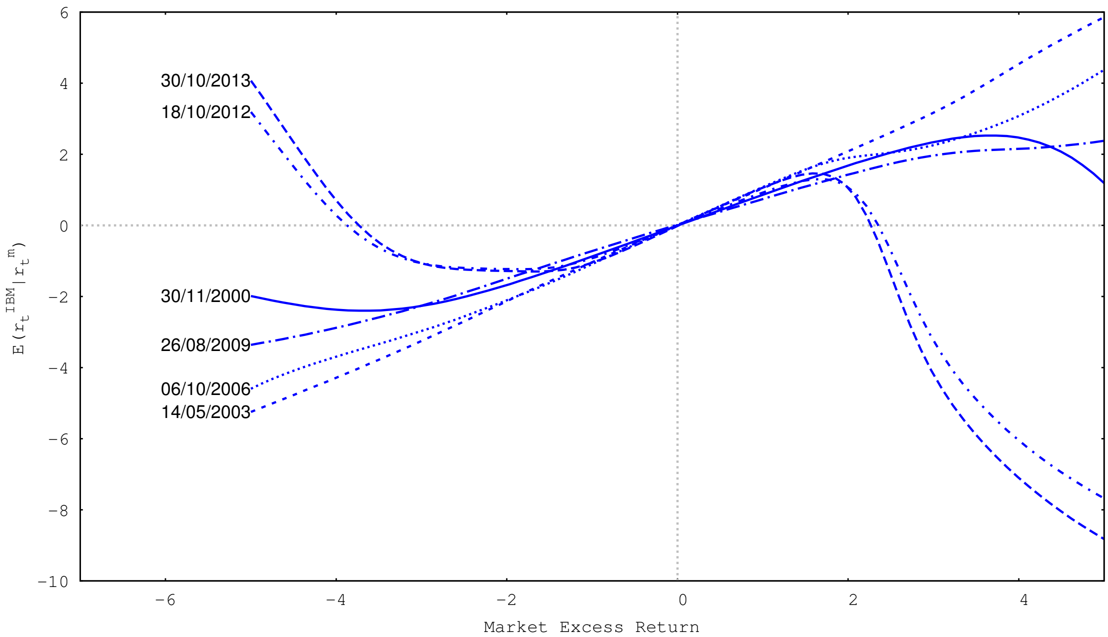

Figure 11: IBM: posterior mean of the conditional expected excess return of IBM given different values of the contemporaneous market excess return for different dates.

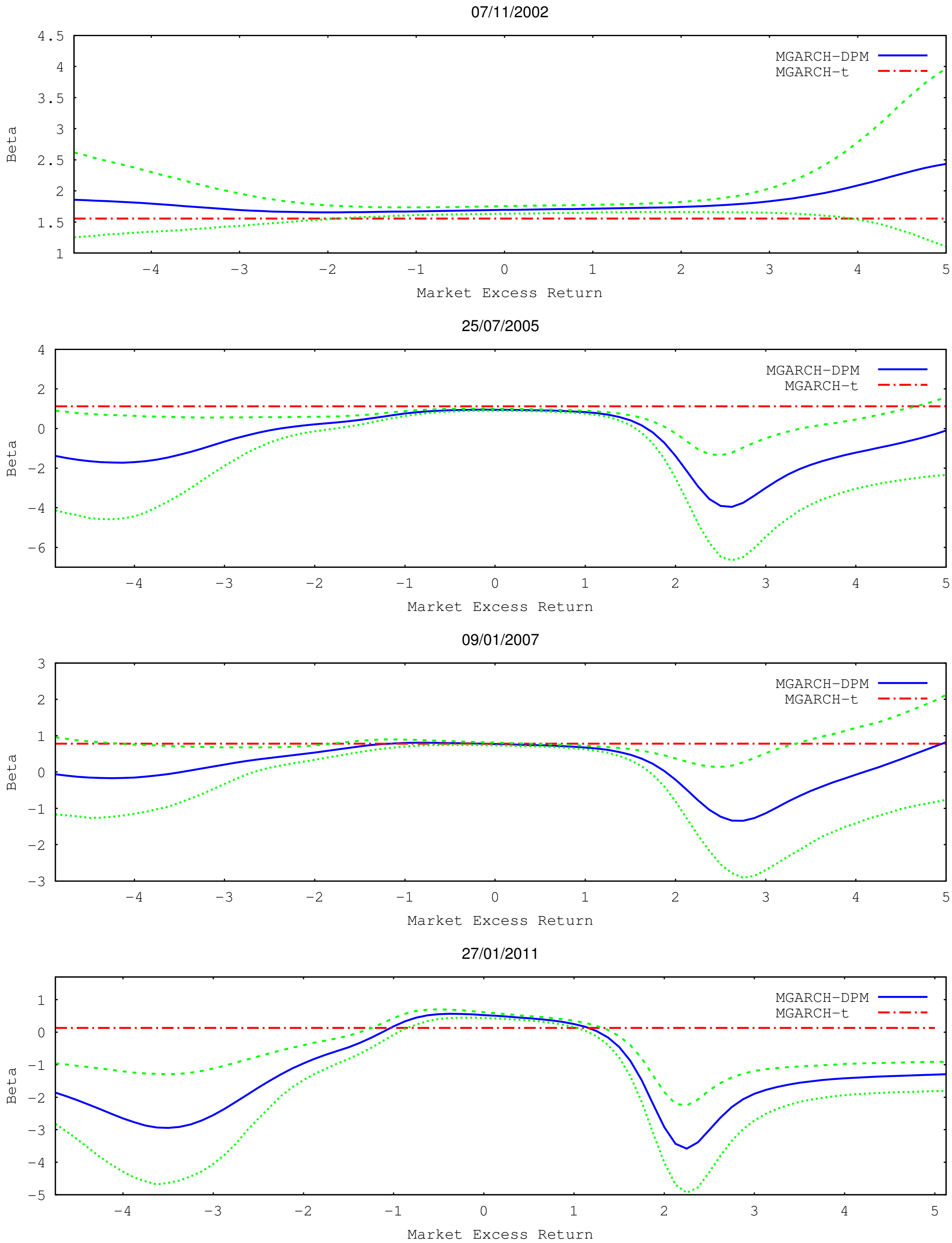

Figure 12: The posterior mean and 0.90 density intervals of IBM's conditional beta as a function of the excess market return from the MGARCH-DPM model. The red line shows the beta coefficients estimated with MGARCH-t model.

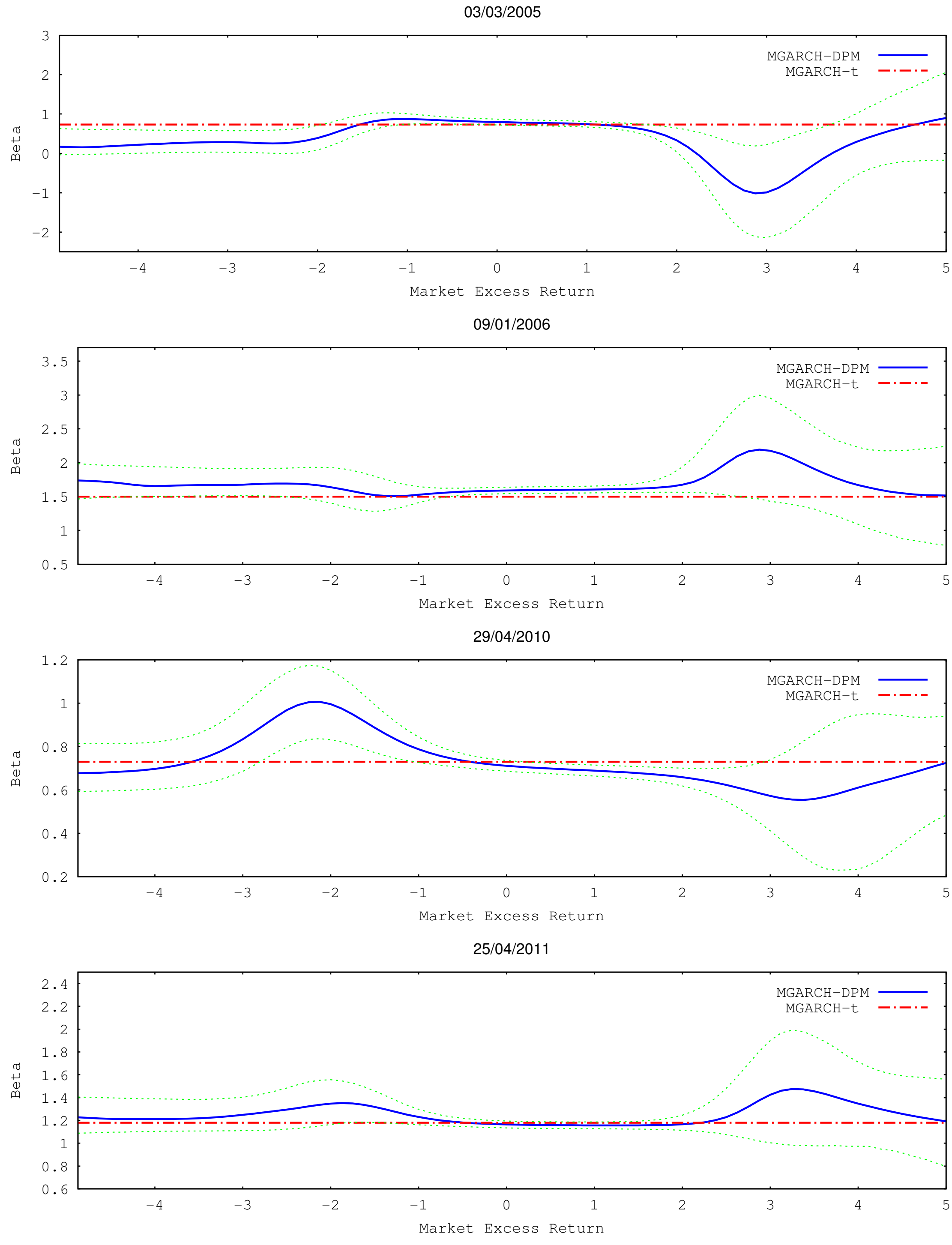

Figure 13: The posterior mean and 0.90 density intervals of XOM's conditional beta as a function of the excess market return from the MGARCH-DPM model. The red line shows the beta coefficients estimated with MGARCH-t model.

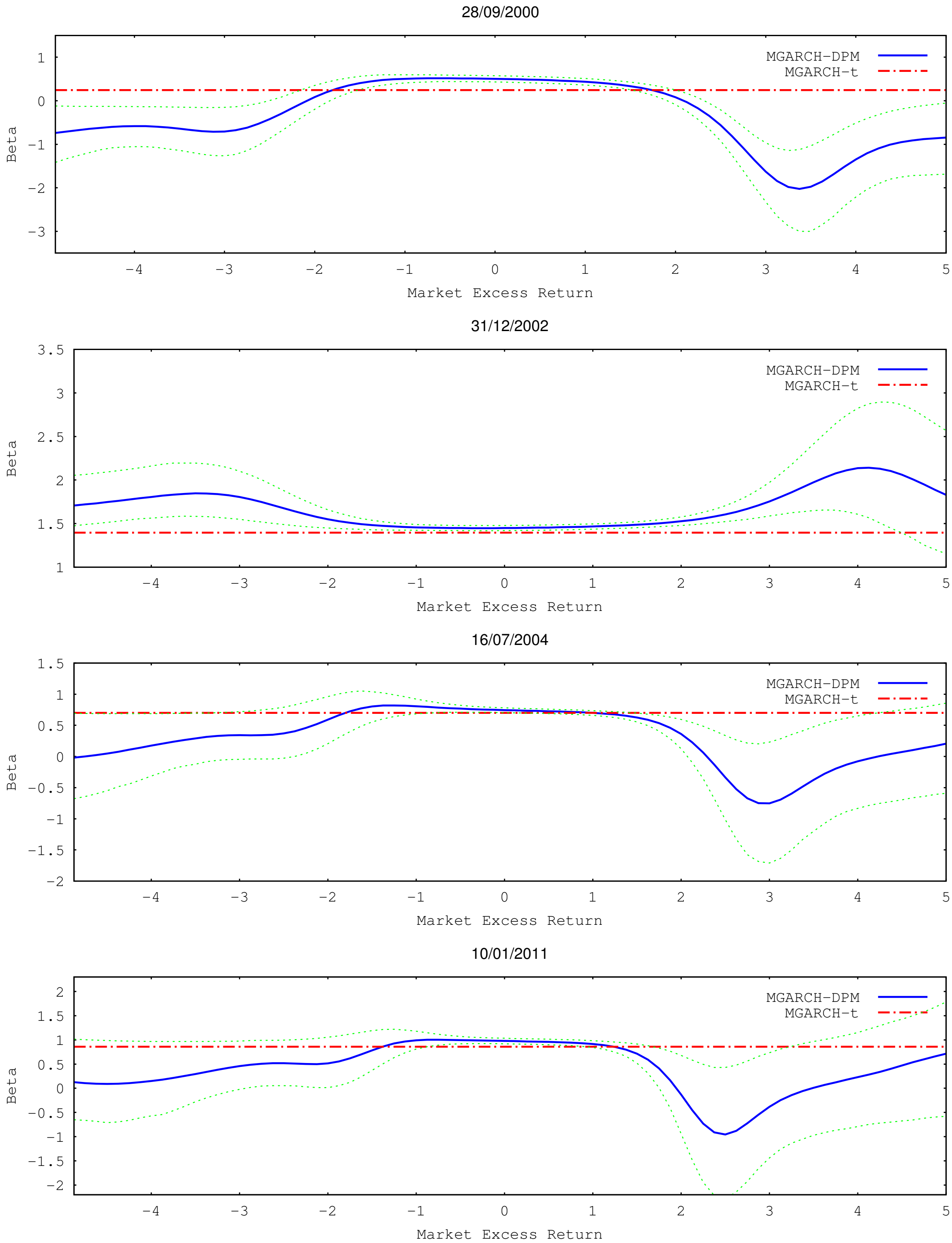

Figure 14: The posterior mean and 0.90 density intervals of GE's conditional beta as a function of the excess market return from the MGARCH-DPM model. The red line shows the beta coefficients estimated with MGARCH-t model.

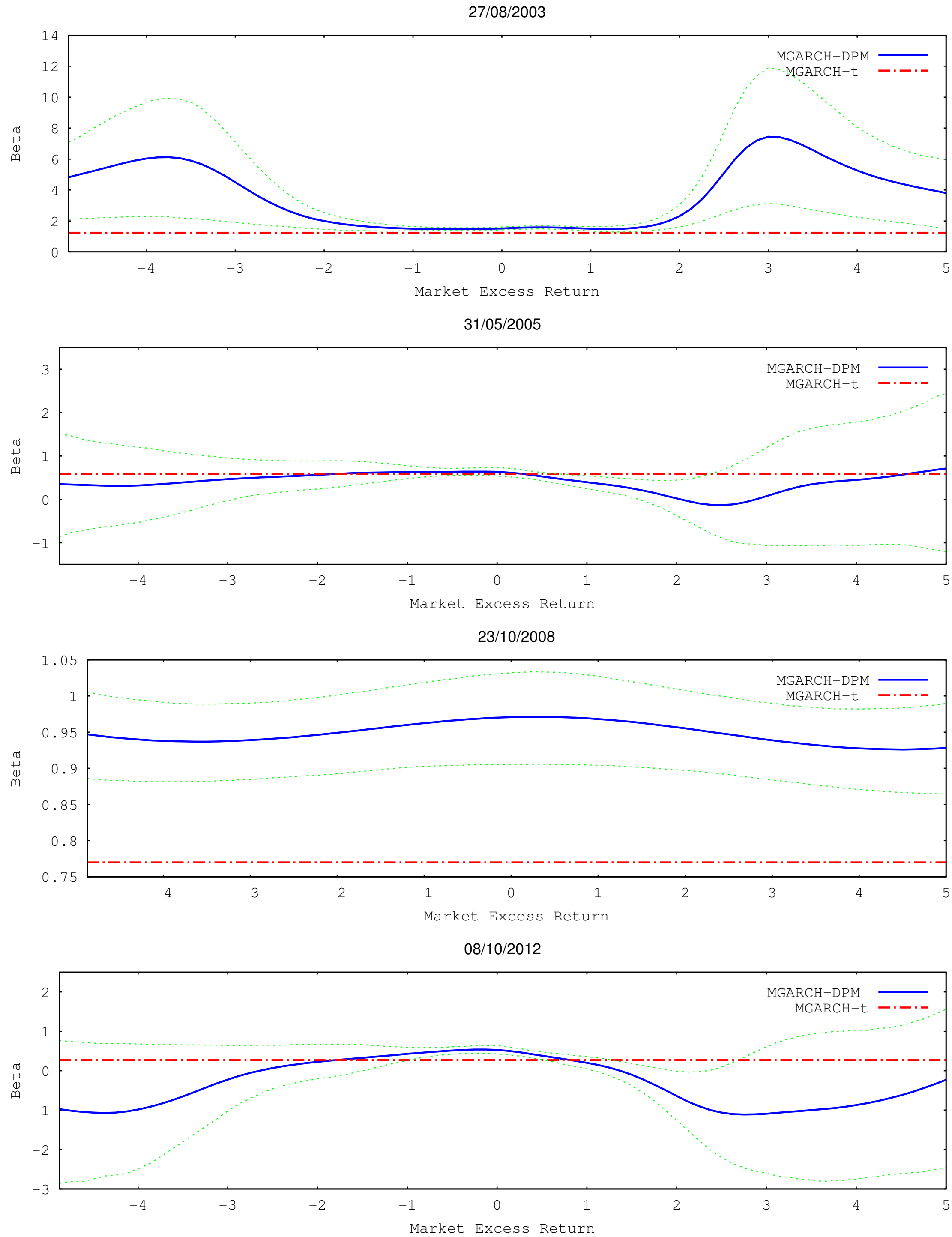

Figure 15: The posterior mean and 0.90 density intervals of AMGN's conditional beta as a function of the excess market return from the MGARCH-DPM model. The red line shows the beta coefficients estimated with MGARCH-t model.

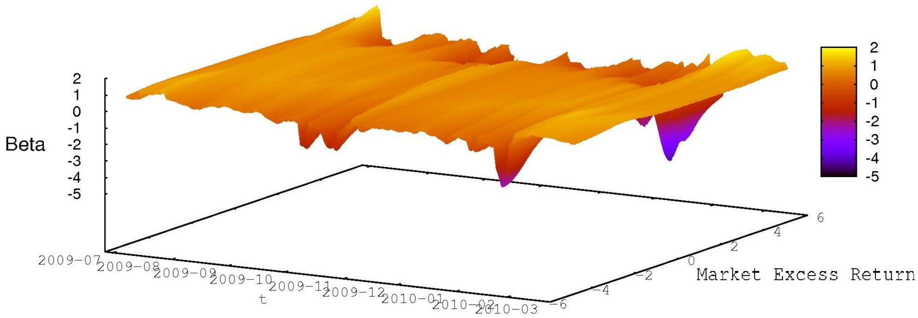

Figure 16: The posterior mean of IBM's nonparametric conditional beta as a function of excess market return and time from 2009-07 to 2010-03 estimated with MGARCH-DPM model.

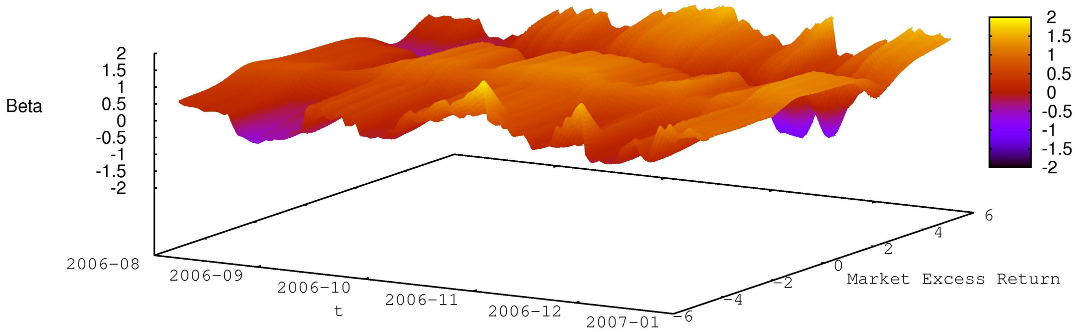

Figure 17: The posterior mean of XOM's nonparametric conditional beta as a function of excess market return and time from 2006-08 to 2007-01 estimated with MGARCH-DPM model.

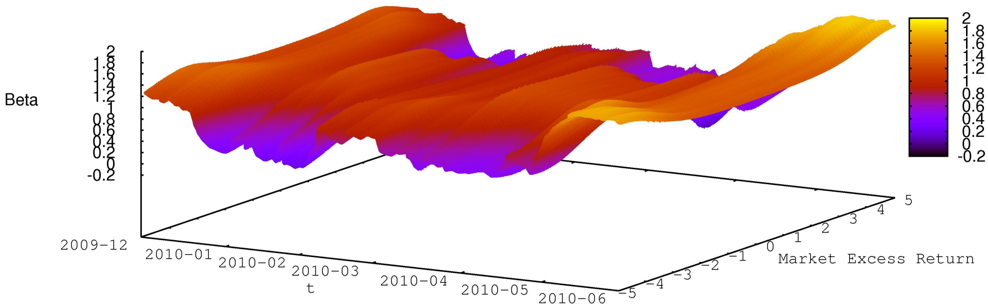

Figure 18: The posterior mean of GE's nonparametric conditional beta as a function of excess market return and time from 2009-12 to 2010-06 estimated with MGARCH-DPM model.

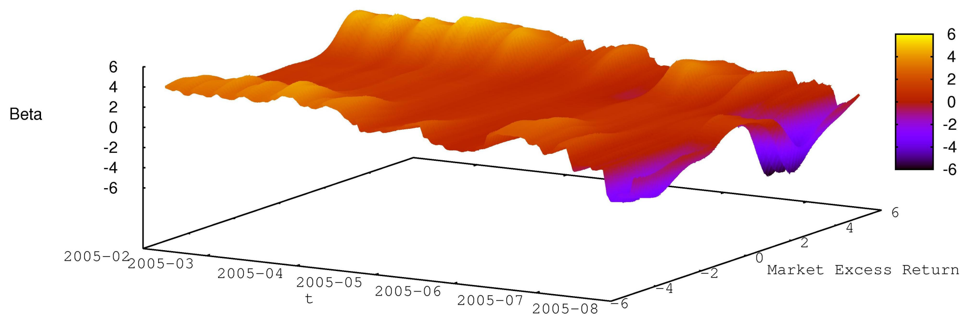

Figure 19: The posterior mean of AMGN's nonparametric conditional beta as a function of excess market return and time from 2005-02 to 2005-08 estimated with MGARCH-DPM model.

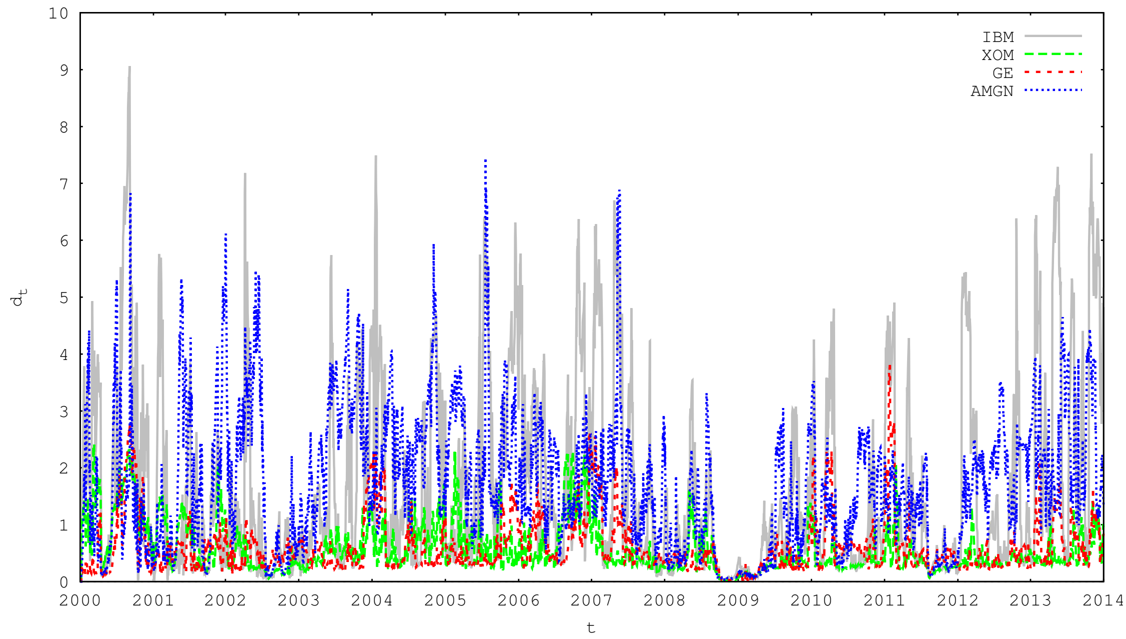

Figure 20: Variability of conditional beta with respect to the contemporaneous value of market excess returns over time for different stocks. $d_t = \max_{\mathbf{r}_{m,t}} b_{m,t}(\mathbf{r}_{m,t}) - \min_{\mathbf{r}_{m,t}} b_{m,t}(\mathbf{r}_{m,t})$.

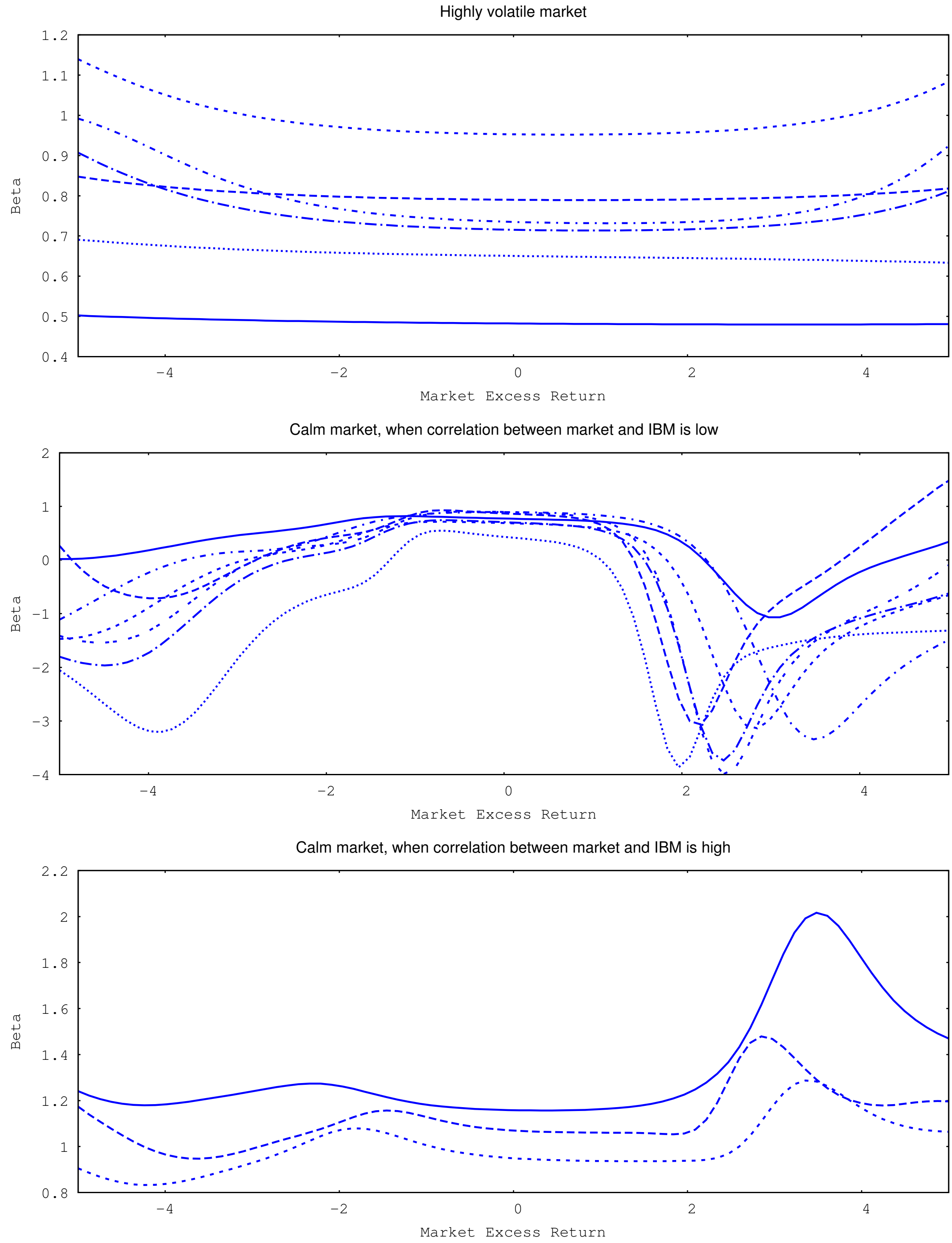

Figure 21: IBM: conditional beta as a function of the market excess return for various dates grouped by market conditions and correlation.

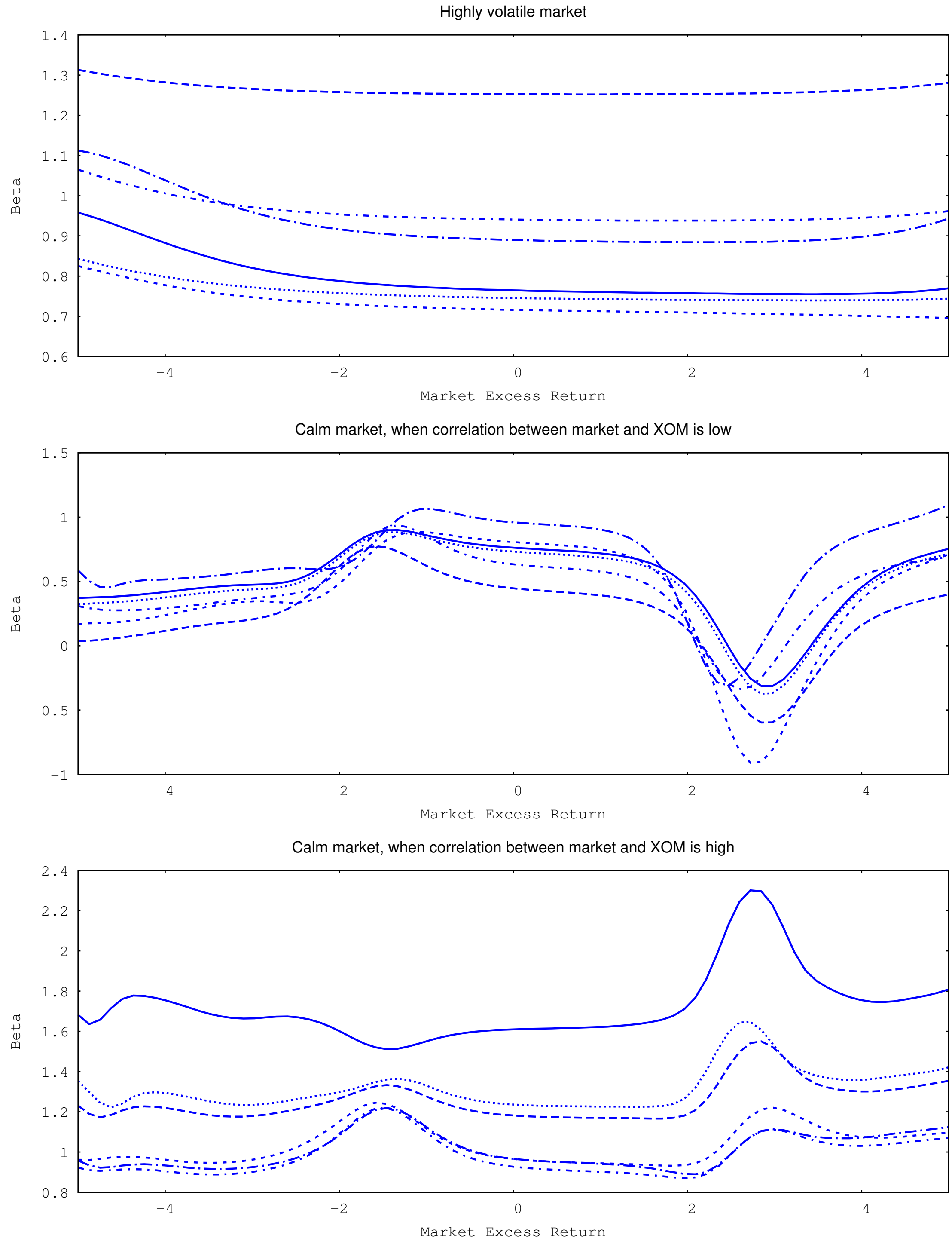

Figure 22: XOM: conditional beta as a function of the market excess return for various dates grouped by market conditions and correlation.

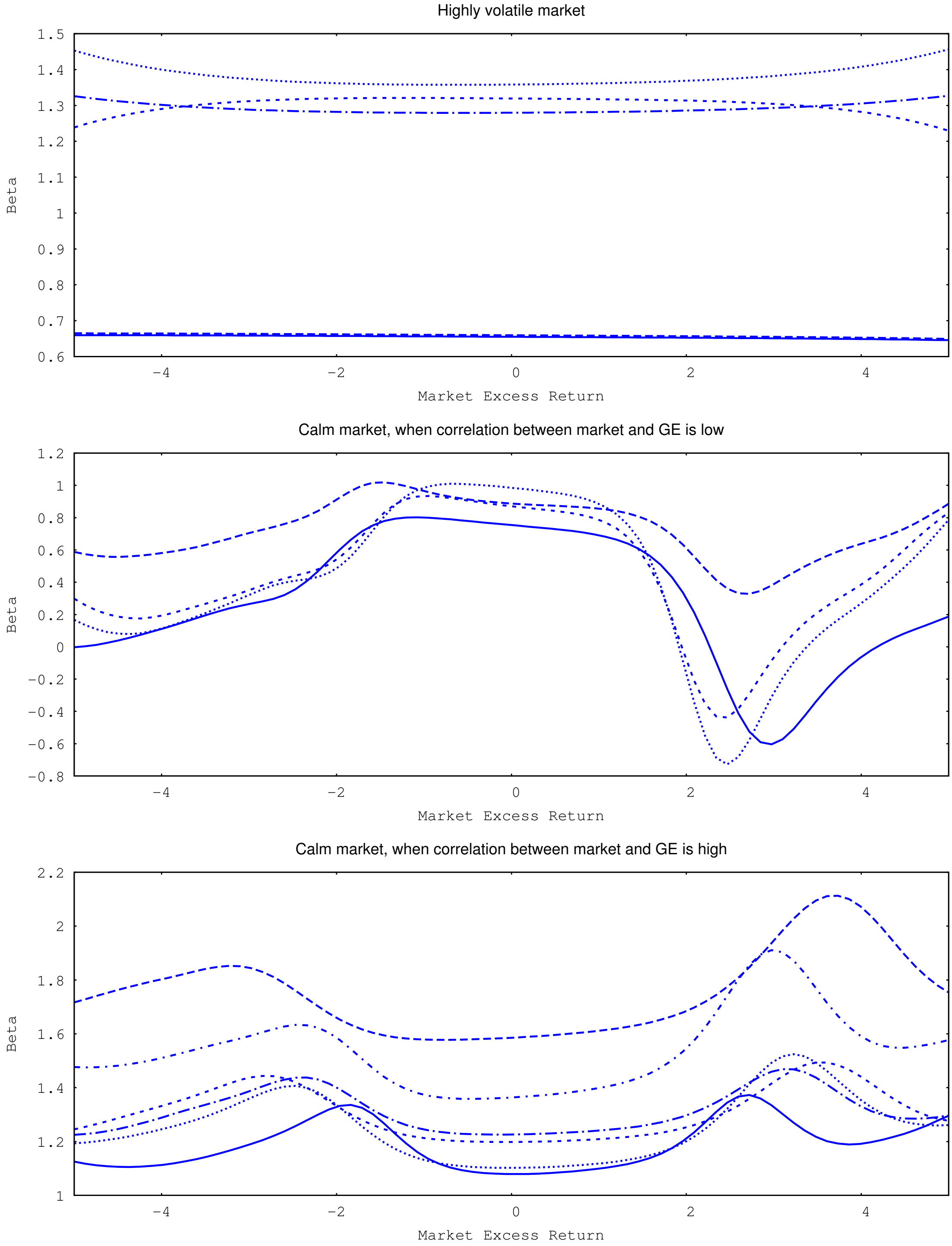

Figure 23: GE: conditional beta as a function of the market excess return for various dates grouped by market conditions and correlation.

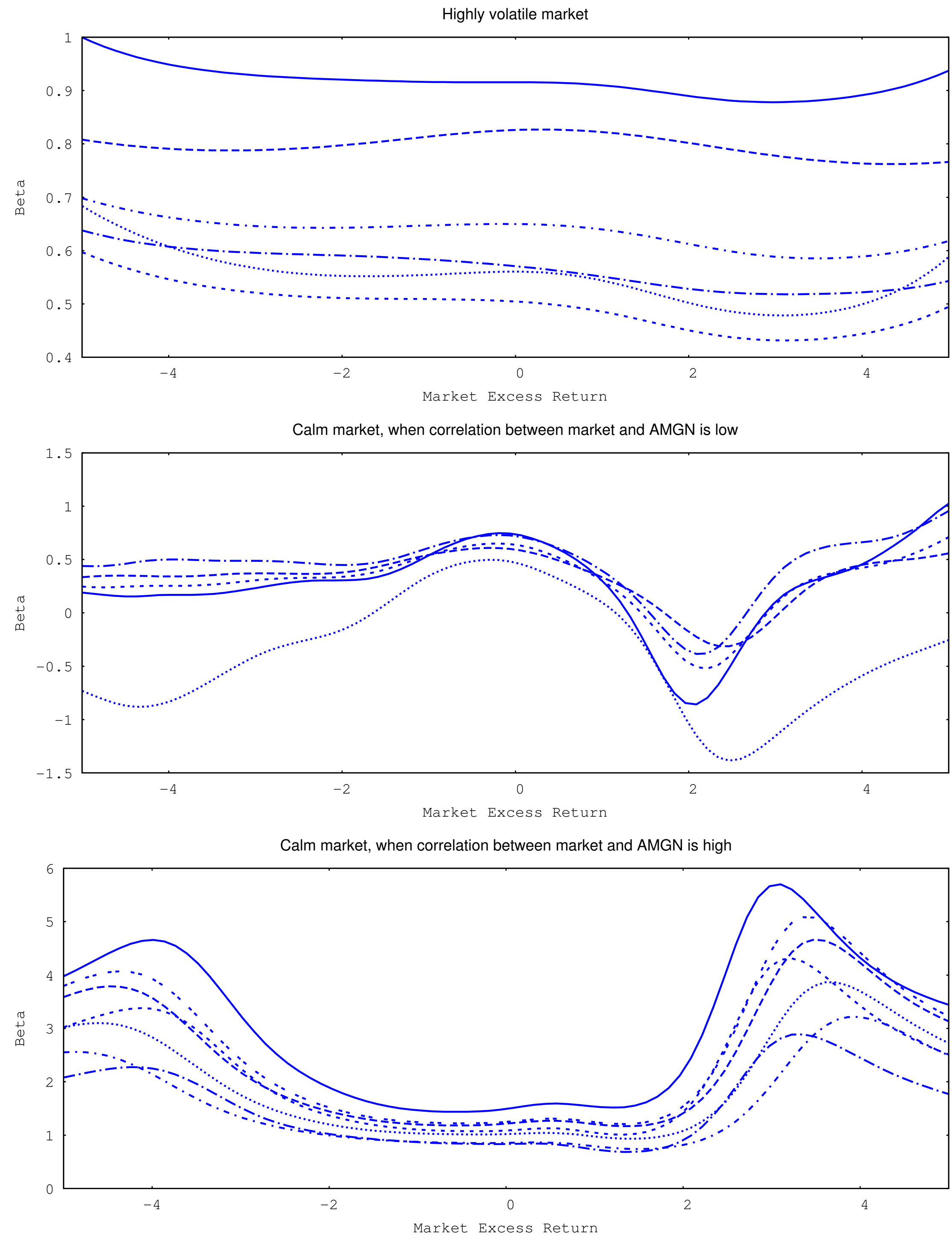

Figure 24: AMGN: conditional beta as a function of the market excess return for various dates grouped by market conditions and correlation.

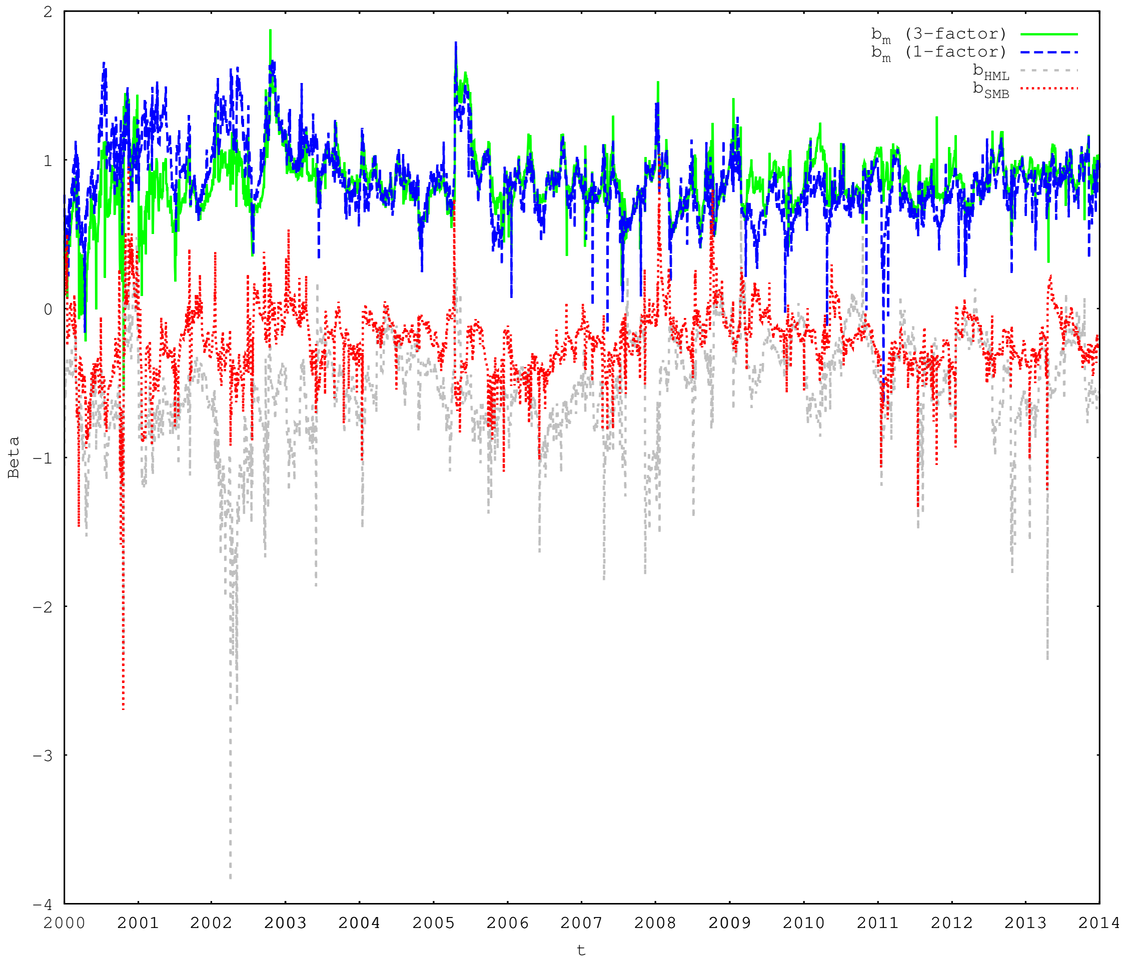

Figure 25: This figure illustrates the time series patterns of the posterior mean of $b_{m,t}$, $b_{SMB,t}$, and $b_{HML,t}$ estimated in Fama-French three-factor model (FF-MGARCH-DPM) using daily returns on IBM, the market, SMB, and HML. Beta from the one-factor model (MGARCH-DPM) is also included.

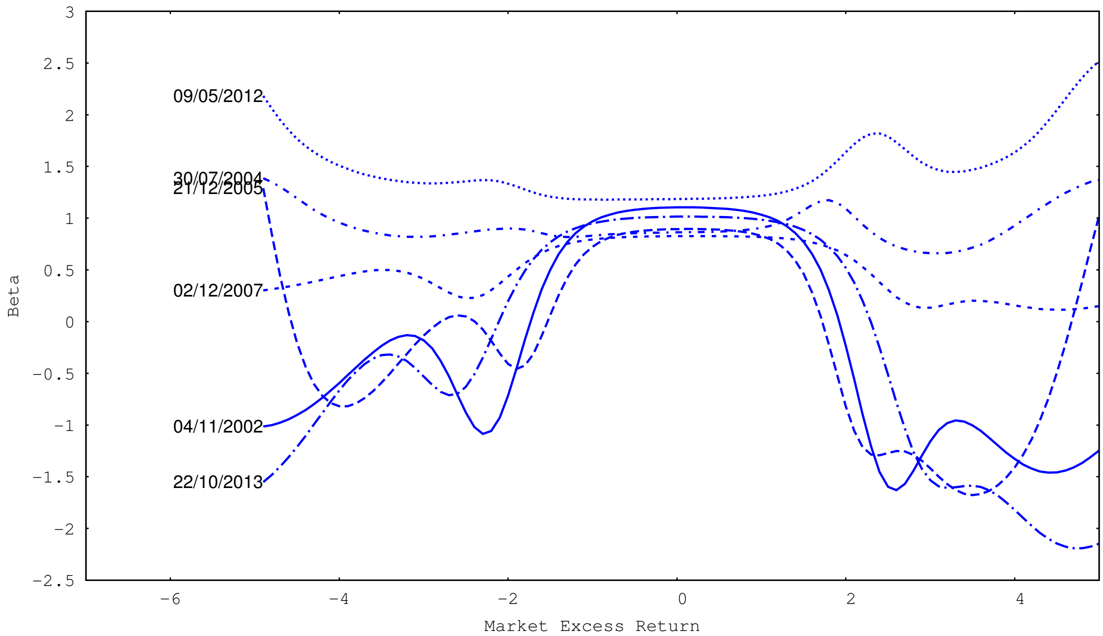

Figure 26: IBM: posterior mean of $b_{m,t}$ as a function of the market excess return from the three-factor model (FF-MGARCH-DPM).

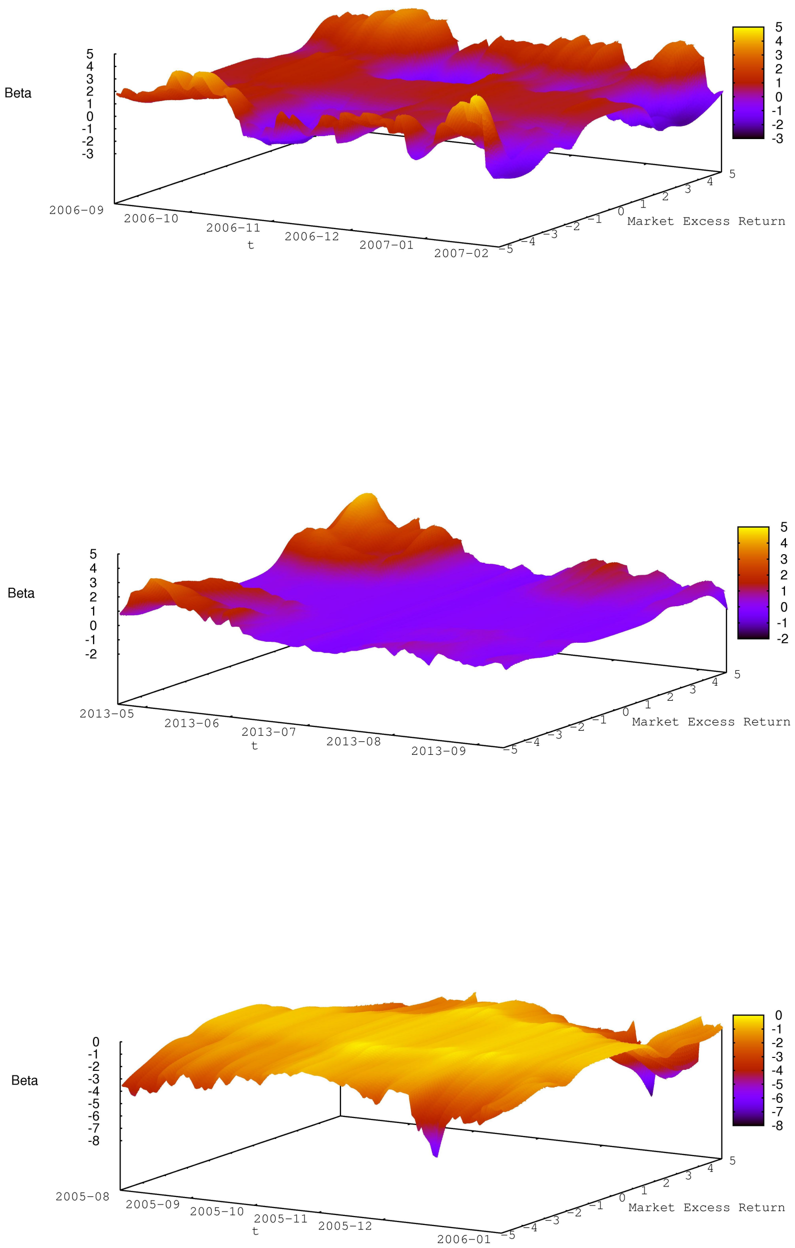

Figure 27: This figure shows the posterior mean of the $b_{m,t}$ (top panel), $b_{SMB,t}$ (middle panel), and $b_{HML,t}$ (bottom panel) as functions of the excess market return from the three-factor model (FF-MGARCH-DPM).