Nonparametric Dynamic Conditional Beta*

John M. Maheu † Azam Shamsi ‡

This version
May 2018

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. We extend the model to several factors and find empirical support for a three factor model with nonlinear factor sensitives.

Key words: Dirichlet Process Mixture; GARCH; Beta.

*We are grateful for comments from an Associate Editor, two anonymous referees, 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. Maheu thanks the SSHRC of Canada for financial support.

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

‡Faculty of Business Administration, University of New Brunswick, Fredericton, NB, Canada, E3B 5A3. E-mail address: azam.shamsi@unb.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 (2016) 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 (2016) 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 linear parametric pricing relationship holds more generally for the elliptic family of distributions.[1] This is an attractive approach but may be limiting if the parametric distributional assumptions are not valid.

This paper extends the conditionally linear parametric pricing relationship to a nonlinear setting which arises from deviations from the elliptic family of distributions. In this framework beta is the sensitivity of the expected excess return to a factor. This is defined as the respective derivative and in general will be a nonlinear function of the factors.[2]

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 and results in beta being nonlinearly dependent on the market. This includes asymmetric distributions which the elliptic family omit being only symmetric.

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. In this setting,

[1]This approach is consistent with a pricing kernel linear in the factors.

[2]Similarly, the nonlinear asset pricing models of Dittmar (2002), Guidolin & Timmermann (2008) and Maheu et al. (2013) lead to asset pricing models with higher order moments and nonlinear pricing components.

beta is defined as the derivative of the conditional expectation of an asset with respect to the factor. 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 nonlinear 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. This nonlinear dependence is robust to different MGARCH specifications as well as more factors in the model.

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 nonlinear dynamics that are absent in other models.

We extend the model to several factors and find empirical support for a three factor model with nonlinear factor sensitives for four of the five firms analyzed. We use a new approach to selecting the number of factors in a model. Since specifications with a different number of factors are not comparable by the usual Bayes factors due to different dimensions of the dependent variable we select the number of factors based on the conditional predictive likelihood. This relies on the conditional predictive likelihood of the individual stock return derived from models with different dimensions and is directly comparable across specifications.

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, key features of the semiparametric MGARCH model, and the use of the Dirichlet process prior. Posterior sampling is detailed in Section 4. The derivation of the nonparametric dynamic conditional beta is presented in Section 5. Data is introduced in Section 6, and Section 7 compares the performance of the proposed model to the benchmark model. Applications of the semiparametric model are found in Section 8, and Section 9 provides some implications of the semiparametric model in finance. In Section 10, we assess models with different number of factors and discuss alternative model specifications. 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 (2016). Engle (2016) 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 an asset be y_t and a vector of q regressors (factors) including the excess market return be $x_t = (x_{1,t}, ..., x_{q,t})^{'}$. $r_t = (y_t, x_t^{'})'$ 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).

Following $r_t = (y_t^{'}, x_t^{'})^{'}$ similarly partition $\mu = (\mu_y^{'}, \mu_x^{'})^{'}$ and

$$H_t = \begin{bmatrix} H_{yy,t} & H_{yx,t} \\ H_{yx,t} & H_{xx,t} \end{bmatrix}.$$

Applying the properties of the Student-t distribution (Roth 2013) the conditional distribution of y_t given x_t is

$$y_t | x_t \sim t(\mu_{y|x}, H_{t,y|x}, \nu_{y|x}), \tag{2.3}$$

$$\mu_{y|x} = \mu_y + H_{yx,t} H_{xx,t}^{-1}(x_t - \mu_x), \tag{2.4}$$

$$H_{t,y|x} = \frac{\nu + (x_t - \mu_x)^{'} H_{xx,t}^{-1}(x_t - \mu_x)}{\nu + q}(H_{yy,t} - H_{yx,t} H_{xx,t}^{-1} H_{yx,t}^{'}), \tag{2.5}$$

$$\nu_{y|x} = \nu + q, \tag{2.6}$$

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

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

$$E[y_t | x_{m,t}, H_t] = \mu_y + H_{ym,t} H_{mm,t}^{-1}(x_{m,t} - \mu_m). \tag{2.7}$$

This tells how the expected excess return of asset y_t reacts to any value of the market. If the market shock is zero ($x_{m,t} = \mu_m$) then the expected value is μ_y 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

$$\beta_t = H_{ym,t} H_{mm,t}^{-1}. \tag{2.8}$$

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

There are several advantages 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 pins 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.

We estimate the model from a Bayesian perspective. The posterior density has the non-standard form

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

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 return 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 $r_t = (y_t, x_{1,t}, ..., x_{q,t})'$ 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. B_j is a symmetric positive definite matrix which scales H_t to yield a better estimate of the joint density of data.

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 $x_t = (x_{1,t}, ..., x_{q,t})'$ then the conditional density of y_t given x_t is a mixture distribution as well and the conditional expectation can be written as the following weighted mixture

$$E(y_t|x_t, H_t) = \sum_{j=1}^{\infty} q_j(x_t) E(y_t|x_t, \mu_j, B_j, H_t). \tag{3.2}$$

The weights, $q_j(x_t) \propto \omega_j f(x_t|\mu_j, B_j, H_t)$ are a function of the factors and affect how much each conditional expectation, $E(y_t|x_t, \mu_j, B_j, H_t)$, 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. We label this as a *beta* but could also call it a nonlinear factor sensitivity. It measures the sensitivity of the conditional mean to a small change in the factor. For instance, if there is one factor, $q = 1$, we have

$$b_t^x = \frac{\partial E(y_t|x_t, H_t)}{\partial x_t}. \tag{3.3}$$

The conditional beta at time t is not constant in general but it changes as the contemporaneous value of the corresponding factor changes. Note that this definition of beta yields (2.8) for elliptical distributions and provides a coherent definition for beta based on conditional expectations that are linear and nonlinear functions of factors. Next we introduce the Dirichlet process prior to estimate the mixture model while Section 5 derives an estimate of the nonparametric conditional beta from posterior simulation.

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 non-negligible 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,

$$r_t \quad |\phi_t, H_t \sim N(\xi_t, H_t^{1/2}\Lambda_t(H_t^{1/2})'), \ t = 1, ..., T \tag{3.4}$$

$$\phi_t \quad \equiv \{\xi_t, \Lambda_t\}|G \sim G, \tag{3.5}$$

$$G| \quad \alpha, G_0 \sim DP(\alpha, G_0), \tag{3.6}$$

$$G_0 \quad \equiv N(\mu_0, D) \times \mathcal{W}^{-1}(B_0, \nu_0), \tag{3.7}$$

$$H_t \quad = \Gamma_0 + \Gamma_1 \odot (r_{t-1} - \eta)(r_{t-1} - \eta)' + \Gamma_2 \odot H_{t-1}. \tag{3.8}$$

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

$$p(r_t|\mu, B, W, H_t) = \sum_{j=1}^{\infty} \omega_j N(r_t|\mu_j, H_t^{1/2} B_j (H_t^{1/2})'), \tag{3.9}$$

$$\omega_1 = v_1, \ \omega_j = v_j \prod_{l=1}^{j-1} (1 - v_l), \ j > 1, \tag{3.10}$$

$$v_j \overset{iid}{\sim} \text{Beta}(1, \alpha), \tag{3.11}$$

$$\mu_j \overset{iid}{\sim} N(\mu_0, D), \quad B_j \overset{iid}{\sim} \mathcal{W}^{-1}(B_0, \nu_0), \tag{3.12}$$

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 (3.5), 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.

4 Posterior Sampling

To estimate the unknown parameters in (3.4)-(3.8), we apply an MCMC sampler along with the slice sampler of Walker (2007). 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

$$f(r_t, u_t|W, \Theta) = \sum_{j=1}^{\infty} \mathbf{1}(u_t < \omega_j) N(r_t|\mu_j, (H_t^{1/2})' B_j H_t^{1/2}). \tag{4.1}$$

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

$$p(r_{1:T}, u_{1:T}, s_{1:T}|W, \Theta) = \Pi_{t=1}^{T} \mathbf{1}(u_t < \omega_{s_t}) N(r_t|\mu_{s_t}, (H_t^{1/2}) B_{s_t} (H_t^{1/2})'), \tag{4.2}$$

and the joint posterior is proportional to

$$p(W_{1:K}) \Pi_{j=1}^{K} p(\mu_j, B_j) \Pi_{t=1}^{T} \mathbf{1}(u_t < \omega_{s_t}) N(r_t|\mu_{s_t}, (H_t^{1/2}) B_{s_t} (H_t^{1/2})') \tag{4.3}$$

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

$$B_j | r_{1:T}, s_{1:T}, \mu_j, \Gamma \sim \mathcal{W}^{-1}\left(n_j + \nu_0, B_0 + \sum_{s_t=j}(z_t - H_t^{-1/2}\mu_j)(z_t - H_t^{-1/2}\mu_j)'\right) \quad (4.4)$$

$$\mu_j | r_{1:T}, s_{1:T}, B_j, \Gamma \sim N(\bar{\mu}, \bar{D}) \quad (4.5)$$

in which

$$\bar{D}^{-1} = D^{-1} + \sum_{t|s_t=j} H_t^{-1/2'} B_j^{-1} H_t^{-1/2},\ \bar{\mu} = \bar{D}\left(\sum_{t|s_t=j} H_t^{-1/2'} B_j^{-1} z_t + D^{-1}\mu_0\right). \quad (4.6)$$

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

$$v_j | S \sim \text{Beta}\left(1 + \sum_{t=1}^{T} \mathbf{1}(s_t = j), \alpha + \sum_{t=1}^{T} \mathbf{1}(s_t > j)\right). \quad (4.7)$$

Then we update $W_{1:K}$ based on (3.10).

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 $t = 1, ..., T$,

$$p(s_t = j | r_{1:T}) \propto \mathbf{1}(\omega_j > u_t) N(r_t | \mu_j, H_t^{1/2} B_j (H_t^{1/2})'), j = 1, ..., K. \quad (4.8)$$

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

$$p(\Gamma|\mu, B, S, r_{1:T}) \propto p(\Gamma) \times \prod_{t=1}^{T} N(r_t|\mu_{s_t}, H_t^{1/2} B_{s_t} (H_t^{1/2})') \tag{4.9}$$

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

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

and otherwise rejected. $\widehat{V}$ is proportional to the inverse Hessian matrix of $\ell = \log[p(\Gamma|\mu, B, r_{1:T})]$ evaluated at its posterior mode, $\widehat{\Gamma}$, with all $\mu_j = 0$ and $B_j = I$. This is computed once at the start of estimation with the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm. Although independent of μ, B and S, this provided a good proposal distribution. $\widehat{V}$ is scaled to achieve an acceptance rate between 0.2 and 0.5.

5 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, $r_t = (y_t, x_t)$ where y_t and x_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:

$$\{(\mu_j^{(g)}, B_j^{(g)}), v_j^{(g)}, j = 1, ..., K^{(g)}\}, \{s_t^{(g)}, u_t^{(g)}, t = 1, ..., T\}, H_{1:T}^{(g)} = \{H_1^{(g)}, ..., H_T^{(g)}\}, \tag{5.1}$$

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

$$G^{(g)} = \sum_{j=1}^{K^{(g)}} \omega_j^{(g)} \delta_{\theta_j^{(g)}} + \left(1 - \sum_{j=1}^{K^{(g)}} \omega_j^{(g)}\right) G_0(\theta), \tag{5.2}$$

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 $(\mathbf{y}_t, \mathbf{x}_t)$ conditional on $G^{(g)}$ as

$$p(\mathbf{y}_t, \mathbf{x}_t|r_{1:T}, G^{(g)}) = \sum_{j=1}^{K^{(g)}} \omega_j^{(g)} f(\mathbf{y}_t, \mathbf{x}_t|\theta_j^{(g)}) + \left(1 - \sum_{j=1}^{K^{(g)}} \omega_j^{(g)}\right) \int f(\mathbf{y}_t, \mathbf{x}_t|\theta) G_0(\theta) d\theta, \tag{5.3}$$

where $f(\mathbf{y}_t, \mathbf{x}_t|\theta)$ is the multivariate normal density.

To assess the nonlinear regression function $E(\mathbf{y}_t|\mathbf{x}_t, r_{1:T})$, or the conditional beta of the individual stock i, we require the conditional density derived from this predictive (joint) density of $(\mathbf{y}_t, \mathbf{x}_t)$. Therefore,[3]

$$p(\mathbf{y}_t|\mathbf{x}_t, r_{1:T}, G^{(g)}) = \frac{p(\mathbf{y}_t, \mathbf{x}_t|r_{1:T}, G^{(g)})}{p(\mathbf{x}_t|r_{1:T}, G^{(g)})} \tag{5.4}$$

$$= \frac{p(\mathbf{y}_t, \mathbf{x}_t|r_{1:T}, G^{(g)})}{\sum_{j=1}^{K^{(g)}} \omega_j^{(g)} f_2(\mathbf{x}_t|\theta_j^{(g)}) + \left(1 - \sum_{j=1}^{K^{(g)}} \omega_j^{(g)}\right) \int f_2(\mathbf{x}_t|\theta) G_0(\theta) d\theta}$$

$$= \sum_{j=1}^{K^{(g)}} q_j^{(g)}(\mathbf{x}_t) f(\mathbf{y}_t|\mathbf{x}_t, \theta_j^{(g)}) + \left(1 - \sum_{j=1}^{K^{(g)}} q_j^{(g)}(\mathbf{x}_t)\right) f(\mathbf{y}_t|\mathbf{x}_t, G_0), \tag{5.5}$$

where

$$q_j^{(g)}(\mathbf{x}_t) = \frac{\omega_j^{(g)} f_2(\mathbf{x}_t|\theta_j^{(g)})}{\sum_{j=1}^{K^{(g)}} \omega_j^{(g)} f_2(\mathbf{x}_t|\theta_j^{(g)}) + \left(1 - \sum_{j=1}^{K^{(g)}} \omega_j^{(g)}\right) \int f_2(\mathbf{x}_t|\theta) G_0(\theta) d\theta} \tag{5.6}$$

and $f_2(\mathbf{x}_t|\theta_j^{(g)})$ is the marginal (normal) density of $\mathbf{x}_t$ and $f(\mathbf{y}_t|\mathbf{x}_t, G_0)$ is the conditional distribution using the base measure. The terms $q_j^{(g)}(\mathbf{x}_t)$ determine which components in the mixture receive more weight. Clusters that have a marginal density $f_2(\mathbf{x}_t|\theta_j^{(g)})$ that has a higher likelihood value for $\mathbf{x}_t$ will receive larger weights. The marginal density, and hence relative weight of clusters, will change with $\mathbf{x}_t$ 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. An estimate of the unknown conditional distribution is obtained as

$$p(\mathbf{y}_t|\mathbf{x}_t, r_{1:T}) \approx \frac{1}{M} \sum_{g=1}^{M} p(\mathbf{y}_t|\mathbf{x}_t, r_{1:T}, G^{(g)}). \tag{5.7}$$

Our focus is on the conditional mean of $\mathbf{y}_t$ given $\mathbf{x}_t$. Using the properties of the normal distribution the conditional mean directly comes from (5.5) and is

$$E(\mathbf{y}_t|\mathbf{x}_t, r_{1:T}, G^{(g)}) = \sum_{j=1}^{K^{(g)}} q_j^{(g)}(\mathbf{x}_t)[\mu_{j,1}^{(g)} + \beta_{jt}^{(g)}(\mathbf{x}_t - \mu_{j,2}^{(g)})] + \tag{5.8}$$

$$\left(1 - \sum_{j=1}^{K^{(g)}} q_j^{(g)}(\mathbf{x}_t)\right) \frac{\int [\mu_1 + \beta_t(\mathbf{x}_t - \mu_2)] N(\mathbf{x}_t|\mu_2, (H_t^{(g)^{1/2}} B H_t^{(g)^{1/2'}})_{22}) p(\mu, B) d\mu dB}{\int N(\mathbf{x}_t|\mu_2, (H_t^{(g)^{1/2}} B H_t^{(g)^{1/2'}})_{22}) p(\mu, B) d\mu dB}.$$

The cluster specific beta is defined as

$$\beta_{jt}^{(g)} = \frac{(H_t^{(g)^{1/2}} B_j H_t^{(g)^{1/2'}})_{12}}{(H_t^{(g)^{1/2}} B_j H_t^{(g)^{1/2'}})_{22}} \tag{5.9}$$

where the subscript (i, j) on $()_{ij}$ denotes element (i, j) of the matrix and β_t in the second line of (5.8) is defined as $\beta_{jt}^{(g)}$ except B_j is replaced with B. The numerator and denominator in the last term of (5.8) can be approximated by simulation.

[3]A referee has pointed out that an alternative and potentially simplier approximation would be to conditional on u_t only as in Section 4.3 of Griffin & Walker (2011).

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

$$E(\mathbf{y}_t|\mathbf{x}_t, r_{1:T}) \approx \frac{1}{M}\sum_{g=1}^{M} E(\mathbf{y}_t|\mathbf{x}_t, r_{1:T}, G^{(g)}). \tag{5.10}$$

The nonparametric beta is the derivative of this conditional expectation of $\mathbf{y}_t$ given $\mathbf{x}_t$, (5.10) with respect to $\mathbf{x}_t$. This is,

$$b_t^m(\mathbf{x}_t) = \frac{\partial E(\mathbf{y}_t|\mathbf{x}_t, r_{1:T})}{\partial \mathbf{x}_t}\Big|_{x_t=\mathbf{r}_{m,t}}. \tag{5.11}$$

An estimate of this is obtained by averaging over the posterior draws of the derivative of (5.8). Similarly, credible intervals can be obtained from the associated quantiles of the derivative of (5.8) for each posterior draw of $G^{(g)}$. Full details on this derivative and estimate are provided in the appendix.

In the case that we have more than one factor, we follow the same process. We first estimate the joint model and back out the conditional distribution of the stock return y_t given all factors. The nonparametric conditional beta in this case is a vector. It is defined analogously to (5.11) as the partial derivative with respect to the factor. For instance in the case of the Fama-French three factor model with size factor ($x_{SMB,t}$), value factor ($x_{HML,t}$), and market factor (Fama & French 1993), beta for size factor is defined as

$$b_t^{SMB}(\mathbf{x}_t) = \frac{\partial E(\mathbf{y}_t|\mathbf{x}_{m,t}, \mathbf{x}_{SMB,t}, \mathbf{x}_{t,HML}, r_{1:T})}{\partial \mathbf{x}_{SMB,t}}\Big|_{\substack{x_{m,t}=\mathbf{r}_{m,t}\\ x_{SMB,t}=\mathbf{r}_{SMB,t}\\ x_{HML,t}=\mathbf{r}_{HML,t}}} \tag{5.12}$$

with a similar expression for the other factor coefficients b_t^m and b_t^{HML}.

6 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 five individual stock excess returns for IBM, General Electric or GE, Exxon or XOM, Amgen or AMGN, and bank of America or BAC 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. Table 1 reports summary statistics. All individual stocks display skewness and excess kurtosis. Daily data for the size factor, $x_{SMB,t}$, value factor, $x_{HML,t}$, and momentum factor, $x_{MOM,t}$, are obtained from Kenneth French's website.[4]

7 Model Performance

The criteria that we use to compare different models is the value of the log-predictive likelihood. For each particular model $\mathcal{M}$ (i.e., MGARCH-t or MGARCH-DPM), the

[4]See http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/f-f_factors.html, http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/det_mom_factor_daily.html and Fama & French (1993).

predictive likelihood for $r_{L:T}$, $L < T$ is expressed in terms of the one-step-ahead predictive likelihoods,

$$m(r_{L:T}|r_{1:L-1}, \mathcal{M}) = \Pi_{t=L}^{T} p(r_t|r_{1:t-1}, \mathcal{M}) \tag{7.1}$$

where $L > 1$ 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

$$\begin{aligned} p(r_t|r_{1:t-1}, \mathcal{M}) &= \int p(r_t|\theta, r_{1:t-1}, \mathcal{M}) p(\theta|r_{1:t-1}, \mathcal{M}) d\theta \qquad (7.2) \\ &\approx \frac{1}{M} \sum_{g=1}^{M} p(r_t|\theta^{(g)}, r_{1:t-1}, \mathcal{M}) \end{aligned}$$

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 $\mathcal{M}$.

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 $\alpha \sim \mathcal{G}(0.1, 0.3)$. 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.

Based on (7.2), the predictive likelihoods for the two models are estimated as

$$p(r_t|r_{1:t-1}, \text{MGARCH-t}) \approx \frac{1}{M} \sum_{g=1}^{M} t(r_t|\mu^{(g)}, H_t^{(g)}, \nu^{(g)}), \tag{7.3}$$

$$p(r_t|r_{1:t-1}, \text{MGARCH-DPM}) \approx \frac{1}{M} \sum_{g=1}^{M} N(r_t|\mu_{s_t^{(g)}}^{(g)}, H_t^{(g)^{1/2}} B_{s_t^{(g)}}^{(g)} H_t^{(g)^{1/2'}}). \tag{7.4}$$

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: $H_t^{(g)} = \Gamma_0^{(g)} + \Gamma_1^{(g)} \odot (r_{t-1} - \eta^{(g)})(r - \eta^{(g)})'_{t-1} + \Gamma_2^{(g)} \odot H_{t-1}^{(g)}$. In MGARCH-DPM model, at each iteration g, $s_t^{(g)}$

Table 2 reports the out-of-sample log-predictive likelihoods for the 1-factor MGARCH-t and MGARCH-DPM models, and the log-Bayes factor over 2012/03/12 to 2013/12/31. Bivariate models based on daily excess returns on IBM, GE, XOM, AMGN and BAC 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 211. This is very strong evidence of significant deviations from the Student-t MGARCH model.

8 Applications of 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 5) 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.

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 3-7 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 very similar. The tables report c, the number of components in the mixture used to estimate the unknown density. On average, the bivariate joint density of IBM, XOM, GE, and BAC with the market is estimated using about 3.6-6.3 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 1 and 2 compare the posterior mean of the *realized* beta over time derived from both models for IBM. For MGARCH-t model, the posterior mean of (2.8) is reported while for the MGARCH-DPM model the posterior mean of (5.11) 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. The 0.95 density intervals are tight and support the time-variation of the conditional betas.

Figures 3–5 illustrate posterior mean of the conditional beta as a function of the contemporaneous market excess return using (5.11) at several dates for three stocks. 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.

The nonlinear relationship between beta and the market transfers directly into the conditional expected excess return. For example, Figure 6 displays the posterior mean of the conditional expected excess return of IBM given different values of the contemporaneous market excess return, derived from (5.10), for dates for which the conditional betas are illustrated in Figure 3. 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 7-9 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 and stocks. 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.

8.1 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 $x_{m,t}$ (beta is almost constant with respect to $x_{m,t}$) while in other time periods beta changes significantly with $x_{m,t}$. To measure the sensitivity of $b_t^m(\mathbf{x}_t)$ to $\mathbf{x}_t$ at each time t consider the following measure

$$d_t = \max_{\mathbf{x}_t} b_t^m(\mathbf{x}_t) - \min_{\mathbf{x}_t} b_t^m(\mathbf{x}_t), \tag{8.1}$$

where $b_t^m(\mathbf{x}_t)$ is defined in (5.11) for the one factor model with $\mathbf{x}_t$ the market. Large values of d_t indicate that $b_t^m(\mathbf{x}_t)$ is strongly sensitive to $\mathbf{x}_t$, while a $d_t = 0$ indicates no sensitivity. The MGARCH-t model has a $d_t = 0$ for all t. Figure 10 illustrates this d_t over time for all individual stocks. Among these four stocks, the dynamic conditional beta for IBM and BAC have the most sensitivity and XOM has the least sensitivity to $\mathbf{x}_t$. 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 $b_t^m(\mathbf{x}_t)$ changes with different market conditions Figure 11 show the broad trends that we find in all stocks using AMGEN as an example. 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 $b_t^m(\mathbf{x}_t)$ from $\mathbf{x}_t$ is asymmetric. Frequently $b_t^m(\mathbf{x}_t)$ is more sensitive to large positive values of $\mathbf{x}_t$ 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.

9 Financial Applications

From Equation (5.7), we are able to examine the whole conditional density of the stock given factors. Therefore we are able to study the individual stock's conditional expected return under different risk scenarios. For example, the semiparametric model allows us to study the effect of big shocks in the factors (i.e., market return) on stock's expected return and risk measures such as value-at-risk.

Consider the predictive conditional expected return of IBM at time t derived from the 1-factor model, $E[y_{IBM,t}|x_{m,1:t-1}, y_{IBM,1:t-1}]$. Using the semiparametric model, this value is a nonlinear function of $x_{m,t}$. Therefore, when a large shock is expected to the market, this shock affects our expectation of the IBM return nonlinearly. While in the benchmark model this effect is linear. For instance, Figure 12 illustrates IBM's predictive conditional expected return for a specific date (2000-11-30). From this figure we can assess the expected impact of a large positive or negative shock to the market on the value of IBM's return.

Consider a second example of a large realized market shock in 2008-10-28. We can study how the semiparametric model is able to predict the effect of this shock on IBM's expected return. Figure 13 shows IBM's predictive conditional expected return for this day derived from the benchmark model and the semiparametric model. The realized market return and IBM on this day are 9.77% and 9.56%, respectively. This point is illustrated on the graph as well. It is clear how the nonlinearity resulting from the semiparametric model reduces the prediction error. This shows how we can benefit from the semiparametric model in the events that we expect big positive or negative shocks in the market (i.e., a political event, a new financial policy).

Figure 13 shows only one specific date. We looked at all dates that the market has realized a big shock (more than 6% in absolute value) and compared the performance of the semiparametric model with the benchmark. The root mean squared error of the prediction for the benchmark and the semiparameric model is 8.394 and 8.131, respectively, showing the outperformance of the semiparametric model by 3.2% improvement in prediction.

In addition to the expected return, the semiparametric model enables us to study the effect of big shocks in the market on IBM's whole conditional density and different risk measures. Figure 14 illustrates the effect of +5% and −5% shocks in the market return on IBM's predictive conditional density on 2011-03-08 derived from the semiparametric 1-factor model. The value at risk of investment in IBM when we have no shock in the market is 0.652%. A +5% shock in the market return decreases the value at risk to 0.050%, while a −5% shocks in the market return increases the value at risk to 1.805%. Therefore, we can carry out different risk scenario analyses in order to indicate the effect of big shocks in the market on our investment in a specific firm.

10 Robustness and Extensions

In this section we discuss alternative MGARCH specifications, selection of the number of factors and results from these models. Table 2 reports the predictive likelihoods for the BEKK model of Engle & Kroner (1995) coupled with DPM innovations (BEKK-DPM). This replaces (3.8) with the richer specification $H_t = C_0 C_0' + A'(r_{t-1} - \eta)(r_{t-1} - \eta)' A + F' H_{t-1} F$, where C_0 is a lower triangular matrix and A and F are square matrices of dimension $(q + 1)$.[5] This model improves upon the MGARCH-t model but is not

[5]Priors for the DPM portion of the model are the same as the MGARCH-DPM model along with independent $N(0, 100)$ for each parameter of the BEKK specification and restrictions for identification of those parameters.

as good as our model. The log-Bayes factors are all greater than 50 in favour of the MGARCH-DPM versus the BEKK-DPM. In addition, the BEKK-DPM model resulted in very similar nonparametric beta dependence and time series patterns as our benchmark model.

Our discussion has focused on the one-factor model. To determine the number of factors to be used in the proposed model, we compare the values of the *conditional* predictive likelihood of the individual stock returns derived from each model, using different factors.[6] The predictive likelihoods discussed in Section 7 are directly comparable for a common r_t. But when comparing a model with 2 factors versus 3 factors the independent variable x_t is 2 dimensional and 3 dimensional, respectively these predictive likelihood values are not comparable. Instead we compare the conditional predictive likelihood for the individual stock return only. This is obtained from averaging over (5.5) given different posterior draws of $G^{(g)}$ while conditioning on the factor data x_t under study. For instance, for excess returns on a stock, we compare the one-factor model against the two-factor model with $\sum_{t\in D} \log p(y_t|y_{1:t-1}, x_{1,1:t})$ and $\sum_{t\in D} \log p(y_t|y_{1:t-1}, x_{1,1:t}, x_{2,1:t})$ for the out-of-sample observations D.[7]

Table 8 reports the conditional log-predictive likelihood values for the stocks for the MGARCH-DPM model, for 2012/03/12 to 2013/12/31 (500 observations) when we use different factors. These factors include market excess return ($x_{m,t}$), size factor ($x_{SMB,t}$) and value factor ($x_{HML,t}$) from the Fama-French three factor model, and the momentum factor. This table shows that for IBM, GE, and AMGN the three factor model results in better conditional predictive likelihoods relative to the 1 and 4-factor models. The evidence for the three factor model is very strong for these three stocks. For instance, for IBM the log-predictive Bayes factor for the three factor model against the 1-factor model is 101.532. For XOM, the 3 and 1 factor models perform almost the same with a log-BF of the 3 factor model versus the 1-factor model equal to 1.981. For BAC, the 1-factor model significantly outperforms the 3 and 4 factor models.

Here, we look at some empirical results for the three factor MGARCH-DPM model for IBM and AMGN. The three factor MGARCH-DPM model for IBM uses about 4.3 distinct clusters (with $\alpha \approx 0.453$) on average to estimate the joint distribution of $y_{IBM,t}$, $x_{m,t}$, $x_{SMB,t}$ and $x_{HML,t}$ while without considering the size and value factors, the MGARCH-DPM model uses about 5.6 distinct clusters (with $\alpha \approx 0.571$) (Table 3). The estimated number of distinct clusters for the case of AMGN is 11.4 (with $\alpha \approx 1.48$), while without considering the size and value factors, the MGARCH-DPM model uses about 15 distinct clusters (with $\alpha \approx 2.41$) (Table 6).

The nonparametric conditional beta in the three factor MGARCH-DPM model is a vector, $(b_t^m, b_t^{SMB}, b_t^{HML})'$. Figure 15 illustrates the time series patterns of the posterior mean of b_t^m (3-factor), b_t^{SMB}, and b_t^{HML} estimated using the three factor MGARCH-DPM model for IBM. Beta on market excess returns is also compared with the conditional beta derived using the one factor MGARCH-DPM model, b_t^m (1-factor).

Figure 16 show the posterior mean of b_t^m as a function of the market excess return at several times for IBM using the three factor MGARCH-DPM model. The size and value factors are set at their sample mean value. As we see, even after accounting for $x_{SMB,t}$, and $x_{HML,t}$ in the model, the market beta coefficient at each time is sensitive to

[6]A previous version of the paper selected factors based on marginal predictive likelihoods. For example, $p(y_t|x_{1,1:t-1})$ and $p(y_t|x_{1,1:t-1}, x_{2,1:t-1})$ are compared for 1 versus 2 factors. These are derived by integrating out the factors from the bivariate and trivariate density respectively.

[7]We thank a referee for suggesting this.

the contemporaneous shocks in the market.

Figure 17 reveals another interesting finding. These three dimensional figures illustrate the posterior mean of b_t^m, b_t^{SMB}, and b_t^{HML} as functions of the contemporaneous value of market excess returns over time. It clearly can be seen that besides b_t^m, b_t^{SMB}, and b_t^{HML} are also nonlinearly dependent on the contemporaneous value of market excess returns.

11 Conclusion

This paper derives a dynamic conditional beta representation using a Bayesian semiparametric multivariate GARCH model. 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. We extend the model to several factors and find empirical support for a three factor model with nonlinear factor sensitives.

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

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

12.2 Derivation of the nonparametric conditional beta

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

$$\left(1 - \sum_{j=1}^{K^{(g)}} q_j^{(g)}(\mathbf{x}_t)\right)\frac{\int[\mu_1 + \beta_t(\mathbf{x}_t - \mu_2)]N(\mathbf{x}_t|\mu_2, (H_t^{(g)^{1/2}}BH_t^{(g)^{1/2'}})_{22})p(\mu, B)d\mu dB}{\int N(\mathbf{x}_t|\mu_2, (H_t^{(g)^{1/2}}BH_t^{(g)^{1/2'}})_{22})p(\mu, B)d\mu dB}.$$

Let

$$A_1 = \int [\mu_1 + \beta_t(\mathbf{x}_t - \mu_2)] N(\mathbf{x}_t|\mu_2, (H_t^{(g)^{1/2}} B H_t^{(g)^{1/2'}})_{22}) p(\mu, B) d\mu dB, \tag{12.2}$$

$$A_2 = \int N(\mathbf{x}_t|\mu_2, (H_t^{(g)^{1/2}} B H_t^{(g)^{1/2'}})_{22}) p(\mu, B) d\mu dB. \tag{12.3}$$

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

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

$$A_2 \approx \frac{1}{N} \sum_{n=1}^{N} N(\mathbf{x}_t|\mu_{n,2}, (H_t^{(g)^{1/2}} B_n H_t^{(g)^{1/2'}})_{22}) \tag{12.5}$$

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:

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

After replacing A_1 and A_2 with their approximations we have

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

where $\beta_{jt}^{(g)}$, $\beta_{nt}^{(g)}$, and $q_j^{(g)}(x_t^m)$ are defined in Equations (5.9), (12.6), and (5.6), respectively, and $N'(x|.)$ is the derivative of the pdf of Normal distribution with respect to x. In the case that we have more than one factor (say q factors), the derivations follow similarly but the derivative will be a vector of size q, each element of which is the coefficient of the corresponding 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. http://archive.nyu.edu/bitstream/2451/26575/2/FIN-01-029.pdf.

Dittmar, R. F. (2002), 'Nonlinear pricing kernels, kurtosis preference, and evidence from the cross section of equity returns', *The Journal of Finance* **57**(1), 369–403.

Engle, R. F. (2016), 'Dynamic conditional beta', *Journal of Financial Econometrics* **14**(4), 643–667.

Engle, R. F. & Kroner, K. F. (1995), 'Simultaneous Multivariate Generalized ARCH', *Econometric Theory* **11**(1), 122–150.

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.

Griffin, J. E. & Walker, S. G. (2011), 'Posterior simulation of normalized random measure mixtures', *Journal of Computational and Graphical Statistics* **20**, 241–259.

Guidolin, M. & Timmermann, A. (2008), 'International asset allocation under regime switching, skew, and kurtosis preferences', *The Review of Financial Studies* **21**(2), 889–935.

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

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.

Maheu, J. M., McCurdy, T. H. & Zhao, X. (2013), 'Do jumps contribute to the dynamics of the equity premium?', *Journal of Financial Economics* **110**(2), 457–477.

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 3059. http://www.diva-portal.org/smash/get/diva2:618567/FULLTEXT02.

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
BAC	0.031	10.701	0.891	23.399	35.261	-28.969

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

log-predictive likelihood					
Model	IBM	GE	XOM	AMGN	BAC
MGARCH-DPM	**−983.27**	**−964.99**	**−875.47**	**−1140.12**	**−1473.11**
BEKK-DPM	−1039.88	−1023.37	−951.14	−1226.54	−1526.39
MGARCH-t	−1353.67	−1369.03	−1300.21	−1571.32	−1684.72
log-Bayes factors					
MGARCH-DPM vs MGARCH-t	370.40	404.04	424.74	431.20	211.61
MGARCH-DPM vs BEKK-DPM	56.61	58.37	75.67	86.42	53.28

Table 2: This table reports the log-predictive likelihood for the bivariate MGARCH-DPM, BEKK-DPM and MGARCH-t and log-Bayes factors, for the last 500 observations, from 2012/03/12 to 2013/12/31. Bold entries denote the largest log-predictive likelihood for each asset. Bivariate data are daily excess market returns coupled with excess returns on IBM, GE, XOM, AMGN, and BAC 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).

BAC	MGARCH-DPM		MGARCH-t	
Parameter	Post. Mean	95% DI	Post. Mean	95% DI
γ_{01}	0.064	$(0.024, 0.115)$	0.065	$(0.045, 0.091)$
γ_{02}	-0.031	$(-0.062, 0.012)$	-0.022	$(-0.036, -0.008)$
γ_{03}	-0.007	$(-0.061, 0.042)$	0.029	$(0.004, 0.051)$
γ_{11}	0.284	$(0.235, 0.351)$	0.219	$(0.206, 0.238)$
γ_{12}	0.212	$(0.183, 0.260)$	0.213	$(0.197, 0.229)$
γ_{21}	0.962	$(0.954, 0.968)$	0.962	$(0.956, 0.966)$
γ_{22}	0.955	$(0.945, 0.963)$	0.963	$(0.956, 0.969)$
μ_1			0.000	$(-0.033, 0.033)$
μ_2			0.040	$(0.021, 0.072)$
ν			6.37	$(5.992, 6.839)$
c	6.24	$(3.000, 12.000)$		
α	0.658	$(0.092, 1.800)$		
η_1	0.436	$(0.326, 0.547)$	0.472	$(0.400, 0.554)$
η_2	0.414	$(0.333, 0.501)$	0.432	$(0.354, 0.558)$

Table 7: BAC 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 BAC and excess market returns. Data is from Jan 3, 2000 to Dec 31, 2013 (3521 observations).

	Log of the Conditional Predictive Likelihood		
Stock	1-factor model	3-factor model	4-factor model
IBM	−651.397	**−549.866**	−637.261
GE	−603.154	**−498.385**	−605.745
BAC	**−585.941**	−597.852	−602.425
XOM	−672.418	**−670.437**	−678.152
AMGN	−635.289	**−615.284**	−642.158

Table 8: This table reports the log of the conditional predictive likelihood for MGARCH-DPM model, for the last 500 observations, from 2012/03/12 to 2013/12/31. Bold entries denote the largest value in each row. Data are daily excess market, HML, SMB returns, and the momentum factor coupled with excess returns on IBM, BAC, XOM, GE, and AMGN from 2000/01/03 to 2013/12/31.

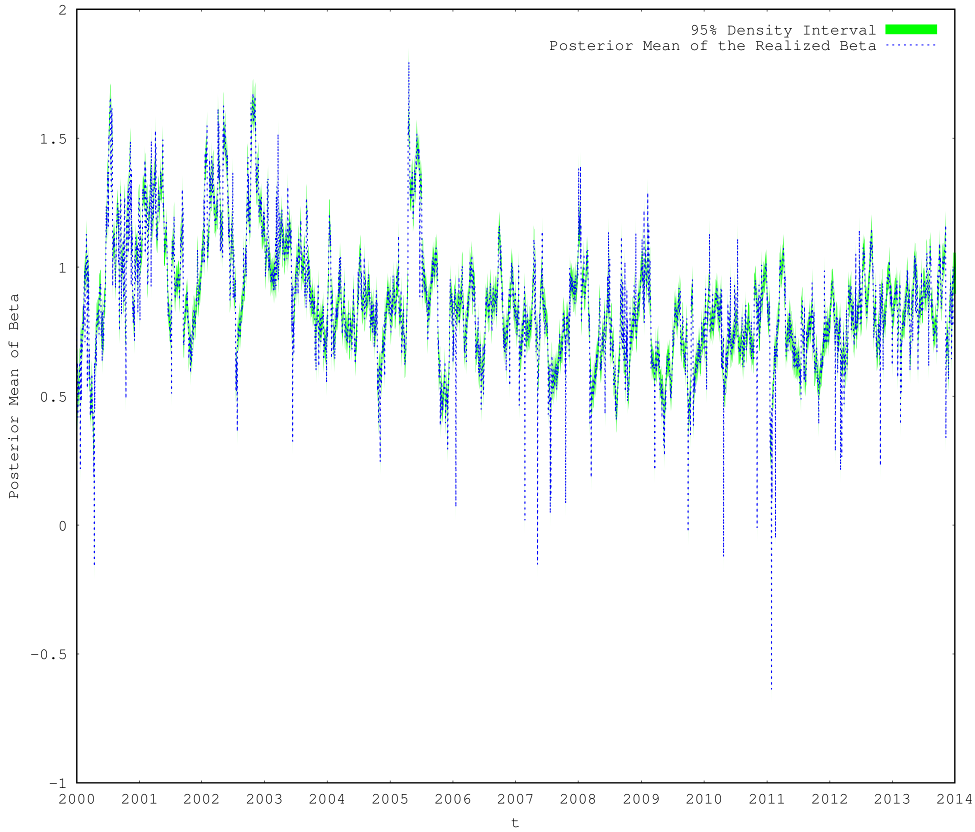

Figure 1: IBM: Realized conditional beta along with the 95% density intervals over time from MGARCH-DPM model.

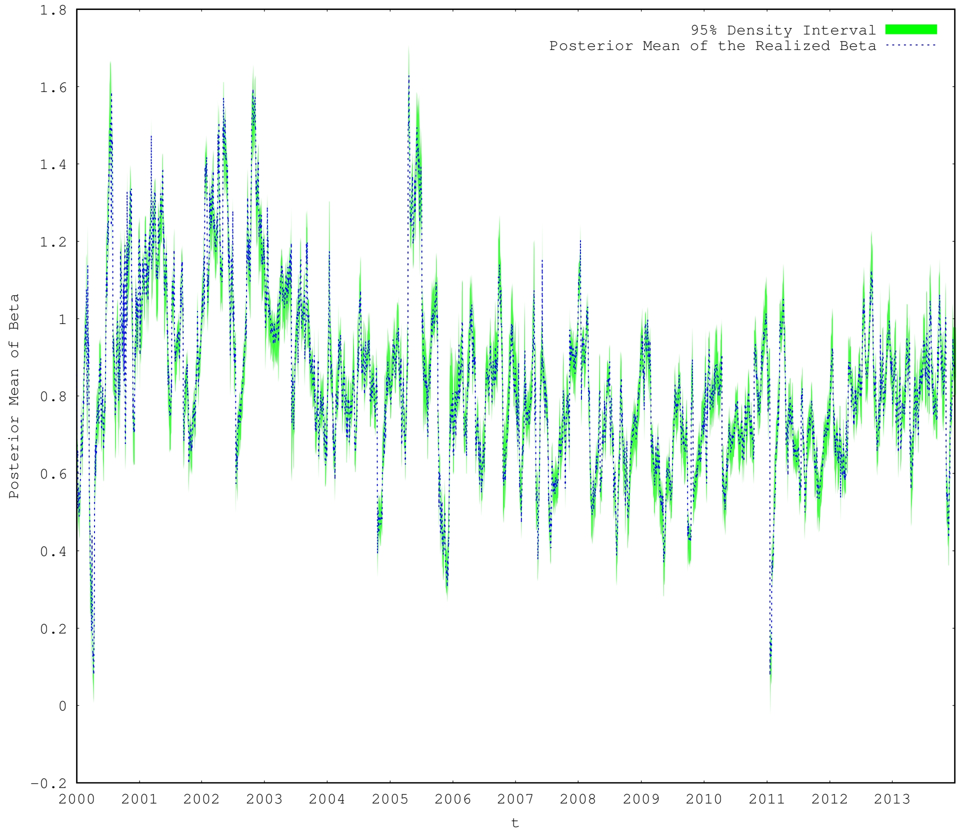

Figure 2: IBM: Realized conditional beta along with the 95% density intervals over time from MGARCH-t model.

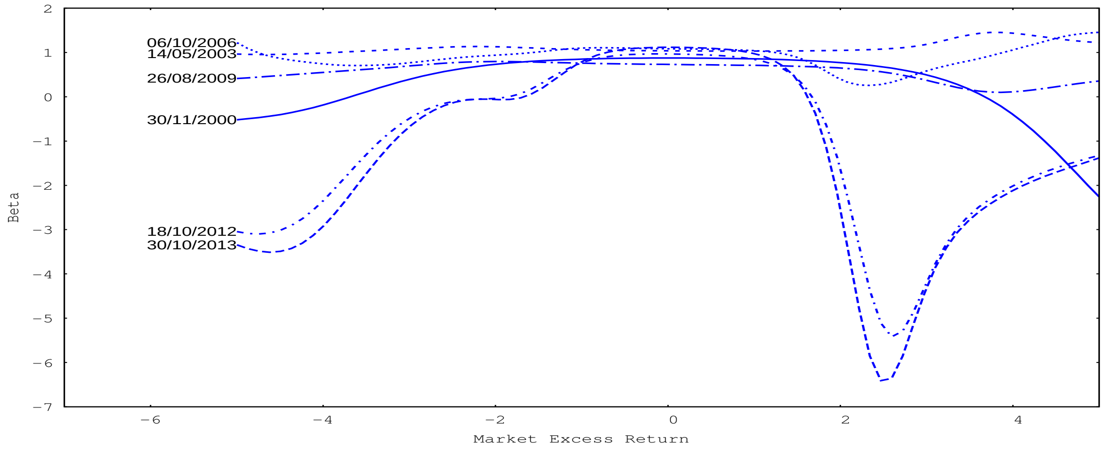

Figure 3: IBM: posterior mean of conditional beta as a function of the market excess return for different dates.

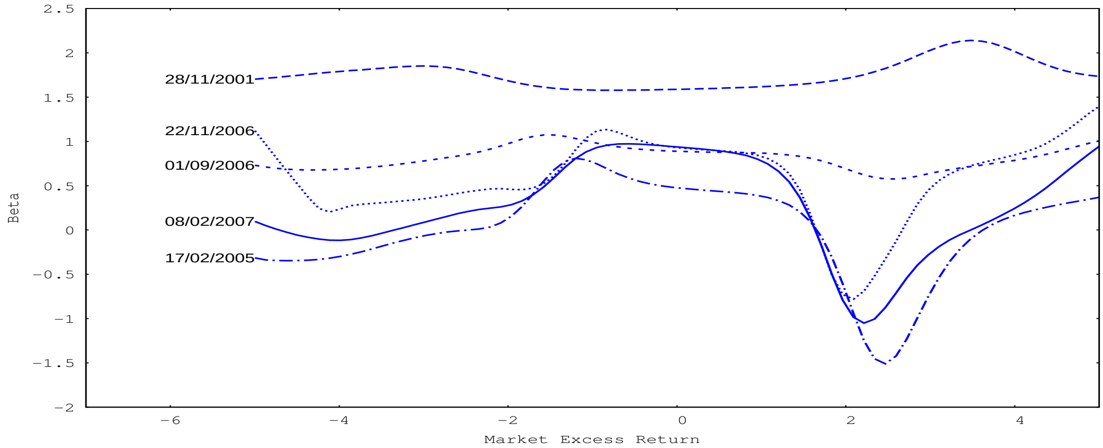

Figure 4: XOM: posterior mean of conditional beta as a function of the market excess return for different dates.

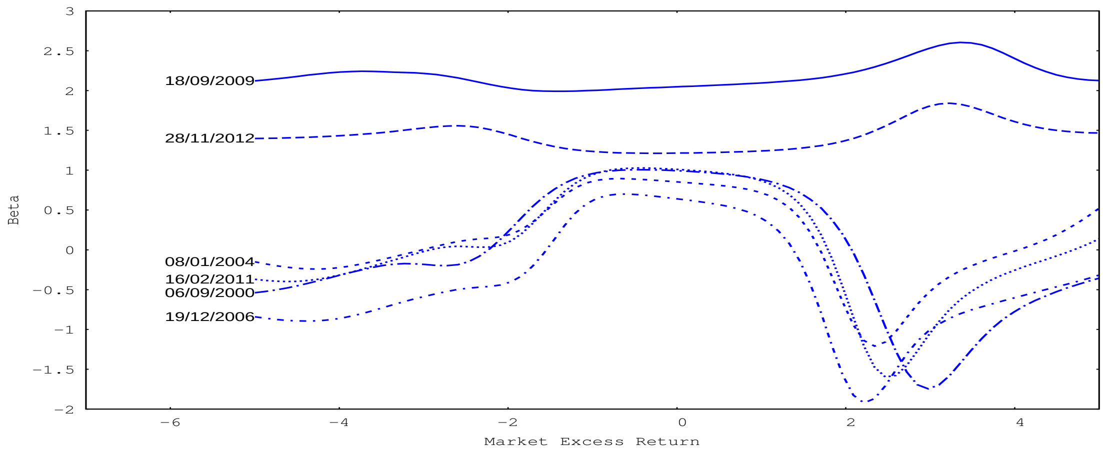

Figure 5: GE: posterior mean of conditional beta as a function of the market excess return for different dates.

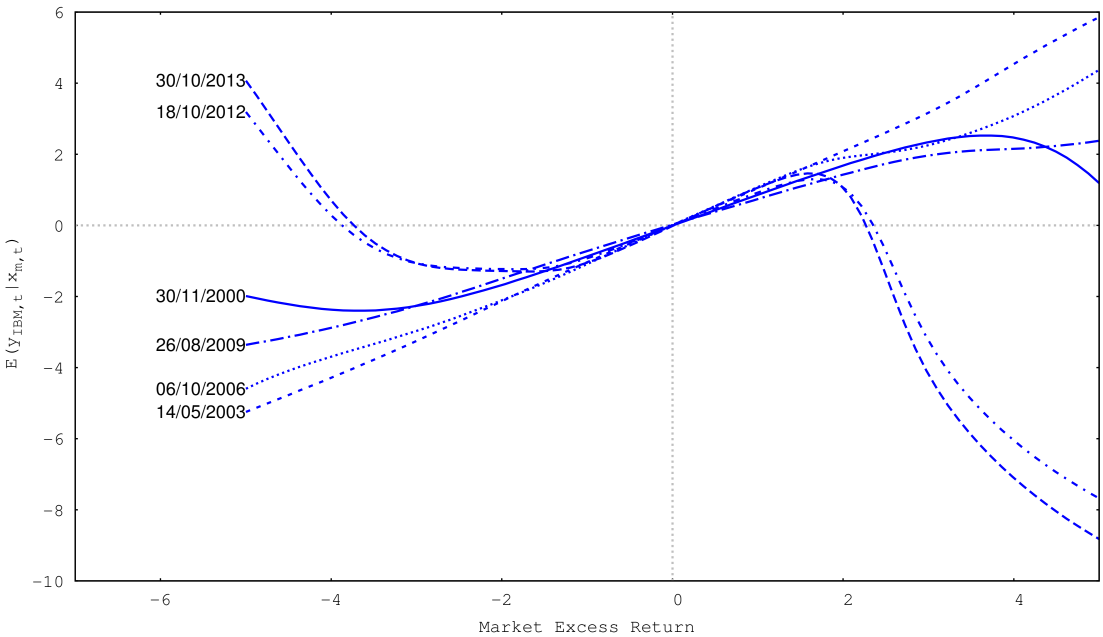

Figure 6: IBM: posterior mean of the conditional expected excess return of IBM given different values of the contemporaneous market excess return for different dates.

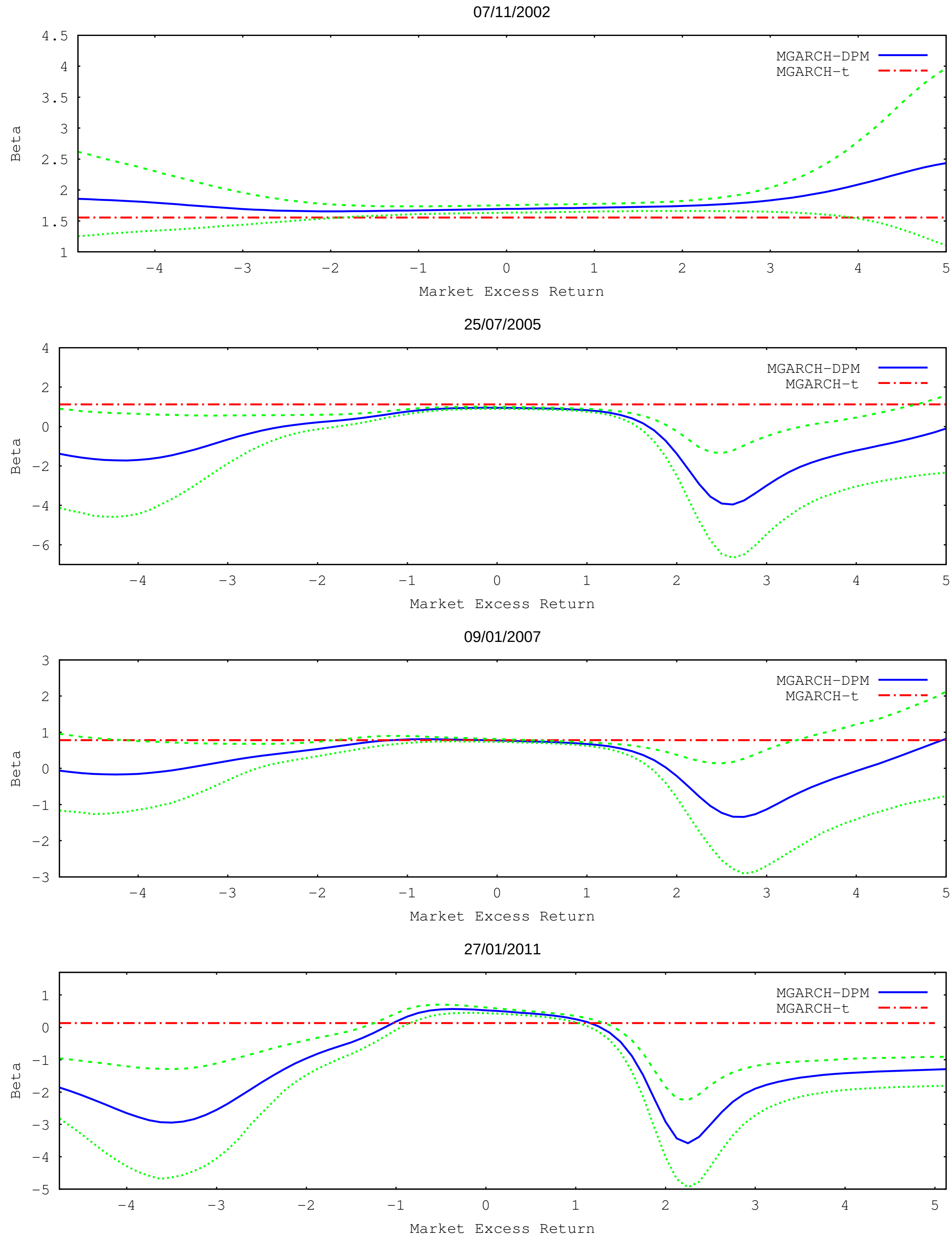

Figure 7: The posterior mean and 0.90 density intervals of IBM's conditional beta as a function of the excess market return from the MGARCH-DPM model. The red line shows the beta coefficients estimated with MGARCH-t model.

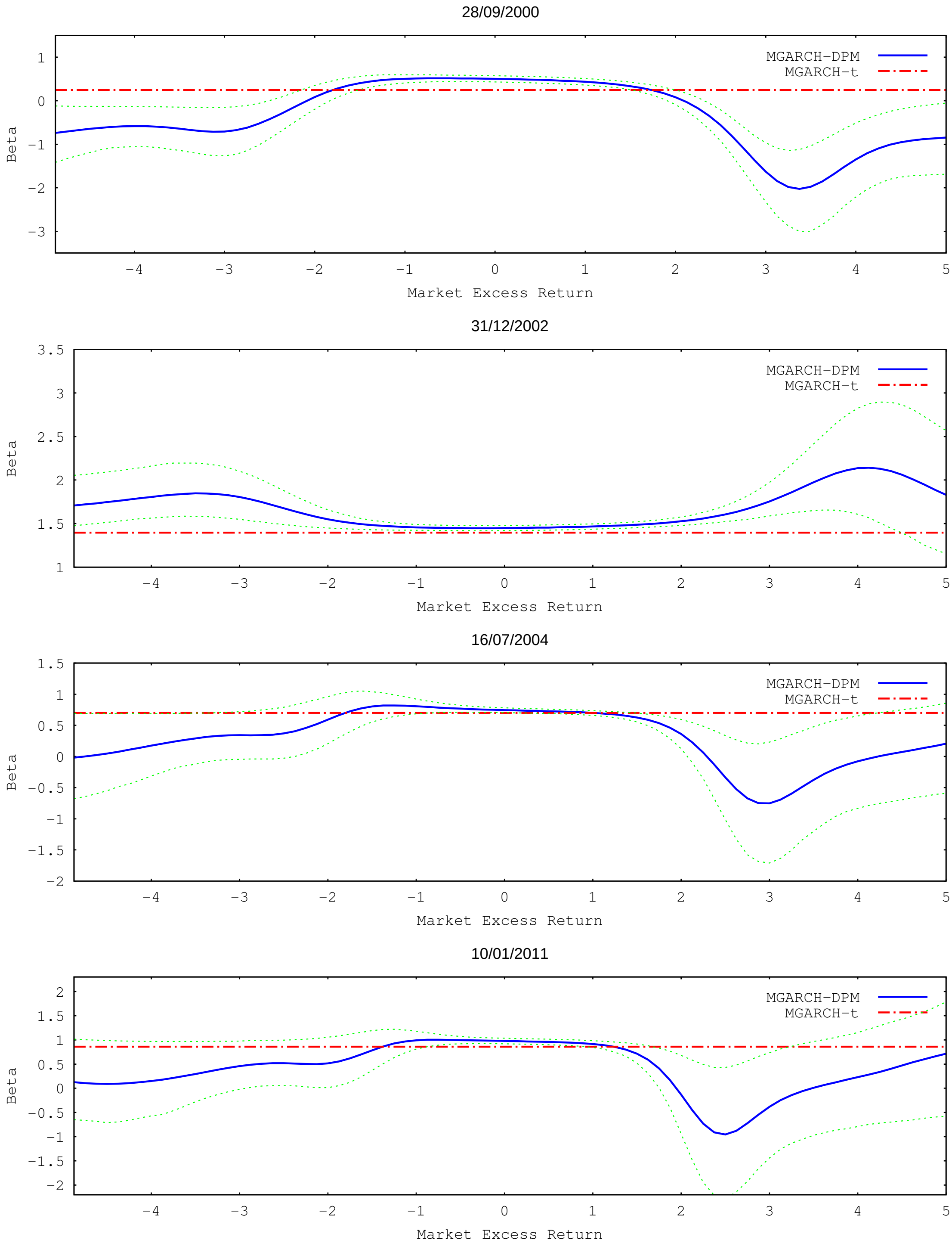

Figure 8: The posterior mean and 0.90 density intervals of GE's conditional beta as a function of the excess market return from the MGARCH-DPM model. The red line shows the beta coefficients estimated with MGARCH-t model.

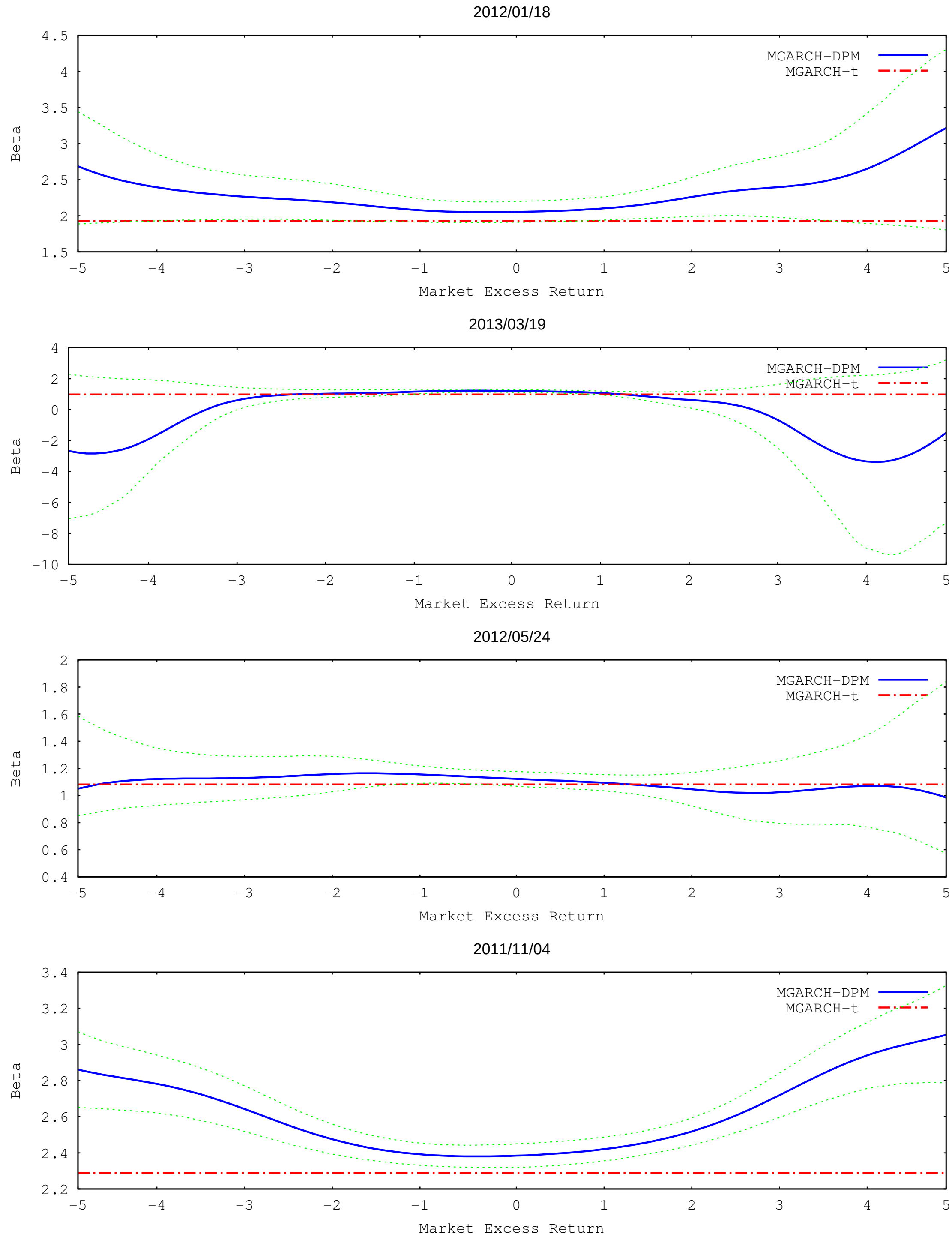

Figure 9: The posterior mean and 0.90 density intervals of BAC's conditional beta as a function of the excess market return from the MGARCH-DPM model. The red line shows the beta coefficients estimated with MGARCH-t model.

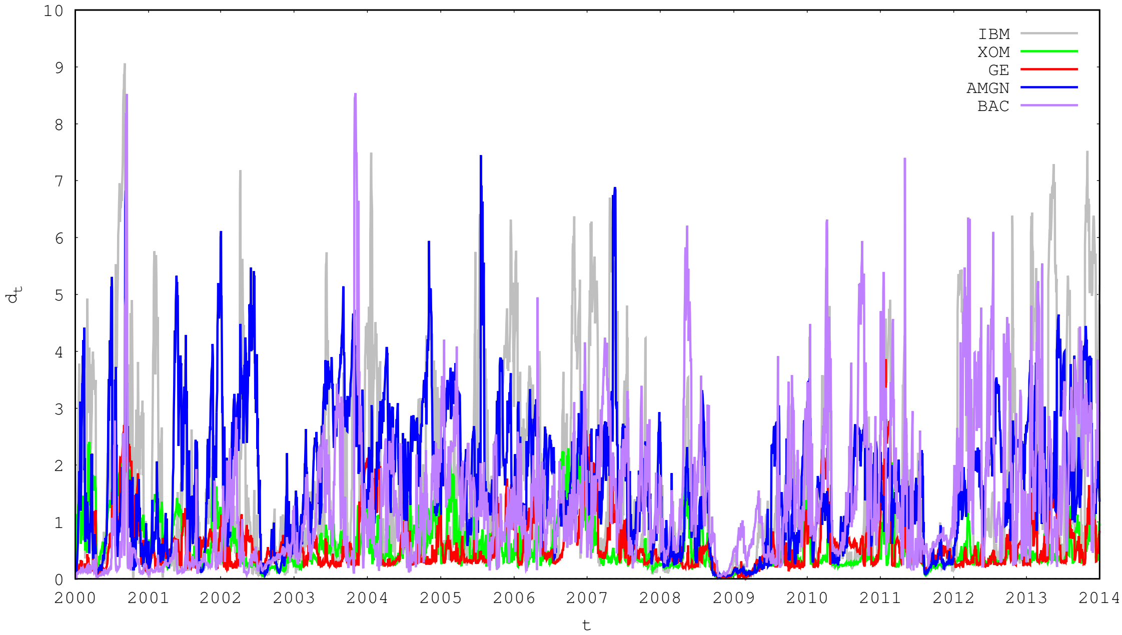

Figure 10: Variability of conditional beta with respect to the contemporaneous value of market excess returns over time for different stocks. $d_t = \max_{\mathbf{x}_t} b_t^m(\mathbf{x}_t) - \min_{\mathbf{x}_t} b_t^m(\mathbf{x}_t)$.

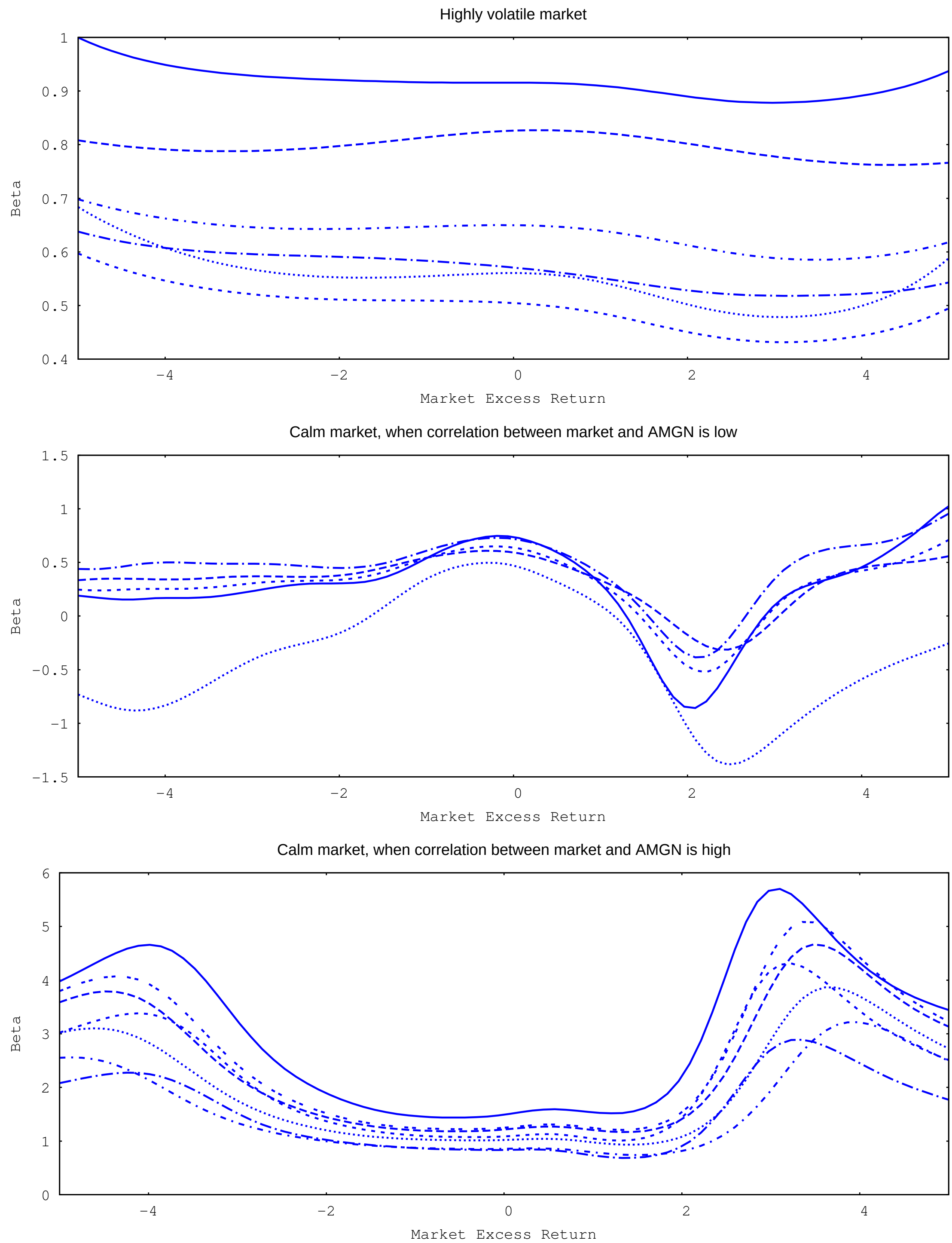

Figure 11: AMGN: conditional beta as a function of the market excess return for various dates grouped by market conditions and correlation.

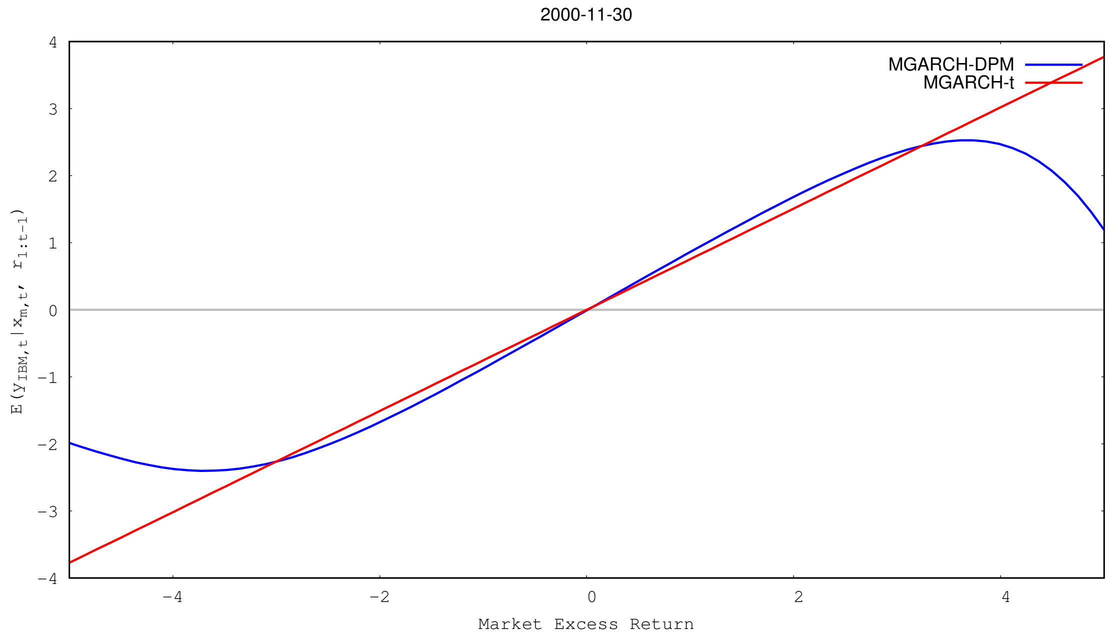

Figure 12: IBM: Predictive conditional expected return of IBM derived from MGARCH-DPM and MGARCH-t model.

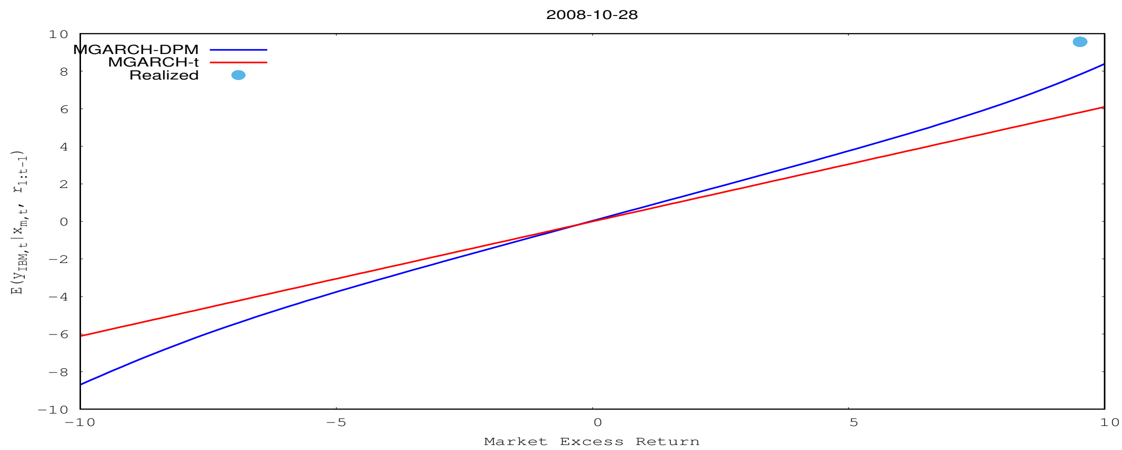

Figure 13: IBM: Predictive expected return from MGARCH-t and MGARCH-DPM models compared with the realized excess return of IBM when we expect a big shock to the market.

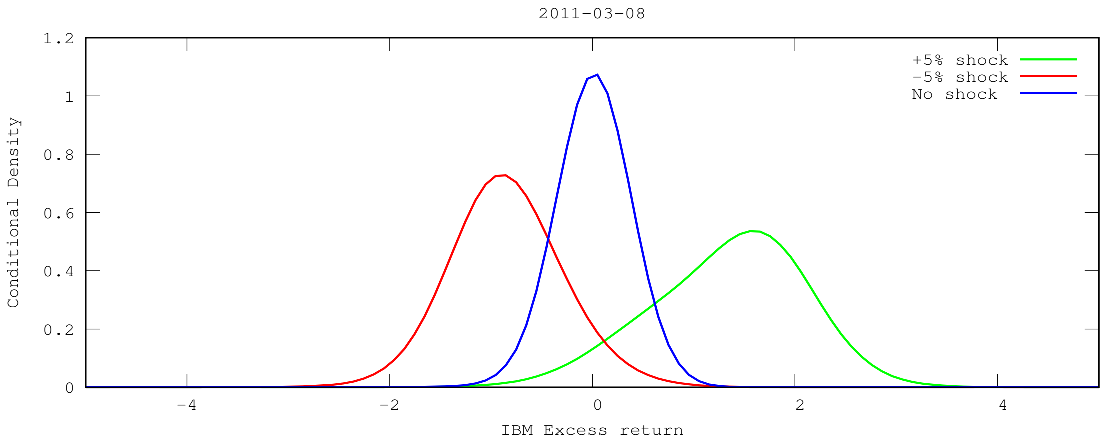

Figure 14: IBM: Effect of big shocks in the market return on IBM's predictive conditional density.

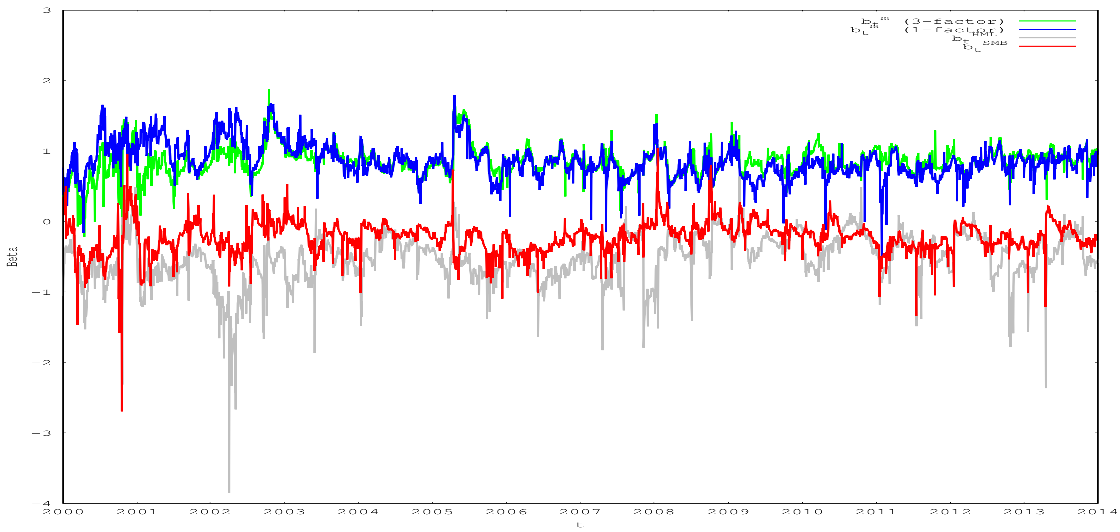

Figure 15: IBM: This figure illustrates the time series patterns of the posterior mean of b_t^m (3-factor), b_t^{SMB}, and b_t^{HML} estimated for the 3-factor MGARCH-DPM model using daily returns on IBM, the market, SMB, and HML. Beta on excess market returns is also compared with IBM's conditional beta derived using the 1-factor MGARCH-DPM model, b_t^m (1-factor).

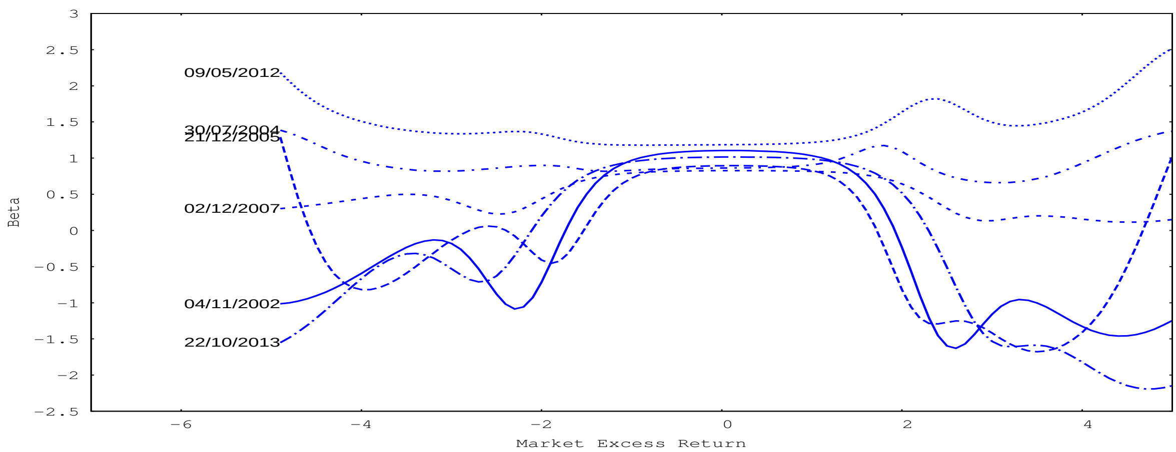

Figure 16: IBM: This figure shows the posterior mean of b_t^m, as a function of the excess market return estimated for the 3-factor MGARCH-DPM model. The size and value factors are set at their sample mean value. Data is daily returns on IBM, the market, SMB, and HML.

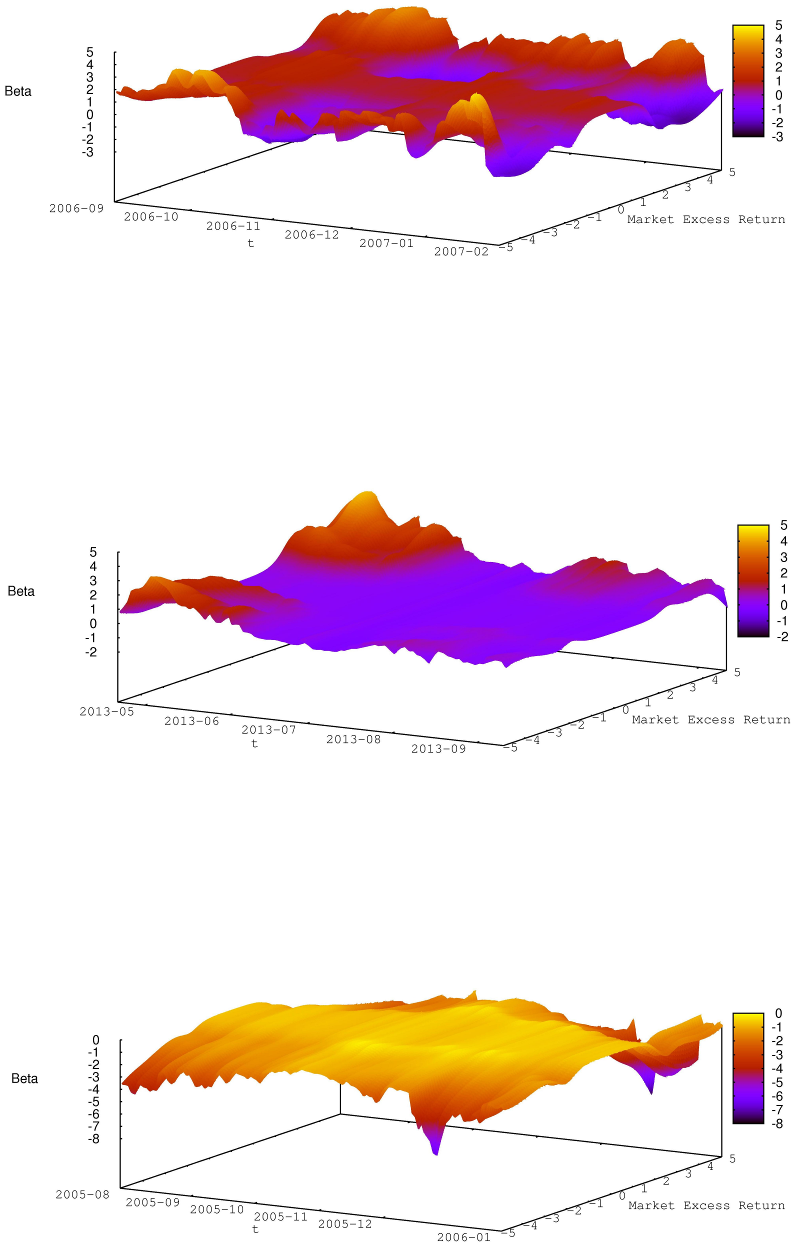

Figure 17: IBM: This figure shows the posterior mean of the b_t^m (top panel), b_t^{SMB} (middle panel), and b_t^{HML} (bottom panel) as functions of the excess market return from the 3-factor MGARCH-DPM model.