Modeling Covariance Breakdowns in Multivariate GARCH

Jin, Xin and Maheu, John M

Shanghai University of Finance and Economics, DeGroote School of Business, McMaster University

10 April 2014

Modeling Covariance Breakdowns in Multivariate GARCH*

Xin Jin† John M. Maheu‡

March 2014

Abstract

This paper proposes a flexible way of modeling dynamic heterogeneous covariance breakdowns in multivariate GARCH (MGARCH) models. During periods of normal market activity, volatility dynamics are governed by an MGARCH specification. A covariance breakdown is any significant temporary deviation of the conditional covariance matrix from its implied MGARCH dynamics. This is captured through a flexible stochastic component that allows for changes in the conditional variances, covariances and implied correlation coefficients. Different breakdown periods will have different impacts on the conditional covariance matrix and are estimated from the data. We propose an efficient Bayesian posterior sampling procedure for the estimation and show how to compute the marginal likelihood of the model. When applying the model to daily stock market and bond market data, we identify a number of different covariance breakdowns. Modeling covariance breakdowns leads to a significant improvement in the marginal likelihood and gains in portfolio choice.

JEL: C32, C51, C58
key words: correlation breakdown, marginal likelihood, particle filter, Markov chain, generalized variance

*We thank Jeroen Rombouts for comments on an earlier version of this paper. We thank Jia Liu and conference participants at CFE'13 for helpful comments. Maheu is grateful to the SSHRC for financial support.

†Shanghai University of Finance and Economics, jin.xin@mail.shufe.edu.cn

‡DeGroote School of Business, McMaster University, 1280 Main Street West, Hamilton, ON, Canada, L8S4M4 and University of Toronto, Canada and RCEA, Italy, maheujm@mcmaster.ca

1 Introduction

This paper proposes a flexible way of modeling dynamic heterogeneous covariance breakdowns in multivariate GARCH (MGARCH) models. During periods of normal market activity, volatility dynamics are governed by an MGARCH specification. A covariance breakdown is any significant temporary deviation of the conditional covariance matrix from its implied MGARCH dynamics. A covariance breakdown is captured through a flexible stochastic component that allows for changes in the conditional variances, covariances and implied correlation coefficients.

It is widely acknowledged that markets face periods that are characterized by abnormal behavior. Several approaches have been used to capture changes in the dynamics of conditional second moments including dynamic copulas (Kenourgios et al. 2011, Christoffersen et al. 2012), and the factor spline GARCH model of Rangel & Engle (2012).[1] Dufays (2013) uses an infinite-state hidden Markov model to allow for parameter change in Engle's (2002) dynamic conditional correlation model. The path dependence that the latent state variable causes in the GARCH recursions is removed following the ideas in Klaassen (2002).[2] Haas & Mittnik (2008) and Chen (2009) extend the univariate MS-GARCH model in Haas et al. (2004) to a multivariate setting. Their model assumes there are K parallel MGARCH models running at the same time, where K is the number of states. Silvennoinen & Tersvirta (2009) apply the smooth transition modeling approach to conditional correlations. Other regime-switching approaches include Ang & Bekaert (2004), Guidolin & Timmermann (2006) and Pelletier (2006).

In contrast to the literature, which has tended to focus on correlation breakdowns, we investigate breakdowns in each component of the conditional covariance matrix. This has several advantages. First, we can see how conditional correlations are affected through variances and covariances. Second, by modeling the full covariance matrix we avoid issues of misspecification by focusing only on correlations (Forbes & Rigobon 2002) and neglecting heteroskedasticity. In our model a covariance breakdown does not necessarily imply a correlation breakdown or contagion effect. It depends on the relative changes in the conditional covariance and conditional variances. Empirically we identify both covariance breakdowns which lead to correlation changes and breakdowns which have little impact on correlations.

To our knowledge this is the first paper to explicitly model the dynamics of conditional covariance breakdowns and estimate their impacts. In our approach a covariance breakdown is any sustained deviation of the conditional second moments from the covariance matrix implied by the MGARCH specification. Each breakdown period is different and estimated from the data. Covariance breakdowns as well as normal periods are assumed to follow a first-order Markov chain. Each breakdown is characterized by a random matrix drawn from an inverse-Wishart distribution that scales (multiplies) the MGARCH covariance matrix.[3]

[1] It is important to account for changes in GARCH dynamics. For instance, in the univariate setting, neglected parameter changes in volatility dynamics can bias GARCH parameter estimates toward higher persistence and lead to poor forecasts of volatility (Lamoureux & Lestrapes 1990, Hillebrand 2005).

[2] Another approach to dealing with path dependence directly is the particle MCMC approach of Bauwens et al. (2014).

[3] To be precise, a positive definite matrix drawn from an inverse-Wishart density is sandwiched between the Cholesky decomposition of the MGARCH matrix.

This approach is very flexible while retaining a positive definite matrix. Since covariance breakdowns are finite, they eventually end and we return to a model in which the MGARCH dynamics solely determine conditional second moments.

Our model can be considered as an extension to Markov switching models. Bayesian inference for Markov regime-switching models is usually carried out based on the forward-backward algorithm of Chib (1996). Our approach is different than the conventional regime-switching specification in which model parameters governing a time period are selected from a fixed parameter set. A covariance breakdown is captured by introducing an exogenous stochastic multiplicative component to the volatility matrix itself. This requires a new posterior sampling approach for the states. We construct an efficient sampling scheme to sample the unobserved state variables as well as other fixed parameters.

Whether covariance breakdowns are supported by the data can be formally assessed in the context of Bayesian model comparison by making use of the marginal likelihood. We show how to compute the marginal likelihood and design a particle filter for the task.

The model is applied to daily excess returns on the S&P 500 index and short-term and long-term bonds over a twenty-five year period. Including fat-tailed return innovations in the model is important in distinguishing between outliers and sustained covariance breakdowns. We compare our model to an MGARCH model with Student-t innovations but with no breakdowns. Bayes factors strongly support the inclusion of covariance breakdowns. The volatility dynamics during breakdown periods are very different for the two models as well as breakdowns being different over time. For example, in the recent financial crisis we identify an initial breakdown which leads to an overall increase in variability. This features large increases in conditional variances and drops in covariances between the stock and bond market. However, the conditional correlations do not show a dramatic change. Following this episode is another breakdown which is characterized as a reduction in overall variability.

Estimates indicate that covariance breakdowns occur 34% of the time and their expected duration is 1-2 months. The impact of a typical covariance breakdown is expected to increase variability. In addition to improving the fit of the data, modeling covariance breakdowns provides improved portfolio choice.

The rest of the paper is organized as follows. In Section 2, we introduce the breakdown model and discuss its properties. Section 3 constructs a sampling procedure for the posterior inference of the model. Section 4 provides simulation study for illustration. Section 5 shows how to compute the marginal likelihood of our model. In Section 6, we apply the model to study the volatility dynamics among the stock market and the bond market and Section 7 concludes. The Appendix contains details on posterior sampling and computation of the marginal likelihood.

2 Multivariate GARCH with Covariance Breakdowns

Consider a k-dimensional vector time series y_t, $t = 1, 2, 3, \ldots$. Let $\mathcal{F}_{t-1}$ be the sigma field generated by the past values of y_t until time $t-1$. Consider the following model

$$y_t = \mu + H_t^{1/2} \Lambda_t^{1/2} z_t, \tag{1}$$

where $\mu = \mathbb{E}(y_t|\mathcal{F}_{t-1})$ is the constant conditional mean[4] vector and $z_t \sim NID(0, I)$.[5] $H_t^{1/2}$ denotes the Cholesky factor of the $k \times k$ positive definite matrix H_t, which is assumed to follow any of the popular specifications for the MGARCH model. Popular examples of MGARCH models include, among others, the vector-diagonal GARCH (VDGARCH) by Ding & Engle (2001) and the dynamic conditional correlation (DCC) by Engle (2002). See Bauwens et al. (2006) for a review.

The latent discrete state variable $s_t \in \{1, 2\}$ is assumed to direct the dynamics of Λ_t and follows a two-state first-order Markov chain with transition matrix

$$P = \begin{pmatrix} \pi_1 & 1-\pi_1 \\ 1-\pi_2 & \pi_2 \end{pmatrix}, \tag{2}$$

which divides the sample path into periods of *normal* states ($s_t = 1$) and periods of *covariance breakdown* states ($s_t = 2$). We impose the constraint $\pi_1 > \pi_2$ to identify normal periods as being more frequent than covariance breakdowns.

If $s_{1:t} = \{s_1, ..., s_t\}$, then Λ_t is determined as follows

$$\Lambda_t|s_{1:t} = \begin{cases} I & \text{if } s_t = 1 \\ \Lambda_{t-1} & \text{if } s_t = 2, s_{t-1} = 2 \\ \sim G_0 & \text{if } s_t = 2, s_{t-1} = 1 \end{cases} \tag{3}$$

G_0 can be any distribution over symmetric positive definite matrices. In this paper, we use an inverse-Wishart distribution $W^{-1}(\nu, Q_0)$, with first moment $Q_0/(\nu - k - 1)$, where $\nu > k - 1$ is the scalar degree of freedom and Q_0 is the $k \times k$ symmetric positive definite scale matrix. The choice of inverse-Wishart distribution aids in posterior sampling and will be discussed in more detail in Section 3.

Some key features of the model are as follows. First, there remains an MGARCH structure in the volatility dynamics that is represented by H_t. However, with the addition of Λ_t in the structure, H_t is no longer the conditional covariance matrix of y_t as is the case with the conventional MGARCH models since $\text{Var}(y_t|\Lambda_t, \mathcal{F}_{t-1}) = H_t^{1/2}\Lambda_t(H_t^{1/2})'$. Second, s_t determines the underlying state of the volatility dynamics of y_t. $s_t = 1$, $\Lambda_t = I$ denotes the *normal state* in which volatility dynamics are solely driven by H_t from the MGARCH structure, where $\text{Var}(y_t|\Lambda_t = I, \mathcal{F}_{t-1}) = H_t$. A change from the underlying MGARCH covariance dynamics begins when the state switches from $s_{t-1} = 1$ to $s_t = 2$, which signals the beginning of a *breakdown period*. In this case, Λ_t is a new draw from G_0 and $\text{Var}(y_t|\Lambda_t, \mathcal{F}_{t-1}) = H_t^{1/2}\Lambda_t(H_t^{1/2})'$ in general will not be equal to H_t. As a result, the volatility starts to deviate from H_t. If $s_{t+1} = 2$ the breakdown period continues and $\Lambda_{t+1} = \Lambda_t$, while the MGARCH component changes to H_{t+1}. In this way we isolate the tendency of a covariance breakdown period to display a constant impact on H_t such as driving up conditional correlations or depressing conditional variances. The breakdown continues until the state switches back to 1, and we are back in the normal state, where the volatility once again coincides with H_t.

In each breakdown period Λ_t is constant and has the same impact on H_t. On the other hand, different breakdown periods will have their unique break patterns, e.g. increased variability or decreased variability, etc., as each breakdown period is characterized by a unique

[4]A time-varying conditional mean can also be used.

[5]Other distributions such as a Student-t could be used for z_t as long as a normal decomposition can be admitted.

Λ_t drawn from G_0. Thus even though s_t is a two-state Markov chain and the volatility dynamics can only alternate between normal periods and breakdown periods, the breakdown periods are "heterogeneous" and there could be infinitely many "types" of covariance breakdowns. Hence this modeling structure distinguishes itself from a standard two-state regime-switching approach which moves between two fixed parameter vectors. The model is also different from the (infinite dimension) structural break model. In structural break models, past states (periods) cannot recur. Whereas in our model, the volatility dynamics can always revert to the normal state $s_t = 1$.

The functional form for covariance breakdowns $H_t^{1/2}\Lambda_t(H_t^{1/2})'$ is convenient in that it remains positive definite and a well defined conditional covariance matrix. In general, Λ_t can inflate or shrink the elements of H_t. One overall measure of variability is the determinant of the covariance matrix, also called the *generalized variance* by Muirhead (1982). The relative change in the generalized variance is $|H_t^{1/2}\Lambda_t(H_t^{1/2})'|/|H_t| = |\Lambda_t|$. Λ_t can either increase or decrease the generalized variance depending on whether its own determinant is greater or less than 1.

In addition to the direct effect on the covariance matrix, Λ_t also has an indirect effect on future volatility through the feedback from the current volatility matrix to future H_t. Any of the usual MGARCH specifications for H_{t+1} depends on the cross-products of y_t, the distribution of which is a function of $H_t^{1/2}\Lambda_t(H_t^{1/2})'$. If $\Lambda_t \neq I$, it will affect H_{t+1} through the realized value of y_t. As a result, during a breakdown period, the instantaneous impact of Λ_t on the conditional second moments is compounded over time with the help of the feedback channel. An important implication is that the model may accommodate breakdown periods that feature local nonstationarity.

This model is flexible in allowing for various changes in the conditional covariance matrix. Consider the following example. Let

$$H = \begin{pmatrix} 2 & 1.5 \\ 1.5 & 3 \end{pmatrix}.$$

We draw several random Λ from $G_0 = W^{-1}(7, 4I)$ and compute $V = H^{1/2}\Lambda(H^{1/2})'$.[6] Table 1 records the changes in the elements as well as the implied correlation coefficient of V relative to those of H for different Λ. Various scenarios can occur. For example, we have cases where the variances, the covariance, and the correlation coefficient all increase (or decrease), and we have cases where the variances and the covariance increase (or decrease) in a way that the correlation coefficient remains relatively constant. There can be times when the variances go up but the covariance goes down and even becomes negative. Other times the two variances can move in opposite directions. This particular form for covariance breakdowns allows for a wide variety of effects on the conditional second moments that can be estimated from the data.

In another example, assume the following VDGARCH specification

$$H_t = CC' + aa' \odot y_{t-1}y_{t-1}' + bb' \odot H_{t-1} \tag{4}$$

where C is a $k \times k$ lower triangular matrix, a and b are $k \times 1$, $\odot$ denotes the element-by-element matrix product (Hadamard product). Let G_0 be the same distribution as above.

[6] $\mathbb{E}(\Lambda) = \frac{4I}{7-2-1} = I$.

Three episodes of the simulated data are shown in Figure 1 to 3 displaying the data, states and elements of the total conditional covariance matrix $V_t = (v_{ij,t}) = H_t^{1/2}\Lambda_t(H_t^{1/2})')$ as well as the corresponding values from $H_t = (h_{ij,t})$. All three episodes have one breakdown period, but the break patterns are different from one another. In the first episode, the breakdown period features a sharp surge in both conditional variances and the conditional covariance. The correlation coefficient also jumps up from below 0.3 to around 0.9. The associated $\Lambda = (\lambda_{ij})$ [7] for this breakdown period is $\lambda_{11} = 6.62$, $\lambda_{21} = 1.86$, $\lambda_{22} = 1.46$. Its determinant is $|\Lambda| = 6.20$, indicating an increase in the overall variability. In the second episode, the breakdown period causes both variances to increase, while inducing a strong drop in the correlation between $y_{1,t}$ and $y_{2,t}$. Here $\lambda_{11} = 1.35$, $\lambda_{21} = -0.98$, $\lambda_{22} = 1.47$, and $|\Lambda| = 1.02$, which means no change in the overall variability. For the third episode, the two variances move in opposite directions during the breakdown period and the correlation switches signs as the breakdown period starts and becomes very negative until the normal period is restored, by which time the correlation becomes positive again.

In summary, this modeling approach is flexible enough to accommodate different abrupt changes that lead to significant deviations from the underlying MGARCH structure.

3 Model Inference

We apply a Bayesian approach to model estimation where Markov chain Monte Carlo (MCMC) methods are used for posterior inference. The unknown parameters of the model consist of $\pi_1, \pi_2, \Theta, \mu$, where Θ denotes the set of parameters in the GARCH specification that governs H_t. For example, if the VDGARCH specification in (4) is used, then $\Theta = \{C, a, b\}$. We refer to this model with covariance breakdowns as VDGARCH-B. The parameter space is augmented with s_t and Λ_t, both of which are jointly estimated with the fixed parameters.

The stochastic nature of Λ_t makes the posterior sampling of states s_t more complicated than a standard Markov switching model in which state-dependent parameters are fixed over the whole sample. Given the observed data of sample size T, let $\mathbf{S} = \{s_t\}_{t=1}^T$, $\mathbf{Y} = \{y_t\}_{t=1}^T$, $\mathbf{\Lambda} = \{\Lambda_t\}_{t=1}^T$, a hybrid MCMC algorithm can be designed to sample from the joint posterior distribution. To sample from $p(\pi_1, \pi_2, \Theta, \mu, \mathbf{S}, \mathbf{\Lambda}|\mathbf{Y})$, we propose an efficient sampler that iteratively samples through the conditional posterior distribution of each block:

1. $p(\mathbf{S}|\Theta, \pi_1, \pi_2, \mu, \mathbf{Y})$
2. $p(\pi_1, \pi_2|\mathbf{S})$,
3. $p(\Theta|\mathbf{S}, \mu, \mathbf{Y})$
4. $p(\mathbf{\Lambda}|\mathbf{S}, \Theta, \mu, \mathbf{Y})$
5. $p(\mu|\mathbf{\Lambda}, \Theta, \mathbf{Y})$.

Taking a draw from all of the conditional distributions constitutes one sweep of the sampler. After dropping an initial set of draws as burn-in we collect M draws $\{\pi_1^{(i)}, \pi_2^{(i)}, \Theta^{(i)}, \mu^{(i)}, \mathbf{S}^{(i)}, \mathbf{\Lambda}^{(i)}\}_{i=1}^M$ for posterior inference. Simulation consistent estimates of posterior moments can be obtained as sample averages of the draws. For instance, the posterior mean of π_1 can be estimated as $M^{-1}\sum_{i=1}^M \pi_1^{(i)}$. Next we discuss each block in more detail.

[7] The subscript t of Λ_t is suppressed.

3.1 Sampling S

Conditional on H_t and μ, apply the transformation

$$\tilde{y}_t = H_t^{-1/2}(y_t - \mu), \tag{5}$$

so that

$$\tilde{y}_t|\Lambda_t \sim N(0, \Lambda_t). \tag{6}$$

Let $\tilde{\mathbf{Y}} = \{\tilde{y}_t\}_{t=1}^T$ be the transformed data. Sampling from $p(\mathbf{S}|\Theta, \pi_1, \pi_2, \mu, \mathbf{Y})$ is equivalent to sampling from $p(\mathbf{S}|\pi_1, \pi_2, \tilde{\mathbf{Y}})$.

We sequentially sample from the two-point discrete distributions $p(s_t|\mathbf{S}_{-t}, \pi_1, \pi_2, \tilde{\mathbf{Y}})$ for $t = 1, \ldots, T$, where $\mathbf{S}_{-t} = \mathbf{S}\backslash s_t$. It requires calculating $Pr(s_t = 1|\mathbf{S}_{-t}, \pi_1, \pi_2, \tilde{\mathbf{Y}})$ and $Pr(s_t = 2|\mathbf{S}_{-t}, \pi_1, \pi_2, \tilde{\mathbf{Y}})$ for each t. Let t_1, t_2 be integers such that $T \geq t_1 \geq 1$, $T \geq t_2 \geq 1$. There are several cases depending on the values of $\mathbf{S}_{-t}$. Suppress π_1, π_2 in the conditioning set for the moment and write $p(\tilde{y}_t|\Lambda) = N(\tilde{y}_t|0, \Lambda)$, $p(\Lambda) = W^{-1}(\Lambda|\nu, Q_0)$, where $N(.|.,.)$ and $W^{-1}(.|.,.)$ are the density functions of the normal distribution and the inverse-Wishart distribution, respectively. The cases are:

- $s_{t-t_1-1} = 1, s_{t-t_1} = \cdots = s_{t-1} = 2, s_{t+1} = \cdots = s_{t+t_2} = 2, s_{t+t_2+1} = 1$

$$\begin{aligned} Pr(s_t = 1|\mathbf{S}_{-t}, \tilde{\mathbf{Y}}) &\propto \left[\int \left(\prod_{j=-t_1}^{-1} p(\tilde{y}_{t+j}|\Lambda)\right) p(\Lambda) d\Lambda\right] p(\tilde{y}_t|s_t = 1) \left[\int \left(\prod_{j=1}^{t_2} p(\tilde{y}_{t+j}|\Lambda)\right) p(\Lambda) d\Lambda\right] \\ &\quad \times Pr(s_t = 1|s_{t-1} = 2) Pr(s_{t+1} = 2|s_t = 1) \\ Pr(s_t = 2|\mathbf{S}_{-t}, \tilde{\mathbf{Y}}) &\propto \left[\int \left(\prod_{j=-t_1}^{t_2} p(\tilde{y}_{t+j}|\Lambda)\right) p(\Lambda) d\Lambda\right] Pr(s_t = 2|s_{t-1} = 2) Pr(s_{t+1} = 2|s_t = 2). \end{aligned}$$

- $s_{t-1} = 1, s_{t+1} = \cdots = s_{t+t_2} = 2, s_{t+t_2+1} = 1$

$$\begin{aligned} Pr(s_t = 1|\mathbf{S}_{-t}, \tilde{\mathbf{Y}}) &\propto p(\tilde{y}_t|s_t = 1) \left[\int \left(\prod_{j=1}^{t_2} p(\tilde{y}_{t+j}|\Lambda)\right) p(\Lambda) d\Lambda\right] \\ &\quad \times Pr(s_t = 1|s_{t-1} = 1) Pr(s_{t+1} = 2|s_t = 1) \\ Pr(s_t = 2|\mathbf{S}_{-t}, \tilde{\mathbf{Y}}) &\propto \left[\int \left(\prod_{j=0}^{t_2} p(\tilde{y}_{t+j}|\Lambda)\right) p(\Lambda) d\Lambda\right] Pr(s_t = 2|s_{t-1} = 1) Pr(s_{t+1} = 2|s_t = 2). \end{aligned}$$

- $s_{t-t_1-1} = 1, s_{t-t_1} = 2 = \cdots = s_{t-1} = 2, s_{t+1} = 1$

$$\begin{aligned} Pr(s_t = 1|\mathbf{S}_{-t}, \tilde{\mathbf{Y}}) &\propto \left[\int \left(\prod_{j=-t_1}^{-1} p(\tilde{y}_{t+j}|\Lambda)\right) p(\Lambda) d\Lambda\right] p(\tilde{y}_t|s_t = 1) \\ &\quad \times Pr(s_t = 1|s_{t-1} = 2) Pr(s_{t+1} = 1|s_t = 1) \\ Pr(s_t = 2|\mathbf{S}_{-t}, \tilde{\mathbf{Y}}) &\propto \left[\int \left(\prod_{j=-t_1}^{0} p(\tilde{y}_{t+j}|\Lambda)\right) p(\Lambda) d\Lambda\right] Pr(s_t = 2|s_{t-1} = 2) Pr(s_{t+1} = 1|s_t = 2). \end{aligned}$$

- $s_{t-1}=1, s_{t+1}=1$

$$\begin{aligned} Pr(s_t=1|\mathbf{S}_{-t},\tilde{\mathbf{Y}}) &\propto p(\tilde{y}_t|s_t=1)Pr(s_t=1|s_{t-1}=1)Pr(s_{t+1}=1|s_t=1) \\ Pr(s_t=2|\mathbf{S}_{-t},\tilde{\mathbf{Y}}) &\propto \left[\int p(\tilde{y}_t|\Lambda)p(\Lambda)d\Lambda\right] Pr(s_t=2|s_{t-1}=1)Pr(s_{t+1}=1|s_t=2). \end{aligned}$$

In all cases,[8] we need to compute the integral $\int\left(\prod_{t=t_3}^{t_4} p(\tilde{y}_t|\Lambda)\right)p(\Lambda)d\Lambda$ for some t_3 and t_4 with $T\geq t_4\geq t_3\geq 0$. We show in the Appendix that

$$\int\left(\prod_{t=t_3}^{t_4} p(\tilde{y}_t|\Lambda)\right)p(\Lambda)d\Lambda = \frac{2^{\frac{nk}{2}}(2\pi)^{-\frac{nk}{2}}|Q_0|^{\frac{\nu}{2}}\prod_{j=1}^{k}\Gamma(\frac{n+\nu+1-j}{2})}{|Q|^{\frac{n+\nu}{2}}\prod_{j=1}^{k}\Gamma(\frac{\nu+1-j}{2})}, \tag{7}$$

where $Q=\sum_{t=t_3}^{t_4}\tilde{y}_t\tilde{y}_t'+Q_0$, and $n=t_4-t_3+1$.

3.2 Sampling π_1, π_2

The conditional posterior distributions are:

$$p(\pi_1|\mathbf{S})\propto p(\mathbf{S}|\pi_1)p(\pi_1)\propto p(\pi_1)\pi_1^{n_1}(1-\pi_1)^{n_2} \tag{8}$$

and

$$p(\pi_2|\mathbf{S})\propto p(\mathbf{S}|\pi_2)p(\pi_2)\propto p(\pi_2)\pi_2^{n_3}(1-\pi_2)^{n_4} \tag{9}$$

where

$$\begin{aligned} n_1 &= \#\{t\in\{1,\ldots,T-1\}|s_t=1,s_{t+1}=1\}, \qquad n_2=\#\{t\in\{1,\ldots,T-1\}|s_t=1,s_{t+1}=2\} \\ n_3 &= \#\{t\in\{1,\ldots,T-1\}|s_t=2,s_{t+1}=2\}, \qquad n_4=\#\{t\in\{1,\ldots,T-1\}|s_t=2,s_{t+1}=1\}, \end{aligned}$$

and $\#$ denotes the number of elements in a set. $p(\pi_1)$ and $p(\pi_2)$ are prior distributions. For the choice of priors, let $\pi_1\sim Beta(\alpha_{\pi_1},\beta_{\pi_1})$ and $\pi_2\sim Beta(\alpha_{\pi_2},\beta_{\pi_2})$, where $Beta(.,.)$ denotes the beta distribution. Then we have the Gibbs sampling step

$$\pi_1|\mathbf{S}\sim Beta(\bar{\alpha}_{\pi_1},\bar{\beta}_{\pi_1}) \text{ and } \pi_2|\mathbf{S}\sim Beta(\bar{\alpha}_{\pi_2},\bar{\beta}_{\pi_2}) \tag{10}$$

where $\bar{\alpha}_{\pi_1}=\alpha_{\pi_1}+n_1$, $\bar{\beta}_{\pi_1}=\beta_{\pi_1}+n_2$, $\bar{\alpha}_{\pi_2}=\alpha_{\pi_2}+n_3$ and $\bar{\beta}_{\pi_2}=\beta_{\pi_2}+n_4$. If we impose the restriction $\pi_1>\pi_2$, we can jointly sample π_1 and π_2 using a Metropolis-Hastings (M-H) step with independent joint proposal of π_1',π_2' from (10). The proposal is accepted with probability $\alpha((\pi_1,\pi_2),(\pi_1',\pi_2')|\mathbf{S})=\mathbb{I}_{\pi_1'>\pi_2'}$.

[8] s_1 and s_T are sampled in similar fashion as other cases by excluding $Pr(s_1=1,2|s_0=1,2)$ and $Pr(s_{T+1}=1,2|s_T=1,2)$ from the corresponding probability kernels, respectively.

3.3 Sampling Θ

The likelihood function $p(\mathbf{Y}|\Theta,\mu,\mathbf{S})$ can be computed by integrating out $\mathbf{\Lambda}$ as

$$\begin{aligned}
&p(\mathbf{Y}|\Theta,\mu,\mathbf{S}) = \int p(\mathbf{Y}|\Theta,\mu,\mathbf{\Lambda})p(\mathbf{\Lambda}|\mathbf{S})d\mathbf{\Lambda} \\
&= \int \left(\prod_{t=1}^{T} p(y_t|\mu,H_t,\Lambda_t)\right) p(\mathbf{\Lambda}|\mathbf{S})d\mathbf{\Lambda} \\
&= \int \left(\prod_{t=1}^{T} (2\pi)^{-\frac{k}{2}} |H_t^{\frac{1}{2}}\Lambda_t(H_t^{\frac{1}{2}})'|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(y_t-\mu)'(H_t^{\frac{1}{2}}\Lambda_t(H_t^{\frac{1}{2}})')^{-1}(y_t-\mu)\right)\right) p(\mathbf{\Lambda}|\mathbf{S})d\mathbf{\Lambda} \\
&= \left[\prod_{t=1}^{T} |H_t|^{-\frac{1}{2}}\right] \int \left(\prod_{t=1}^{T} (2\pi)^{-\frac{k}{2}} |\Lambda_t|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(H_t^{-\frac{1}{2}}(y_t-\mu))'\Lambda_t^{-1}(H_t^{-\frac{1}{2}}(y_t-\mu))\right)\right) p(\mathbf{\Lambda}|\mathbf{S})d\mathbf{\Lambda} \\
&= \left[\prod_{t=1}^{T} |H_t|^{-\frac{1}{2}}\right] \int \left(\prod_{t=1}^{T} (2\pi)^{-\frac{k}{2}} |\Lambda_t|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\tilde{y}_t'\Lambda_t^{-1}\tilde{y}_t\right)\right) p(\mathbf{\Lambda}|\mathbf{S})d\mathbf{\Lambda} \\
&= \left[\prod_{t=1}^{T} |H_t|^{-\frac{1}{2}}\right] p(\tilde{\mathbf{Y}}|\mathbf{S}) \qquad (11)
\end{aligned}$$

The last line in (11) requires the computation of $p(\tilde{\mathbf{Y}}|\mathbf{S})$ which depends on the number of breakdown periods in the sample. Suppose, given $\mathbf{S}$, there are $B \geq 0$ breakdown period(s) over the whole sample denoted as $\mathcal{BP}_1, \mathcal{BP}_2, \ldots, \mathcal{BP}_B$. Each $\mathcal{BP}_q$ starts at date $t_{q,s}$ and ends at date $t_{q,e} \geq t_{q,s}$, with duration $N_q = t_{q,e} - t_{q,s} + 1$. If $\mathcal{BP}^c$ denotes the union of all the normal periods then $\mathcal{BP}^c$, $\mathcal{BP}_1, \mathcal{BP}_2, \ldots, \mathcal{BP}_B$ form a partition of $\{1, 2, \ldots, T\}$.

Now

$$\begin{aligned}
p(\tilde{\mathbf{Y}}|\mathbf{S}) &= \prod_{t\in\mathcal{BP}^c} N(\tilde{y}_t|0,I) \prod_{q=1}^{B} \int \left(\prod_{t\in\mathcal{BP}_q} p(\tilde{y}_t|\Lambda)\right) p(\Lambda)d\Lambda \\
&= \prod_{t\in\mathcal{BP}^c} \left((2\pi)^{-\frac{k}{2}} \exp(-\frac{1}{2}\tilde{y}_t'\tilde{y}_t)\right) \prod_{q=1}^{B} \frac{2^{\frac{N_q k}{2}} (2\pi)^{-\frac{N_q k}{2}} |Q_0|^{\frac{\nu}{2}} \prod_{j=1}^{k} \Gamma(\frac{N_q+\nu+1-j}{2})}{|Q_q|^{\frac{N_b+\nu}{2}} \prod_{j=1}^{k} \Gamma(\frac{\nu+1-j}{2})}
\end{aligned} \qquad (12)$$

where $Q_q = \sum_{t\in\mathcal{BP}_q} \tilde{y}_t\tilde{y}_t' + Q_0$. The conditional posterior distribution is:

$$p(\Theta|\mathbf{S},\mu,\mathbf{Y}) \;\propto\; p(\mathbf{Y}|\Theta,\mu,\mathbf{S})p(\Theta) \propto \left[\prod_{t=1}^{T} |H_t|^{-\frac{1}{2}}\right] p(\tilde{\mathbf{Y}}|\mathbf{S})p(\Theta). \qquad (13)$$

The choice of prior distribution $p(\Theta)$ depends on the chosen MGARCH specification. We use a multivariate Gaussian proposal distribution in a random walk M-H algorithm.

3.4 Sampling $\mathbf{\Lambda}$

Given $\mathbf{S}$, when $s_t = 1$, $\Lambda_t = I$. During a breakdown period, $\Lambda_t \neq I$ but remains constant. The number of unique non-identity Λ_t is equal to the number of breakdown periods B. Let $\tilde{\Lambda}_q$ be the unique value of Λ_t realized in $\mathcal{BP}_q$. Then the conditional posterior is

$$\tilde{\Lambda}_q \sim W^{-1}(\nu_q, Q_q), \tag{14}$$

where $\nu_q = \nu + N_q$ and $Q_q = \sum_{t \in \mathcal{BP}_q} \tilde{y}_t \tilde{y}_t' + Q_0$.

3.5 Sampling μ

Given $\mathbf{\Lambda}$, Θ and μ, the likelihood function is

$$p(\mathbf{Y}|\mathbf{\Lambda}, \Theta, \mu) \;=\; \prod_{t=1}^{T} N(y_t|\mu, H_t^{\frac{1}{2}} \Lambda_t (H_t^{\frac{1}{2}})'), \tag{15}$$

and the conditional posterior density is

$$p(\mu|\mathbf{\Lambda}, \Theta, \mathbf{Y}) \;\propto\; p(\mu) p(\mathbf{Y}|\mathbf{\Lambda}, \Theta, \mu). \tag{16}$$

Given prior distribution $p(\mu)$ an M-H step is used to sample from the posterior distribution. We specify an independent Gaussian prior.

This covers the steps for posterior simulation of the model. We now consider two important extensions to the basic covariance breakdown model.

3.6 Student-t innovations

A multivariate Student-t distribution can be used in place of the Gaussian assumption to account for fat tails. In the case of a VDGARCH specification, we refer to the fat-tailed version of the breakdown model as VDGARCH-t-B. Only minor adjustments on the original samplers are needed as a Student-t random variable can be written as a ratio of a Gaussian variable and the square root of a Gamma random variable.[9] More specifically, if y_t follows a Student-t distribution with mean μ, scale matrix $H_t^{1/2} \Lambda_t (H_t^{1/2})'$ and degree of freedom $d > 2$, or $y_t \sim t(\mu, H_t^{1/2} \Lambda_t (H_t^{1/2})', d)$, then it can be written as

$$y_t \;=\; \mu + u_t^{-1/2}(y_t^* - \mu), \tag{17}$$

where

$$y_t^* \;\sim\; N(\mu, H_t^{1/2} \Lambda_t (H_t^{1/2})') \tag{18}$$

$$u_t \;\sim\; G(d/2, d/2). \tag{19}$$

To facilitate posterior sampling, data augmentation is implemented again by treating $\mathbf{U} = \{u_t\}_{t=1}^T$ as unknown parameters. In this case, the full set of conditional posterior

[9] Denote a Gamma distribution as $G(a, b)$ which has mean a/b, and denote the associated density function as $G(.|a, b)$.

distributions consists of two more components: $p(\mathbf{U}|\Theta, \mathbf{\Lambda}, \mu, d, \mathbf{Y})$ and $p(d|\mathbf{U})$. To sample $\mathbf{U}$, let $V_t = H_t^{1/2}\Lambda_t(H_t^{1/2})'$ and note

$$\begin{aligned}
p(u_t|y_t, \Theta, \mathbf{\Lambda}, \mu, d) &\propto p(u_t)p(y_t|u_t, \Theta, \mathbf{\Lambda}, \mu) \\
&\propto G(u_t|d/2, d/2)N(y_t|\mu, u_t^{-1}V_t) \\
&= u_t^{d/2-1}e^{-du_t/2}(2\pi)^{-k/2}|u_t^{-1}V_t|^{-1/2}e^{-\frac{1}{2}(y_t-\mu)'(u_t^{-1}V_t)^{-1}(y_t-\mu)} \\
&\propto u_t^{\frac{k+d}{2}-1}e^{-\frac{1}{2}u_t(d+(y_t-\mu)'V_t^{-1}(y_t-\mu))} \\
&\propto G\left(u_t\middle|\frac{k+d}{2}, \frac{1}{2}(d+(y_t-\mu)'V_t^{-1}(y_t-\mu))\right). \qquad (20)
\end{aligned}$$

To sample d, let the prior of d follows a truncated exponential distribution with density function $p(d) \propto \text{Exp}(d|\lambda_0)\mathbb{I}_{d>2}$, where $\text{Exp}(d|\lambda_0) = \lambda_0 e^{-\lambda_0 d}$ is the density function of an exponential distribution with mean equal to $\frac{1}{\lambda_0}$. Then

$$\begin{aligned}
p(d|\mathbf{U}) &\propto p(d)p(\mathbf{U}|d) \\
&\propto \text{Exp}(d|\lambda_0)\mathbb{I}_{d>2}\prod_{t=1}^{T}\frac{(\frac{d}{2})^{d/2}}{\Gamma(\frac{d}{2})}u_t^{d/2-1}e^{-du_t/2} \\
&\propto \lambda_0 e^{-\lambda_0 d}\mathbb{I}_{d>2}\left(\frac{(\frac{d}{2})^{d/2}}{\Gamma(\frac{d}{2})}\right)^T \exp\left(-\frac{d}{2}\sum_{t=1}^{T}(u_t - \log(u_t))\right). \qquad (21)
\end{aligned}$$

The posterior can be sampled using an M-H step.

Given μ and u_t, let $y_t^* = (y_t-\mu)u_t^{1/2}+\mu$ and write $\mathbf{Y}^* = \{y_t^*\}_{t=1}^T$. The sampling procedure for the Student-t model consists of sequential draws from the following conditional posterior distributions[10]:

1. $p(\mathbf{S}|\Theta, \pi_1, \pi_2, \mu, \mathbf{Y}^*)$,
2. $p(\pi_1, \pi_2|\mathbf{S})$,
3. $p(\Theta|\mathbf{S}, \mu, \mathbf{Y}^*)$,
4. $p(\mathbf{\Lambda}|\mathbf{S}, \Theta, \mu, \mathbf{Y}^*)$,
5. $p(\mu|\mathbf{\Lambda}, \Theta, \mathbf{Y}^*)$,
6. $p(\mathbf{U}|\Theta, \mathbf{\Lambda}, \mu, d, \mathbf{Y})$,
7. $p(d|\mathbf{U})$.

Steps 1-5 are a repeat of the sampling steps in the Gaussian model but conditional on $\mathbf{Y}^*$ and require no additional coding.

3.7 Learning about Covariance Breakdowns

So far we have assumed $G_0 = W^{-1}(\nu, Q_0)$ with known parameters ν and Q_0. We can also introduce a hierarchy and place prior distributions on ν and Q_0 . By incorporating both

[10]Note that the conditional posterior distributions of $\mathbf{S}, \pi_1, \pi_2, \Theta, \mathbf{\Lambda}, \mu$ are each conditioned on the transformed data $\mathbf{Y}^*$, while the conditional posterior distributions of $\mathbf{U}$ is conditional on the un-transformed data $\mathbf{Y}$.

parameters into the posterior sampling scheme, we can learn about the typical effect of Λ_t. As shown in Section 8.2, sampling of ν and Q_0 can be included in the posterior sampling algorithm as an M-H step and a Gibbs step, respectively.

4 Simulation Study

To analyze the performance of the proposed estimation algorithm, we conduct a simulation study in a 2×2 dimension. We simulate 5000 observations from a VDGARCH-B model. The parameters are

$$C = \begin{pmatrix} 0.04 & 0 \\ 0.01 & 0.03 \end{pmatrix}, a = (0.12, 0.1)', b = (0.97, 0.98)', \pi_1 = 0.99, \pi_2 = 0.98, \mu = 0.$$

A Gaussian distribution is assumed for the innovations and learning about G_0 is ignored and instead specified as $G_0 = W^{-1}(7, 4I)$ so that $E[\Lambda_t] = I$. The prior distributions are as follows: $\pi_1 \sim Beta(3, 0.1)$, $\pi_2 \sim Beta(2, 0.1)$; $C_{ii} \sim TN_+(0, 100)$ for $i = 1, 2$, $C_{21} \sim N(0, 100)$; $a_1 \sim TN_+(0, 100)$, $a_2 \sim N(0, 100)$; $b_1 \sim TN_+(0, 100)$, $b_2 \sim N(0, 100)$. $TN_+(.)$ denotes the truncated Gaussian distribution on the positive real line. In the posterior sampling, the first 10000 MCMC draws are discarded as burn-in and the next 10000 draws are used for inference.

Parameter estimates are reported in Table 2. All parameter values are accurately recovered. Figure 4 displays the posterior mean of s_t, compared with the true states over time and shows that the model identifies the breakdown periods and normal periods well.

For comparison, we also estimate a plain VDGARCH model that does not allow for covariance breakdowns with the same data. This model can be seen as a special case or restricted version of our model with $\pi_1 = 1$ and $s_1 = 1$. The parameter estimates are reported in Table 2.[11] The results are very different from the true values, which is not surprising due to misspecification. For example, the posterior means of a_1 and a_2 are much higher than their true values while b_1 and b_2 have smaller estimates.

Figure 5 plots the smoothed states and smoothed variances from the models during a covariance breakdown episode. All elements of the volatility matrix are included in the comparison as well as the correlation coefficient. It is clear that the VDGARCH-B model is closer to the true volatility dynamics in general, and particularly so during the breakdown period. The VDGARCH model has conditional moments that are more erratic and deviate from the truth.

Table 3 repeats the exercise by estimating the models on simulated data with no covariance breakdowns. Here data are simulated from a plain VDGARCH model with parameters specified earlier and both specifications are estimated. The VDGARCH-B model does a good job in identifying no breakdowns and recovering the model parameters.

To quantitatively compare the fit of the models based on the volatility estimates in the presence of covariance breakdowns, we calculate the root mean squared error (RMSE) and report it in Table 4. Over the whole sample the VDGARCH-B model has a $RMSE$ of 0.230

[11]The prior distributions for the parameters are the same as those of the common parameters in the breakdown model.

and the VDGARCH model has a $RMSE$ of 0.630. Modeling the covariance breakdowns is important for accurate volatility estimation. Focusing only on the RMSE from covariance breakdown periods ($\{t|s_t = 2\}$) the results show that both models have a higher value but the loss from ignoring the covariance breakdowns is larger. The final columns of the table report the model losses when the data generating process contains no breakdowns ($s_t = 1, \forall t$).

Ignoring covariance breakdowns results in biased parameter estimates and poor volatility estimates.

5 Marginal Likelihood

The marginal likelihood is a key input in Bayesian model comparison. It is defined as the integral of the likelihood function with respect to the prior density. Our approach to computing the marginal likelihood is based on the method proposed by Chib (1995), which exploits the fact that the marginal likelihood can be expressed as:

$$f(\mathbf{Y}) = \frac{f(\mathbf{Y}|\Psi)f(\Psi)}{f(\Psi|\mathbf{Y})}, \tag{22}$$

where Ψ is the set of parameters, $f(\mathbf{Y}|\Psi)$ is the likelihood function, $f(\Psi)$ and $f(\Psi|\mathbf{Y})$ are the prior density and posterior density of the parameters, respectively. Equation (22) is called the basic marginal likelihood identity. It holds for any Ψ, but is most efficiently estimated at some high density point Ψ^*, such as the posterior mean or median. Calculation of the marginal likelihood amounts to computing three quantities: $f(\mathbf{Y}|\Psi^*)$, $f(\Psi^*)$ and $f(\Psi^*|\mathbf{Y})$. After evaluating the three quantities at some given parameter value Ψ^*, the marginal likelihood on the log scale can be estimated as

$$\log f(\mathbf{Y}) = \log f(\mathbf{Y}|\Psi^*) + \log f(\Psi^*) - \log f(\Psi^*|\mathbf{Y}). \tag{23}$$

To estimate $\log f(\mathbf{Y}|\Psi^*)$ the latent state variables $\mathbf{S}$ and $\mathbf{\Lambda}$ are integrated out of the likelihood function. We design a particle filter based on the auxiliary particle filter of Pitt & Shephard (1999) to achieve this purpose for our model. The second term, $\log f(\Psi^*)$, is the log-prior evaluated at Ψ^*. This is straightforward to compute given our priors. However, simulations are used to calculate integrating constants from prior restrictions such as $\pi_1 > \pi_2$. The final term, $\log f(\Psi^*|\mathbf{Y})$, is the log-posterior ordinate. We follow Chib & Jeliazkov (2001), who provide an approach that can be used for M-H sampling steps while Chib (1995) can be used for the Gibbs sampling steps. Full details of the marginal likelihood estimation are found in Section 8.3.

6 Empirical Application

In this section we apply the model to study the volatility dynamics among the stock market and the bond market.[12] We use daily excess returns on the S&P 500 index ($y_{1,t}$), a ten-year

[12] For applications of MGARCH models to stock and bond markets without covariance breakdowns see Cappiello et al. (2006) and De Goeij & Marquering (2004).

Treasury bond ($y_{2,t}$), and a one-year Treasury bond ($y_{3,t}$). The return data are obtained from the Center for Research on Security Prices (CRSP). The excess returns are obtained by subtracting the risk-free return approximated by the three-month Treasury bill rate. The sample period runs from 1987/01/02 to 2011/12/30, delivering 6244 observations. Figure 6 plots the three excess return series and Table 5 provides summary statistics. Returns are in percentage.

We estimate the VDGARCH-B model and the VDGARCH-t-B model using the return data. For the GARCH parameter Θ, the priors on the elements of a, b, C are all independent $N(0, 100)$, except that a_1, b_1 and the diagonal elements of C are truncated to be positive for identification purposes. The other prior distributions are set as follows: $\nu \sim \text{Exp}_{\nu>k-1}(0.1)$, an exponential distribution with support truncated to be greater than $k-1$; $Q_0 \sim W(5, I)$, a Wishart distribution with 5 degrees of freedom and scale matrix equal to I; $\mu \sim N(0, 100I)$. The degree of freedom of the Student-t distribution in VDGARCH-t-B follows a truncated exponential distribution, $d \sim \text{Exp}_{d>2}(0.1)$.

The priors of π_1 and π_2 are set with $\alpha_{\pi_1} = 20$, $\beta_{\pi_1} = 0.1$, $\alpha_{\pi_2} = 2$, $\beta_{\pi_2} = 0.1$ and favor infrequent covariance breakdowns. To ensure that covariance breakdowns are meaningful and not just capturing outliers the posterior sampling of s_t is restricted to span a minimum duration of D days for each state (normal and breakdown). The original case corresponds with $D = 1$. In this analysis, we set $D = 5$ which represents a normal business week.[13] Finally, the restriction of $\pi_1 > \pi_2$ ensures normal periods dominate breakdown periods so that the GARCH dynamics would prevail in the long-run, serving as the main driving force of volatility dynamics. The first 10000 draws of the MCMC chain are discarded as burn-in and the next 10000 draws are used for inference.

First, we discuss the VDGARCH-B model. The parameter estimates are reported in Table 6. Figure 7a plots the posterior mean of s_t. A visual inspection suggests that a large number of breakdown periods are identified by the model. The posterior average number of breakdown periods is 199. The posterior average of the breakdown period duration is 14 days. The top panel of Table 7 shows the empirical posterior distribution of the duration of the breakdown periods (state 2 duration). More than 88% of the breakdown periods have a duration of less than or equal to 30 days. 14% of the breakdown periods have the minimum duration of 5 days. Figure 7b plots the posterior mean of $\log(|\Lambda_t|)$. Recall that $\log(|\Lambda_t|) > 0$ ($\log(|\Lambda_t|) < 0$) means a scale-up (scale-down) effect on the volatility matrix. The figure shows that $\log(|\Lambda_t|) > 0$ most of the time. This suggests during most of the breakdown periods, the overall variability is scaled up. The above results indicate that many short-lived breakdown periods occur to scale up the volatility in order to pick up tail realizations of the return distribution.

Next, we turn our attention to the VDGARCH-t-B model. The parameter estimates are reported in Table 8. Relative to Table 6, the notable difference is that with the t-distribution both π_1 and π_2 have much higher estimated values. This means that state durations will tend to be longer with fewer breakdown periods (and normal periods). This is evident in the plot of the posterior mean of s_t in Figure 8a. The posterior average number of breakdown periods is 44, less than one fourth of the number in the Gaussian case. The breakdown periods cover less days. The posterior average number of days in state 2 is 2282, accounting

[13]This requires only minor modifications to the original sampler of s_t.

for 36% of the total sample. The posterior average duration of a breakdown is 58 days, four times in length as compared to the Gaussian case.

The empirical posterior distribution of the duration of the breakdown periods is shown in the bottom half of Table 7. More than 30% of the breakdown periods have durations of more than 60 days and on average only 2 breakdown periods have the minimum duration of 5 days. According to the estimates of the Markov chain parameters π_1 and π_2 in Table 8, covariance breakdowns occur 34% of the time and their expected duration is 1-2 months.

Evidently, there is a substitution effect between the t-distribution and Λ_t. A large number of those short-lived breakdown periods with jumps in the overall variability with Gaussian innovations are classified as tail observations under the fat-tailed Student-t assumption. A fat-tailed innovation distribution is important in separating transient outliers from sustained covariance breakdowns.

Figure 8a also shows clusters of breakdown periods at the beginning, in the middle part and also towards the end of the whole sample. Some of these are discussed later. Figure 8b plots the posterior mean of $\log(|\Lambda_t|)$. When there is likely a normal state indicated by the posterior mean of s_t, $\log(|\Lambda_t|)$ is close to one as it should be; while during breakdown periods, $\log(|\Lambda_t|)$ have both positive and negative values. This means that the model has identified both breakdown periods with overall increased variability and those with reduced variability.

Table 9 reports posterior summary statistics for the expected impact of covariance breakdowns. The posterior mean of 0.5808 implies an average increase in variability when a breakdown occurs. However, the 0.95 density interval shows that reductions in variability do occur (a negative value of $\log(|E[\Lambda]|)$ implies a decrease in the generalized variance). The density interval shows that the distribution of expected breakdowns is asymmetric in that increases in variability are much more likely than decreases.

We also estimate a plain VDGARCH model with Student-t innovations (VDGARCH-t)[14]. See Table 8 for parameter estimates. Compared to VDGARCH-t-B, VDGARCH-t has larger estimates of a_i. Meanwhile, the estimate of the degree of freedom d is smaller, evidence of the substitution effect between the t-distribution and covariance breakdowns.

6.1 Model Comparison

To formally assess whether covariance breakdowns admitted in our model are supported by the data, we compute the marginal likelihoods and compare models based on Bayes factors. The marginal likelihood for the covariance breakdown specifications is evaluated as discussed in Section 5.

Computing the marginal likelihood for the models without covariance breakdowns is straightforward using the method in Chib & Jeliazkov (2001), where evaluating the likelihood function and the prior ordinates is trivial, and the posterior ordinate can be evaluated using a single block proposal within M-H steps.

Table 10 reports the results for several MGARCH specifications with Gaussian and Student-t innovations, and with or without covariance breakdowns. The final column of the table reports the log-Bayes factor (the difference of the log marginal likelihoods of the two

[14]The prior distributions for the parameters are the same as those for the common parameters in VDGARCH-t-B.

models) for the VDGARCH-t-B against each of its alternatives. This model is strongly favored against all alternatives. Adding covariance breakdowns improves each MGARCH specification. The log Bayes factor in favor of the VDGARCH-t-B model versus the VDGARCH-t is 59.575, which is overwhelming evidence.[15]

6.2 Covariance Breakdown Episodes

In this section we discuss several identified covariance breakdowns from the VDGARCH-t-B model. In particular, we compare the volatility dynamics under the full specification which includes the impact of the breakdown $V_t = H_t^{1/2}\Lambda_t(H_t^{1/2})'$ against the MGARCH component H_t, of the model.

6.2.1 2001-2005

The first episode is between 2001/01/02 and 2005/12/30 and found in Figure 9. Two adjacent extended breakdown periods are identified (Figure 9a). Together they span over one year in time from May 2002 to May 2003, a period featuring significant stock market downturn after the "Internet bubble bursting". The first breakdown period finishes at around mid October 2002 and the second period starts about a week later. Although the two periods occur closely in time they are distinguished from each other, evident by the large difference in $\log(|\Lambda_t|)$. In both periods $\log(|\Lambda_t|) < 0$, indicating overall reduced variability relative to H_t. During these breakdowns we see a large deviation between the conditional moments of V_t and H_t.

The first breakdown period witnesses the bankruptcy case of WorldCom.[16] This covariance breakdown results in significant jumps in the conditional variances of the S&P 500 and the ten-year bond while the one-year bond decreases relative to H_t. There are large drops in the conditional covariances of excess returns between the stock and bond markets. These changes cause substantial drops in conditional correlations. For instance, the conditional correlation between S&P 500 and the ten-year bond drops from about 0 to below -0.6 with a similar effect between the stock market and the one-year bond. On the other hand, the conditional correlation on excess returns from the bond markets spike to well above 0.7.

The second breakdown is from November 2002 to April 2003. The main effect is on the conditional variance of the one-year bond which drops to lower levels than H_t. This results in a reduction in the conditional correlation between S&P and the one-year bond (Figure 9i) and an increase between the two bonds (Figure 9j). Note, that the apparent breakdowns in these correlations have their source in the drop of the conditional variance of the one-year bond and not covariances, since the conditional covariances are very close to those from H_t (after November 2002). Immediately after the end of the second breakdown both H_t and V_t are essentially the same.

[15] Kass & Raftery (1995) suggest interpreting the evidence for model $\mathcal{A}$ as: not worth more than a bare mention if $0 \leq \log(BF_{\mathcal{AB}}) < 1$; positive if $1 \leq \log(BF_{\mathcal{AB}}) < 3$; strong if $3 \leq \log(BF_{\mathcal{AB}}) < 5$; and very strong if $\log(BF_{\mathcal{AB}}) \geq 5$.

[16] On May 9, 2002, Standard & Poor's and Moody's cut WorldCom's credit rating to junk status. On July 19, 2002, the bankruptcy filing is expected on the next business day (Akhigbe et al. 2005).

In summary, during this period we have covariance breakdowns that impact conditional variances, covariances and correlations.

6.2.2 2008-2011

The second episode is from 2008/01/02 to 2011/12/30 and is found in Figure 10. This includes the recent financial crisis of 2008. Not surprisingly, two consecutive sustained breakdown periods are identified by the model starting from September 2008 until November 2009 while there are shorter breakdowns in 2010 and 2011. The first breakdown period lasts for three months. During this period the covariance breakdown implies an increase in overall variability of $\exp(2.7) \approx 15$ times.[17]

Some of the covariance breakdowns over this sample period increase variability while some decrease it. The differences in the conditional variances and covariances of V_t and H_t are more pronounced than the differences between the conditional correlations. In other words, these are covariance breakdowns that have little to no impact on the conditional correlations. Relative to H_t, there is no evidence of a correlation breakdown in V_t.

6.2.3 1987-1989

The final episode is from 1987/01/02 to 1989/12/30 and is plotted in Figure 11. This features the stock market crash in October of 1987, which is within a one-month breakdown period identified by the model.

The main feature of this period is the large increase in conditional variances of all excess returns. The breakdown model implies at least a doubling of the conditional variances compared to those from H_t. The second breakdown after the crash provides a relief valve that puts the breakdown variances below those of H_t. This allows for a faster return to normal levels of volatility. The conditional covariances show a spike associated with the crash and these translate into sustained breakdowns in the conditional correlations between stock and bond markets. There is no evidence of a breakdown in the conditional correlation of excess returns between the two bonds. In other words, the spikes in the conditional variances and covariance in the bond market largely cancel out in the conditional correlation.

The covariance breakdown model is flexible enough to capture complex and erratic temporary structural change/deviation from the long run volatility dynamics which is otherwise difficult to account for comprehensively. It provides a relief value to release the excessive volatility built into MGARCH models after a shock as well as a mechanism to capture abrupt increases/decreases in variance, covariance and correlation dynamics.

6.3 Portfolio Choice

We evaluate the out-of-sample performance of the models from a portfolio optimization perspective. We consider a risk-averse investor who allocates funds among three risky assets, namely, the stock market portfolio, the ten-year bond, the one-year bond, and the risk-free asset. The investor bases her decision on the mean-variance criterion and rebalances her

[17] $\log(|\Lambda_t|) \approx 2.7$ during this time period (Figure 10a).

portfolio daily using a volatility-timing strategy. Specifically, at each day t, she solves for the minimum variance portfolio subject to a required return constraint:

$$\min_{w_{t+1}} w'_{t+1}\Sigma_{t+1}w_{t+1}, \tag{24}$$

$$\text{s.t.} \quad w'_{t+1}\mu = \mu_0. \tag{25}$$

w_{t+1} is the 3×1 vector of portfolio weights to be chosen at time t, Σ_{t+1} is the one-period ahead forecast of the time $t+1$ covariance matrix of y_{t+1}, μ is the assumed vector of expected excess returns over the risk-free return, and μ_0 is the required (target) portfolio return in excess of the risk-free return. The solution renders the optimal portfolio weight

$$w_{t+1} = \frac{\Sigma_{t+1}^{-1}\mu}{\mu'\Sigma_{t+1}^{-1}\mu}\mu_0. \tag{26}$$

The realized portfolio return (in excess of the risk-free rate) is given by

$$R_{t+1} = w'_{t+1}y_{t+1}. \tag{27}$$

Note that $\sum_{i=1}^{3} w_{t,i}$ will not equal one in general, and $1 - \sum_{i=1}^{3} w_{t,i}$ is the share invested in the risk-free asset.

To evaluate the economic gains of allowing for covariance breakdowns in the MGARCH volatility dynamics in the context of portfolio selection, we use the utility-based approach following Fleming et al. (2001) and Clements & Silvennoinen (2013). Let $\{R_{1t}\}_{t=T_0}^{T}$ be the realized portfolio returns over the out-of-sample period using volatility forecasts based on the VDGARCH-t model, and $\{R_{2t}\}_{t=T_0}^{T}$ be those based on the covariance breakdown (VDGARCH-t-B) model. [18] Given a utility function $\mathcal{U}(.)$, we find a constant Δ that equates the total realized utility in

$$\sum_{t=T_0}^{T} \mathcal{U}(R_{1t}) = \sum_{t=T_0}^{T} \mathcal{U}(R_{2t} - \Delta). \tag{28}$$

Δ is the daily maximum return the investor would be willing to give up in exchange for the economic gains obtained by switching from the model with no covariance breakdowns to one with breakdowns. As such, Δ measures the incremental benefit of allowing for covariance breakdowns as opposed to otherwise. A positive value of Δ means that allowing for covariance breakdowns will generate extra economic benefit for the investor. Here we consider two types of utility functions. One is the quadratic utility function in Fleming et al.(2001, 2003)

$$\mathcal{U}_q(R_t) = (1 + r_{ft} + R_t) - \frac{\tau}{2(1+\tau)}(1 + r_{ft} + R_t)^2, \tag{29}$$

and the other is the negative exponential utility used in Clements & Silvennoinen (2013) and Skouras (2007)

$$\mathcal{U}_e(R_t) = -\exp(-\tau(1 + r_{ft} + R_t)). \tag{30}$$

[18] The forecasts of Σ_{t+1} are computed using parameter estimates conditioning on information up to time t for either model. In other words, the models are estimated recursively over the whole out-of-sample observations, mimicking real-time forecasting.

r_{ft} is the risk-free return and τ is the investor's coefficient of risk aversion.

To focus on the difference that volatility dynamics make we demean the data and estimate the models with a zero intercept, and we set μ in (25) to be the sample mean for both models. Any differences in portfolio choice are directly related to the differences in the covariance dynamics. We consider two out-of-sample periods. The first sample period focuses on the financial crisis while the second is extended to a longer period prior to the crisis. Table 11 reports the results for portfolio performance from the covariance timing strategies for several required return values μ_0. Overall, an investor is willing to pay for the covariance breakdown model. It achieves a higher Sharpe ratio in both samples. The performance fee is largest for larger μ_0. These results show that the superior predictability of the covariance breakdown model translates into economic gains in portfolio choice.

7 Conclusion

This paper proposes a flexible way of accommodating dynamic heterogeneous breakdown periods in the conditional covariance matrix of multivariate GARCH models. During periods of normal market activity, volatility dynamics are governed by an MGARCH specification. A covariance breakdown is any significant temporary deviation of the conditional covariance matrix from its implied MGARCH dynamics. This is captured through a flexible stochastic component that allows for changes in the conditional variances, covariances and implied correlation coefficients. Bayesian inference is used and we propose an efficient posterior sampling procedure. We show how to compute the marginal likelihood of our model. Application in daily stock and bond return data shows the benefit of our approach. The new model is strongly supported by Bayes factors while gains to portfolio choice are also documented.

8 Appendix

8.1 Derivation of Equation (7)

We compute the integral $\int \left(\prod_{t=1}^{n} p(y_t|\Lambda)\right) p(\Lambda) d\Lambda$ for some $n \geq 1$ where $\Lambda \sim W^{-1}(\nu, Q_0)$ and $y_t \sim NID(0, \Lambda)$. First note that

$$\begin{aligned}
\left(\prod_{t=1}^{n} p(y_t|\Lambda)\right) &= \prod_{t=1}^{n} \frac{1}{(2\pi)^{\frac{k}{2}}|\Lambda|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} y_t' \Lambda^{-1} y_t\right) \\
&= (2\pi)^{-\frac{nk}{2}} |\Lambda|^{-\frac{n}{2}} \exp\left(-\frac{1}{2}\sum_{t=1}^{n} y_t' \Lambda^{-1} y_t\right) \\
&= (2\pi)^{-\frac{nk}{2}} |\Lambda|^{-\frac{n}{2}} \exp\left(-\frac{1}{2}\sum_{t=1}^{n} Tr(y_t' \Lambda^{-1} y_t)\right) \\
&= (2\pi)^{-\frac{nk}{2}} |\Lambda|^{-\frac{n}{2}} \exp\left(-\frac{1}{2}\sum_{t=1}^{n} Tr(\Lambda^{-1} y_t y_t')\right).
\end{aligned}$$

Therefore,

$$
\begin{aligned}
\left(\prod_{t=1}^{n} p(y_t|\Lambda)\right) p(\Lambda) &= (2\pi)^{-\frac{nk}{2}} |\Lambda|^{-\frac{n}{2}} \exp\left(-\frac{1}{2}\sum_{t=1}^{n} Tr(\Lambda^{-1} y_t y_t')\right) \\
&\quad \times \frac{|Q_0|^{\frac{\nu}{2}} |\Lambda|^{-\frac{\nu+k+1}{2}}}{2^{\frac{\nu k}{2}} \prod_{j=1}^{k} \Gamma(\frac{\nu+1-j}{2})} \exp\left(-\frac{1}{2} Tr(\Lambda^{-1} Q_0)\right) \\
&= \frac{(2\pi)^{-\frac{nk}{2}} |\Lambda|^{-\frac{n+\nu+k+1}{2}} |Q_0|^{\frac{\nu}{2}}}{2^{\frac{\nu k}{2}} \prod_{j=1}^{k} \Gamma(\frac{\nu+1-j}{2})} \exp\left(-\frac{1}{2} Tr[\Lambda^{-1}(\sum_{t=1}^{n} y_t y_t' + Q_0)]\right) \\
&= \frac{|Q|^{\frac{n+\nu}{2}} |\Lambda|^{-\frac{n+\nu+k+1}{2}}}{2^{\frac{(n+\nu)k}{2}} \prod_{j=1}^{k} \Gamma(\frac{n+\nu+1-j}{2})} \exp\left(-\frac{1}{2} Tr(\Lambda^{-1} Q)\right) \\
&\quad \times \frac{2^{\frac{nk}{2}} (2\pi)^{-\frac{nk}{2}} |Q_0|^{\frac{\nu}{2}} \prod_{j=1}^{k} \Gamma(\frac{n+\nu+1-j}{2})}{|Q|^{\frac{n+\nu}{2}} \prod_{j=1}^{k} \Gamma(\frac{\nu+1-j}{2})} \\
&= W^{-1}(\Lambda|n+\nu, Q) \times \frac{2^{\frac{nk}{2}} (2\pi)^{-\frac{nk}{2}} |Q_0|^{\frac{\nu}{2}} \prod_{j=1}^{k} \Gamma(\frac{n+\nu+1-j}{2})}{|Q|^{\frac{n+\nu}{2}} \prod_{j=1}^{k} \Gamma(\frac{\nu+1-j}{2})} \qquad (31)
\end{aligned}
$$

where $Q = \sum_{t=1}^{n} y_t y_t' + Q_0$. So integrating this final result with respect to Λ leaves the second term on the right hand side of (31) since the first term is a well defined inverse-Wishart density that integrates to 1. That is,

$$
\int \left(\prod_{t=1}^{n} p(y_t|\Lambda)\right) p(\Lambda) d\Lambda = \frac{2^{\frac{nk}{2}} (2\pi)^{-\frac{nk}{2}} |Q_0|^{\frac{\nu}{2}} \prod_{j=1}^{k} \Gamma(\frac{n+\nu+1-j}{2})}{|Q|^{\frac{n+\nu}{2}} \prod_{j=1}^{k} \Gamma(\frac{\nu+1-j}{2})}. \qquad (32)
$$

8.2 Sampling ν and Q_0

For ν, the conditional posterior distribution is

$$
\begin{aligned}
p(\nu|\mathbf{Y}, Q_0, \Theta, \mu, \mathbf{S}) &\propto p(\mathbf{Y}|\Theta, \mu, \nu, \mathbf{S}, Q_0) p(\nu) \\
&\propto p(\tilde{\mathbf{Y}}|\mathbf{S}, Q_0, \nu) p(\nu) \\
&\propto \left(\prod_{q=1}^{B} \frac{|Q_0|^{\frac{\nu}{2}} \prod_{j=1}^{k} \Gamma(\frac{N_q+\nu+1-j}{2})}{|Q_q|^{\frac{\nu}{2}} \prod_{j=1}^{k} \Gamma(\frac{\nu+1-j}{2})}\right) p(\nu). \qquad (33)
\end{aligned}
$$

The prior of ν is an exponential distribution with support truncated to be greater than $k-1$, $p(\nu) \propto \text{Exp}(\nu|\lambda_1)\mathbb{I}_{\nu>k-1}$. Sampling from $p(\nu|\mathbf{Y}, Q_0, \Theta, \mathbf{S})$ can be achieved by an M-H step with a random walk proposal.

For Q_0, let $p(Q_0) = W(Q_0|\gamma_0, A)$, where $W(.|.,.)$ is the Wishart density function with scalar degree of freedom γ_0 and scale matrix A. then

$$\begin{aligned}
p(Q_0|\nu,\{\tilde{\Lambda}_q\}_{q=1}^B) &\propto p(\{\tilde{\Lambda}_q\}_{q=1}^B|Q_0,\nu)p(Q_0) \\
&\propto \left(\prod_{q=1}^B |Q_0|^{\frac{\nu}{2}}\exp(-\frac{1}{2}Tr(\tilde{\Lambda}_q^{-1}Q_0))\right)|Q_0|^{\frac{\gamma_0-k-1}{2}}\exp(-\frac{1}{2}Tr(A^{-1}Q_0)) \\
&\propto |Q_0|^{\frac{B\nu}{2}}\exp\left(-\frac{1}{2}Tr(\sum_{q=1}^B\tilde{\Lambda}_q^{-1}Q_0)\right)\times|Q_0|^{\frac{\gamma_0-k-1}{2}}\exp(-\frac{1}{2}Tr(A^{-1}Q_0)) \\
&\propto |Q_0|^{\frac{B\nu+\gamma_0-k-1}{2}}\exp\left(-\frac{1}{2}Tr\left((\sum_{q=1}^B\tilde{\Lambda}_q^{-1}+A^{-1})Q_0\right)\right) \\
&\propto W_k(Q_0|\gamma,\widetilde{A}) \qquad (34)
\end{aligned}$$

where $\gamma = B\nu+\gamma_0, \widetilde{A} = (\sum_{q=1}^B \tilde{\Lambda}_q^{-1} + A^{-1})^{-1}$.

8.3 Marginal Likelihood Estimation

Below the estimation of each of the components of (23) is provided.

8.3.1 Estimating $f(\mathbf{Y}|\Psi)$

We assume Student-t innovation for the data:[19]

$$y_t \sim t(\mu, H_t^{1/2}\Lambda_t(H_t^{1/2})', d).$$

The parameter set Ψ includes $\Theta,\mu,\pi_1,\pi_2,d,\nu,Q_0$. Write $\tilde{Y}_t=\{\tilde{y}_t\}_{t=1}^T$ and $\tilde{y}_t = H_t^{-1/2}(y_t-\mu)$. Note that

$$f(\mathbf{Y}|\Psi) = \left[\prod_{t=1}^T |H_t|^{-1/2}\right] f(\tilde{\mathbf{Y}}|\pi_1,\pi_2,d,\nu,Q_0). \qquad (35)$$

The likelihood function can be obtained by computing $\prod_{t=1}^T |H_t|^{-1/2}$ and $f(\tilde{\mathbf{Y}}|\Psi_1)$, where $\Psi_1 = \{\pi_1,\pi_2,d,\nu,Q_0\}$. Computing $\prod_{t=1}^T |H_t|^{-1/2}$ is straightforward given Θ and μ. $f(\tilde{\mathbf{Y}}|\Psi_1)$ is the likelihood function of the transformed data $\tilde{\mathbf{Y}}$. It can be shown that $\tilde{\mathbf{Y}}$ correspond with the following "transformed" model:

$$\tilde{y}_t|\Lambda_t \sim t(0,\Lambda_t,d) \qquad (36)$$

$$\Lambda_t|(\Lambda_{t-1}=I) \quad \begin{cases} = I & \text{with probability} \quad \pi_1 \\ \sim W^{-1}(\nu,Q_0) & \text{with probability} \quad 1-\pi_1 \end{cases} \qquad (37)$$

$$\Lambda_t|(\Lambda_{t-1}\neq I) \quad \begin{cases} = \Lambda_{t-1} & \text{with probability} \quad \pi_2 \\ = I & \text{with probability} \quad 1-\pi_2 \end{cases} \qquad (38)$$

[19] For the case of Gaussian innovations, the likelihood function can be obtained in a similar and simpler way.

Once again using the fact that a Student-t random variable can be written as a ratio of a Gaussian variable and the square root of a Gamma random variable, the model can be further converted into a conditionally Gaussian state space model:

$$\tilde{y}_t|\Lambda_t, u_t \sim N(0, u_t^{-1}\Lambda_t) \tag{39}$$

$$\Lambda_t|(\Lambda_{t-1} = I) \quad \begin{cases} = I & \text{with probability} \quad \pi_1 \\ \sim W^{-1}(\nu, Q_0) & \text{with probability} \quad 1-\pi_1 \end{cases} \tag{40}$$

$$\Lambda_t|(\Lambda_{t-1} \neq I) \quad \begin{cases} = \Lambda_{t-1} & \text{with probability} \quad \pi_2 \\ = I & \text{with probability} \quad 1-\pi_2 \end{cases} \tag{41}$$

$$u_t \overset{iid}{\sim} G(d/2, d/2) \tag{42}$$

The state variables are (Λ_t, u_t). The transformed likelihood function can be written as

$$f(\tilde{\mathbf{Y}}|\Psi_1) = \prod_{t=1}^{T} f(\tilde{y}_t|\tilde{\mathbf{Y}}_{t-1}, \Psi_1) = \prod_{t=1}^{T} \int f(\tilde{y}_t|\Lambda_t, u_t) f(\Lambda_t, u_t|\tilde{\mathbf{Y}}_{t-1}, \Psi_1) d(\Lambda_t, u_t), \tag{43}$$

where $\tilde{\mathbf{Y}}_t = \{\tilde{y}_i\}_{i=1}^t$. To approximate the likelihood function, we design an Auxiliary Particle Filter (APF) (Pitt & Shephard 1999) to sequentially sample from the filtering distribution $f(\Lambda_t, u_t|\tilde{\mathbf{Y}}_t, \Psi_1)$, $t = 1, \ldots, T$. Given M particles $\{(\Lambda_t^{(j)}, u_t^{(j)})\}_{j=1}^M$ from $f(\Lambda_t, u_t|\tilde{\mathbf{Y}}_t, \Psi_1)$, each with the same discrete probability mass $1/M$ and suppressing Ψ_1 from the conditioning set, the predictive density is

$$\begin{aligned} f(\Lambda_{t+1}, u_{t+1}|\tilde{\mathbf{Y}}_t) &= \int f(\Lambda_{t+1}, u_{t+1}|\Lambda_t, u_t) f(\Lambda_t, u_t|\tilde{\mathbf{Y}}_t) d(\Lambda_t, u_t) \\ &= \int f(u_{t+1}) f(\Lambda_{t+1}|\Lambda_t) f(\Lambda_t, u_t|\tilde{\mathbf{Y}}_t) d(\Lambda_t, u_t) \\ &\approx f(u_{t+1}) \sum_{j=1}^{M} f(\Lambda_{t+1}|\Lambda_t^{(j)}) \frac{1}{M}. \end{aligned} \tag{44}$$

Therefore

$$f(\Lambda_{t+1}, u_{t+1}|\tilde{\mathbf{Y}}_{t+1}) \propto f(\tilde{y}_{t+1}|\Lambda_{t+1}, u_{t+1}) f(u_{t+1}) \sum_{j=1}^{M} f(\Lambda_{t+1}|\Lambda_t^{(j)}). \tag{45}$$

To sample from $f(\Lambda_{t+1}, u_{t+1}|\tilde{\mathbf{Y}}_{t+1})$, introduce an auxiliary discrete variable $m \in \{1, \ldots, M\}$ and define

$$f(\Lambda_{t+1}, u_{t+1}, m|\tilde{\mathbf{Y}}_{t+1}) \propto f(\tilde{y}_{t+1}|\Lambda_{t+1}, u_{t+1}) f(u_{t+1}) f(\Lambda_{t+1}|\Lambda_t^{(m)}). \tag{46}$$

If we draw from this joint distribution in (46) and then discard the index m, we will produce a sample from the distribution in (45).

To sample from $f(\Lambda_{t+1}, u_{t+1}, m|\tilde{\mathbf{Y}}_{t+1})$, we use Gibbs steps to iteratively sample from the conditional distributions of $f(u_{t+1}|\Lambda_{t+1}, m, \tilde{\mathbf{Y}}_{t+1})$ and $f(\Lambda_{t+1}, m|u_{t+1}, \tilde{\mathbf{Y}}_{t+1})$. Note that

$$\begin{aligned} f(u_{t+1}|\Lambda_{t+1}, m, \tilde{\mathbf{Y}}_{t+1}) &\propto f(\tilde{y}_{t+1}|\Lambda_{t+1}, u_{t+1}) f(u_{t+1}) \\ &\propto G\left(u_{t+1} \middle| \frac{k+d}{2}, \frac{d + \tilde{y}_{t+1}' \Lambda_{t+1}^{-1} \tilde{y}_{t+1}}{2}\right) \end{aligned} \tag{47}$$

where $G(.|.,.)$ is the Gamma probability density function. We discuss the sampling of $f(\Lambda_{t+1}, m|u_{t+1}, \tilde{\mathbf{Y}}_{t+1})$ in detail:

• Case I: If $\Lambda_t^{(m)} = I$

$$f(\tilde{y}_{t+1}|\Lambda_{t+1}, u_{t+1})f(\Lambda_{t+1}|\Lambda_t^{(m)}) = \begin{cases} N(\tilde{y}_{t+1}|0, u_{t+1}^{-1}I)\pi_1 & \text{if} \quad \Lambda_{t+1} = I \\ N(\tilde{y}_{t+1}|0, u_{t+1}^{-1}\Lambda_{t+1})W^{-1}(\Lambda_{t+1}|\nu, Q_0)(1-\pi_1) & \text{if} \quad \Lambda_{t+1} \neq I. \end{cases} \tag{48}$$

Thus we have

$$Pr(\Lambda_{t+1} = I|\tilde{\mathbf{Y}}_{t+1}, \Lambda_t^{(m)} = I, u_{t+1}) \propto N(\tilde{y}_{t+1}|0, u_{t+1}^{-1}I)\pi_1 \tag{49}$$

$$\begin{aligned} Pr(\Lambda_{t+1} \neq I|\tilde{\mathbf{Y}}_{t+1}, \Lambda_t^{(m)} = I, u_{t+1}) &\propto \int N(\tilde{y}_{t+1}|0, \lambda_{t+1}^{-1}\Lambda_{t+1})W^{-1}(\Lambda_{t+1}|\nu, Q_0)d\Lambda_{t+1}(1-\pi_1) \\ &= \frac{\Pi^{-\frac{k}{2}}u_{t+1}^{\frac{k}{2}}|Q_0|^{\frac{\nu}{2}}\prod_{j=1}^k\Gamma(\frac{\nu+2-j}{2})}{|\tilde{Q}_{t+1}|^{\frac{1+\nu}{2}}\prod_{j=1}^k\Gamma(\frac{\nu+1-j}{2})}(1-\pi_1), \end{aligned} \tag{50}$$

where $\tilde{Q}_{t+1} = u_{t+1}\tilde{y}_{t+1}\tilde{y}_{t+1}' + Q_0$. Define

$$g_m(\tilde{y}_{t+1}, u_{t+1}) = N(\tilde{y}_{t+1}|0, u_{t+1}^{-1}I)\pi_1 + \frac{\Pi^{-\frac{k}{2}}u_{t+1}^{\frac{k}{2}}|Q_0|^{\frac{\nu}{2}}\prod_{j=1}^k\Gamma(\frac{\nu+2-j}{2})}{|\tilde{Q}_{t+1}|^{\frac{1+\nu}{2}}\prod_{j=1}^k\Gamma(\frac{\nu+1-j}{2})}(1-\pi_1). \tag{51}$$

• Case II: If $\Lambda_t^{(m)} \neq I$

$$f(\tilde{y}_{t+1}|\Lambda_{t+1}, u_{t+1})f(\Lambda_{t+1}|\Lambda_t^{(m)}) = \begin{cases} N(\tilde{y}_{t+1}|0, u_{t+1}^{-1}I)(1-\pi_2) & \text{if} \quad \Lambda_{t+1} = I \\ N(\tilde{y}_{t+1}|0, u_{t+1}^{-1}\Lambda_t^{(m)})\pi_2 & \text{if} \quad \Lambda_{t+1} = \Lambda_t^{(m)}. \end{cases} \tag{52}$$

In this case, define

$$g_m(\tilde{y}_{t+1}, u_{t+1}) = N(\tilde{y}_{t+1}|0, u_{t+1}^{-1}I)(1-\pi_2) + N(\tilde{y}_{t+1}|0, u_{t+1}^{-1}\Lambda_t^{(m)})\pi_2. \tag{53}$$

Putting this together, to sample from $f(\Lambda_{t+1}, m|\tilde{\mathbf{Y}}_{t+1}, u_{t+1})$, first sample the index m with probability in proportion to $g_m(\tilde{y}_{t+1}, u_{t+1})$. Given m, 1) if $\Lambda_t^{(m)} \neq I$, then $\Lambda_{t+1} = I$ or $\Lambda_{t+1} = \Lambda_t^{(m)}$, with probability in proportion to $N(\tilde{y}_{t+1}|0, u_{t+1}^{-1}I)(1-\pi_2)$ and $N(\tilde{y}_{t+1}|0, u_{t+1}^{-1}\Lambda_t^{(m)})\pi_2$, respectively (see equation (52)). 2) If $\Lambda_t^{(m)} = I$, then $\Lambda_{t+1} = I$ or $\Lambda_{t+1} \neq I$, with probability in proportion to $N(\tilde{y}_{t+1}|0, u_{t+1}^{-1}I)\pi_1$ and $\frac{\Pi^{-\frac{k}{2}}u_{t+1}^{\frac{k}{2}}|Q_0|^{\frac{\nu}{2}}\prod_{j=1}^k\Gamma(\frac{\nu+2-j}{2})}{|\tilde{Q}_{t+1}|^{\frac{1+\nu}{2}}\prod_{j=1}^k\Gamma(\frac{\nu+1-j}{2})}(1-\pi_1)$, respectively (equation (49) and (50)); given $\Lambda_{t+1} \neq I$, sample Λ_{t+1} from the distribution $W^{-1}(\nu+1, \tilde{Q}_{t+1})$, where $\tilde{Q}_{t+1} = u_{t+1}\tilde{y}_{t+1}\tilde{y}_{t+1}' + Q_0$.

Use the above method to sample M draws of $(\Lambda_{t+1}^{(j)}, m^{(j)}, u_{t+1}^{(j)})$, $j = 1, \ldots, M$. $\{(\Lambda_{t+1}^{(j)}, u_{t+1}^{(j)})\}_{j=1}^M$ constitute a sample from $f(\Lambda_{t+1}, u_{t+1}|\tilde{\mathbf{Y}}_{t+1})$ and are the desired particles.

Given the above filtering procedure, the likelihood is estimated using the decomposition $f(\tilde{\mathbf{Y}}|\Psi_1) = \prod_{t=1}^T f(\tilde{y}_t|\tilde{\mathbf{Y}}_{t-1}, \Psi_1)$:

1. At time t, obtain M particles of $\Lambda_{t-1}^{(j)}$, $j = 1, \ldots, M$.

2. For each $\Lambda_{t-1}^{(j)}$, sample $\Lambda_t^{(j)}|\Lambda_{t-1}^{(j)}$ according to the transition distribution defined in (37) and (38).

3. Approximate $f(\tilde{y}_t|\tilde{\mathbf{Y}}_{t-1}, \Psi_1)$ as $\hat{f}(\tilde{y}_t|\tilde{\mathbf{Y}}_{t-1}, \Psi_1) = \frac{1}{M}\sum_{j=1}^{M} t(\tilde{y}_t|0, \Lambda_t^{(j)}, d)$, where $t(.|.,.,.)$ is the density of the Student-t distribution.

Repeat the above steps for each t, and estimate $\log f(\tilde{\mathbf{Y}}|\Psi_1)$ by $\sum_{t=1}^{T}\log\hat{f}(\tilde{y}_t|\tilde{\mathbf{Y}}_{t-1}, \Psi_1)$.

8.3.2 Estimating $f(\Psi|\mathbf{Y})$

To compute the posterior ordinate at some Ψ^*, we use the method of Chib & Jeliazkov (2001). Write

$$\begin{aligned} f(\Psi^*|\mathbf{Y}) &= f(\Theta^*|\mathbf{Y}) \times f(\mu^*|\mathbf{Y}, \Theta^*) \times f(\pi_1^*, \pi_2^*|\mathbf{Y}, \Theta^*, \mu^*) \times f(d^*|\mathbf{Y}, \Theta^*, \mu^*, \pi_1^*, \pi_2^*) \\ &\quad \times f(\nu^*|\mathbf{Y}, \Theta^*, \mu^*, \pi_1^*, \pi_2^*, d^*) \times f(Q_0^*|\mathbf{Y}, \Theta^*, \mu^*, \pi_1^*, \pi_2^*, d^*, \nu^*). \end{aligned} \tag{54}$$

As described in Section 3, each block of the parameters is sampled using an M-H step, except for Q_0, which is sampled using a Gibbs step. To estimate $f(\Theta^*|\mathbf{Y})$, denote by $q(\Theta, \Theta^*)$ the proposal density for the transition from Θ to Θ^* and by $\alpha(\Theta, \Theta^*|\mathbf{Y}, \mu, \nu, Q_0, \mathbf{S}, \mathbf{U})$ the M-H probability to move. Using the results from Chib & Jeliazkov (2001), the ordinate $f(\Theta^*|\mathbf{Y})$ can be expressed as

$$\begin{aligned} &f(\Theta^*|\mathbf{Y}) = \\ &\frac{\int \alpha(\Theta, \Theta^*|\mathbf{Y}, \mu, \nu, Q_0, \mathbf{S}, \mathbf{U}) q(\Theta, \Theta^*) f(\Theta, \mu, \nu, Q_0, \mathbf{S}, \mathbf{U}|\mathbf{Y}) d(\Theta, \mu, \nu, Q_0, \mathbf{S}, \mathbf{U})}{\int \alpha(\Theta^*, \Theta|\mathbf{Y}, \mu, \nu, Q_0, \mathbf{S}, \mathbf{U}) q(\Theta^*, \Theta) f(\mu, \nu, Q_0, \mathbf{S}, \mathbf{U}|\mathbf{Y}, \Theta^*) d(\Theta, \mu, \nu, Q_0, \mathbf{S}, \mathbf{U})} \end{aligned} \tag{55}$$

The numerator can be estimated as

$$\frac{1}{M}\sum_{j=1}^{M} \alpha(\Theta^{(j)}, \Theta^*|\mathbf{Y}, \mu^{(j)}, \nu^{(j)}, Q_0^{(j)}, \mathbf{S}^{(j)}, \mathbf{U}^{(j)}) q(\Theta^{(j)}, \Theta^*), \tag{56}$$

where $(\Theta^{(j)}, \mu^{(j)}, \nu^{(j)}, Q_0^{(j)}, \mathbf{S}^{(j)}, \mathbf{U}^{(j)})$ is from the j^{th} draw of the full MCMC run of the posterior distribution $f(\Psi, \mathbf{S}, \mathbf{U}, \mathbf{\Lambda}|\mathbf{Y})$, which consists of $\{\Theta^{(j)}, \mu^{(j)}, \pi_1^{(j)}, \pi_2^{(j)}, d^{(j)}, \nu^{(j)}, Q_0^{(j)}, \mathbf{S}^{(j)}, \mathbf{U}^{(j)}, \mathbf{\Lambda}^{(j)}\}$. For the denominator, the integral is with respect to the distribution $f(\mu, \nu, Q_0, \mathbf{S}, \mathbf{U}|\mathbf{Y}, \Theta^*) \times q(\Theta^*, \Theta)$. To estimate this integral, fix Θ at Θ^* and conduct a reduced run of another M iterations sampling the conditional posterior distributions of all the state variables $(\mathbf{S}, \mathbf{U}, \mathbf{\Lambda})$ and parameters except Θ. At each iteration of the reduced run, also draw Θ from the proposal density $q(\Theta^*, \Theta)$. The results will provide M draws of $\{\mu^{(l)}, \nu^{(l)}, Q_0^{(l)}, \mathbf{S}^{(l)}, \mathbf{U}^{(l)}, \Theta^{(l)}\}$ from the distribution $f(\mu, \nu, Q_0, \mathbf{S}, \mathbf{U}|\mathbf{Y}, \Theta^*)q(\Theta^*, \Theta)$. Then the denominator is estimated as

$$\frac{1}{M}\sum_{l=1}^{M} \alpha(\Theta^*, \Theta^{(l)}|\mathbf{Y}, \mu^{(l)}, \nu^{(l)}, Q_0^{(l)}, \mathbf{S}^{(l)}, \mathbf{U}^{(l)}). \tag{57}$$

The ordinate $f(\mu^*|\mathbf{Y}, \Theta^*)$ can be expressed as

$$f(\mu^*|\mathbf{Y}, \Theta^*) = \frac{\int \alpha(\mu, \mu^*|\mathbf{Y}, \Theta^*, d, \mathbf{\Lambda}) q(\mu, \mu^*) f(\mu, d, \mathbf{\Lambda}|\mathbf{Y}, \Theta^*) d(\mu, \mathbf{\Lambda}, d)}{\int \alpha(\mu^*, \mu|\mathbf{Y}, \Theta^*, d, \mathbf{\Lambda}) q(\mu^*, \mu) f(d, \mathbf{\Lambda}|\mathbf{Y}, \Theta^*, \mu^*) d(\mu, \mathbf{\Lambda}, d)} \tag{58}$$

The numerator can be estimated as

$$\frac{1}{M}\sum_{j=1}^{M}\alpha(\mu^{(j)},\mu^*|\mathbf{Y},\Theta^*,d^{(j)},\mathbf{\Lambda}^{(j)})q(\mu^{(j)},\mu^*), \tag{59}$$

where $(\mu^{(j)},d^{(j)},\mathbf{\Lambda}^{(j)})$ is from the j^{th} draw of the reduced MCMC run of the posterior distribution with Θ fixed at Θ^*, which were used in the previous step to estimate the denominator in (55). For the denominator in (58), the integral is with respect to $f(d,\mathbf{\Lambda}|\mathbf{Y},\Theta^*,\mu^*)\times q(\mu^*,\mu)$. To estimate this integral, fix Θ at Θ^* and μ at μ^*, and conduct a second reduced run of M iterations sampling the conditional posterior distributions of all the state variables and parameters except Θ and μ. At each iteration, draw μ from the proposal density $q(\mu^*,\mu)$. The results will provide M draws of $(\mu^{(l)},d^{(l)},\mathbf{\Lambda}^{(l)})$ from the distribution $f(d,\mathbf{\Lambda}|\mathbf{Y},\Theta^*,\mu^*)q(\mu^*,\mu)$. The denominator is estimated as

$$\frac{1}{M}\sum_{l=1}^{M}\alpha(\mu^*,\mu^{(l)}|\mathbf{Y},\Theta^*,d^{(l)},\mathbf{\Lambda}^{(l)}). \tag{60}$$

The ordinate $f(\pi_1^*,\pi_2^*|\mathbf{Y},\Theta^*,\mu^*)$ can be expressed as

$$f(\pi_1^*,\pi_2^*|\mathbf{Y},\Theta^*,\mu^*)=\frac{\int\alpha(\pi,\pi^*|\mathbf{S})q(\pi^*|\mathbf{S})f(\pi,\mathbf{S}|\mathbf{Y},\Theta^*,\mu^*)d(\mathbf{S},\pi)}{\int\alpha(\pi^*,\pi|\mathbf{S})q(\pi|\mathbf{S})f(\mathbf{S}|\mathbf{Y},\Theta^*,\mu^*,\pi^*)d(\mathbf{S},\pi)}, \tag{61}$$

where $\pi=\{\pi_1,\pi_2\}$. $q(.|\mathbf{S})$ is the density of the independent proposal which does not use the last iteration in the proposal of a new value of π, but depends on the value of $\mathbf{S}$. The numerator can be estimated as

$$\frac{1}{M}\sum_{j=1}^{M}\alpha(\pi^{(j)},\pi^*|\mathbf{S}^{(j)})q(\pi^*|\mathbf{S}^{(j)}), \tag{62}$$

where $(\pi^{(j)},\mathbf{S}^{(j)})$ is from the j^{th} draw of the second reduced MCMC run of the posterior distribution with Θ fixed at Θ^* and μ fixed at μ^*, used in the previous step to estimate the denominator of (58). To estimate the integral in the denominator of (61), fix Θ,μ,π at Θ^*,μ^*,π^*, respectively, and conduct a third reduced run of M iterations sampling the conditional posterior distributions of all the state variables and parameters except Θ, μ and π. At each iteration l, given the value of $\mathbf{S}^{(l)}$, draw $\pi^{(l)}$ from the proposal density $q(.|\mathbf{S}^{(l)})$. The resulting $(\mathbf{S}^{(l)},\pi^{(l)}),l=1,\ldots,M$ are draws from the distribution $f(\mathbf{S}|\mathbf{Y},\Theta^*,\mu^*,\pi^*)q(\pi|\mathbf{S})$. The denominator is estimated as

$$\frac{1}{M}\sum_{l=1}^{M}\alpha(\pi^*,\pi^{(l)}|\mathbf{S}^{(l)}). \tag{63}$$

The ordinate $f(d^*|\mathbf{Y},\Theta^*,\mu^*,\pi_1^*,\pi_2^*)$ can be expressed as

$$f(d^*|\mathbf{Y},\Theta^*,\mu^*,\pi_1^*,\pi_2^*)=\frac{\int\alpha(d,d^*|\mathbf{U})q(d,d^*)f(d,\mathbf{U}|\mathbf{Y},\Theta^*,\mu^*,\pi^*)d(\mathbf{U},d)}{\int\alpha(d^*,d|\mathbf{U})q(d^*,d)f(\mathbf{U}|\mathbf{Y},\Theta^*,\mu^*,\pi^*,d^*)d(\mathbf{U},d)}, \tag{64}$$

The numerator can be estimated as

$$\frac{1}{M}\sum_{j=1}^{M}\alpha(d^{(j)},d^*|\mathbf{U}^{(j)})q(d^{(j)},d^*), \tag{65}$$

where $(d^{(j)},\mathbf{U}^{(j)})$ is from the j^{th} draw of the third reduced MCMC run of the posterior distribution used in the previous step to estimate the denominator of (61). To estimate the integral in the denominator of (64), fix Θ,μ,π,d at Θ^*,μ^*,π^*,d^*, respectively, and conduct a fourth reduced run of M iterations sampling the conditional posterior distributions of all the state variables and parameters except Θ, μ, π and d. At each iteration l, also draw $d^{(l)}$ from the proposal density $q(d^*,d)$. The resulting $(\mathbf{U}^{(l)},d^{(l)}),l=1,\ldots,M$ are draws from the distribution $f(\mathbf{U}|\mathbf{Y},\Theta^*,\mu^*,\pi^*,d^*)q(d^*,d)$. The denominator is estimated as

$$\frac{1}{M}\sum_{l=1}^{M}\alpha(d^*,d^{(l)}|\mathbf{U}^{(l)}). \tag{66}$$

The ordinate $f(\nu^*|\mathbf{Y},\Theta^*,\mu^*,\pi_1^*,\pi_2^*,d^*)$ can be expressed as

$$f(\nu^*|\mathbf{Y},\Theta^*,\mu^*,\pi_1^*,\pi_2^*,d^*)=\frac{\int\alpha(\nu,\nu^*|\mathbf{Y},\Theta^*,\mu^*,Q_0,\mathbf{S},\mathbf{U})q(\nu,\nu^*)f(\nu,Q_0,\mathbf{S},\mathbf{U}|\mathbf{Y},\Theta^*,\mu^*,\pi^*,d^*)d(\nu,Q_0,\mathbf{S},\mathbf{U})}{\int\alpha(\nu^*,\nu|\mathbf{Y},\Theta^*,\mu^*,Q_0,\mathbf{S},\mathbf{U})q(\nu^*,\nu)f(Q_0,\mathbf{S},\mathbf{U}|\mathbf{Y},\Theta^*,\mu^*,\pi^*,d^*,\nu^*)d(\nu,Q_0,\mathbf{S},\mathbf{U})}. \tag{67}$$

The numerator can be estimated as

$$\frac{1}{M}\sum_{j=1}^{M}\alpha(\nu^{(j)},\nu^*|\mathbf{Y},\Theta^*,\mu^*,Q_0^{(j)},\mathbf{S}^{(j)},\mathbf{U}^{(j)})q(\nu^{(j)},\nu^*), \tag{68}$$

where $(\nu^{(j)},Q_0^{(j)},\mathbf{S}^{(j)},\mathbf{U}^{(j)})$ is from the j^{th} draw of the fourth reduced MCMC run of the posterior distribution used in the previous step to estimate the denominator of (64). To estimate the integral in the denominator of (67), fix Θ,μ,π,d,ν at $\Theta^*,\mu^*,\pi^*,d^*,\nu^*$, respectively, and conduct a fifth reduced run of M iterations sampling the conditional posterior distributions of all the state variables and parameters except Θ, μ, π, d and ν. At each iteration l, also draw $\nu^{(l)}$ from the proposal density $q(\nu^*,\nu)$. The resulting $(\mathbf{U}^{(l)},Q_0^{(l)},\mathbf{S}^{(l)},\nu^{(l)}),l=1,\ldots,M$ are draws from the distribution $f(Q_0,\mathbf{S},\mathbf{U}|\mathbf{Y},\Theta^*,\mu^*,\pi^*,d^*,\nu^*)q(\nu^*,\nu)$. The denominator can be estimated as

$$\frac{1}{M}\sum_{l=1}^{M}\alpha(\nu^*,\nu^{(l)}|\mathbf{Y},\Theta^*,\mu^*,Q_0^{(l)},\mathbf{S}^{(l)},\mathbf{U}^{(l)}). \tag{69}$$

Since the conditional posterior distribution of Q_0 is sampled using a Gibbs step, $f(Q_0^*|\mathbf{Y},\Theta^*,\mu^*,\pi_1^*,\pi_2^*,d^*,\nu^*)$ can be estimated by

$$f(Q_0^*|\mathbf{Y},\Theta^*,\mu^*,\pi_1^*,\pi_2^*,d^*,\nu^*)\approx\frac{1}{M}\sum_{j=1}^{M}f(Q_0^*|\mathbf{\Lambda}^{(j)},\nu^*), \tag{70}$$

where $f(Q_0|\mathbf{\Lambda},\nu)=W_k(Q_0|\gamma,\widetilde{A})$ (both γ and $\widetilde{A}$ depend on ν and $\mathbf{\Lambda}$, see Section 8.2), and $\mathbf{\Lambda}^{(j)}$ is from the j^{th} draw of the fifth reduced MCMC run of the posterior distribution used in the previous step to estimate the denominator of (67).

8.3.3 Estimating $f(\Psi)$

The prior distributions are independent for different blocks of parameters and therefore,

$$f(\Psi^*) = f(\Theta^*)f(\mu^*)f(\pi_1^*, \pi_2^*)f(d^*)f(\nu^*)f(Q_0^*).$$

Evaluating the prior ordinates is straightforward for most blocks, as they are standard distributions such as (truncated) normal or exponential. If we impose the restriction $\pi_1 > \pi_2$, then

$$f(\pi_1, \pi_2) = \frac{f_B(\pi_1)f_B(\pi_2)\mathbb{I}_{\pi_1>\pi_2}}{\int_{\pi_1>\pi_2} f_B(\pi_1)f_B(\pi_2)d(\pi_1, \pi_2)}, \tag{71}$$

where $f_B(.)$ is the density of a beta distribution. To evaluate $f(\pi_1^*, \pi_2^*)$, we need to compute the denominator of equation (71), which is the normalizing constant. We simulate $M_1 = 10000$ pairs of (π_1, π_2), and calculate the ratio of the number of pairs where $\pi_1 > \pi_2$ to M_1, this ratio is a simulation consistent estimate of the denominator.

References

Akhigbe, A., Martin, A. D. & Whyte, A. M. (2005), 'Contagion effects of the world's largest bankruptcy: the case of worldcom', *The Quarterly Review of Economics and Finance* **45**(1), 48–64.

Ang, A. & Bekaert, G. (2004), 'How regimes affect asset allocation', *Financial Analysts Journal* **60**(2), 86–99.

Bauwens, L., Dufays, A. & Rombouts, J. V. (2014), 'Marginal likelihood for markov-switching and change-point garch models', *Journal of Econometrics* **178**, 508–522.

Bauwens, L., Laurent, S. & Rombouts, J. (2006), 'Multivariate GARCH models: a survey', *Journal of applied econometrics* **21**(1), 79–109.

Cappiello, L., Engle, R. & Sheppard, K. (2006), 'Asymmetric dynamics in the correlations of global equity and bond returns', *Journal of Financial econometrics* **4**(4), 537–572.

Chen, R. (2009), Regime switching in volatilities and correlation between stock and bond markets. manuscript, London School of Economics.

Chib, S. (1995), 'Marginal likelihood from the gibbs output', *Journal of the American Statistical Association* **90**(432), 1313–1321.

Chib, S. (1996), 'Calculating posterior distributions and modal estimates in markov mixture models', *Journal of Econometrics* **75**, 79–97.

Chib, S. & Jeliazkov, I. (2001), 'Marginal likelihood from the metropolis-hasting output', *Journal of the American Statistical Association* **96**, 270–281.

Christoffersen, P., Errunza, V., Jacobs, K. & Langlois, H. (2012), 'Is the potential for international diversification disappearing? a dynamic copula approach', *Review of Financial Studies* **25**(12), 3711–3751.

Clements, A. & Silvennoinen, A. (2013), 'Volatility timing: How best to forecast portfolio exposures', *Journal of Empirical Finance* **24**(0), 108 – 115.

De Goeij, P. & Marquering, W. (2004), 'Modeling the conditional covariance between stock and bond returns: A multivariate garch approach', *Journal of Financial Econometrics* **2**(4), 531–564.

Ding, Z. & Engle, R. (2001), 'Large scale conditional covariance matrix modeling, estimation and testing', *NYU Working Paper No. FIN-01-029* .

Dufays, A. (2013), Infinite-state markov-switching for dynamic volatility and correlation models. https://sites.google.com/site/websiteofarnauddufays/research.

Engle, R. (2002), 'Dynamic conditional correlation', *Journal of Business and Economic statistics* **20**(3), 339–350.

Fleming, J., Kirby, C. & Ostdiek, B. (2001), 'The economic value of volatility timing.', *Journal of Finance* **56**, 329–352.

Fleming, J., Kirby, C. & Ostdiek, B. (2003), 'The economic value of volatility timing using realized volatility', *Journal of Financial Economics* **67**(3), 473–509.

Forbes, K. J. & Rigobon, R. (2002), 'No contagion, only interdependence: Measuring stock market comovements', *The Journal of Finance* **57**(5), 2223–2261.

Guidolin, M. & Timmermann, A. (2006), 'An econometric model of nonlinear dynamics in the joint distribution of stock and bond returns', *Journal of Applied Econometrics* **21**(1), 1–22.

Haas, M. & Mittnik, S. (2008), Multivariate regimeswitching GARCH with an application to international stock markets, CFS Working Paper Series 2008/08, Center for Financial Studies (CFS).
URL: *http://ideas.repec.org/p/zbw/cfswop/200808.html*

Haas, M., Mittnik, S. & Paolella, M. (2004), 'A new approach to markov-switching garch models', *Journal of Financial Econometrics* **2**(4), 493–530.

Hillebrand, E. (2005), 'Neglecting parameter changes in GARCH models', *Journal of Econometrics* **129**(1-2), 121 – 138.

Kass, R. E. & Raftery, A. E. (1995), 'Bayes factors', *Journal of the American Statistical Association* **90**(420), 773–795.

Kenourgios, D., Samitas, A. & Paltalidis, N. (2011), 'Financial crises and stock market contagion in a multivariate time-varying asymmetric framework', *Journal of International Financial Markets, Institutions and Money* **21**(1), 92 – 106.

Klaassen, F. (2002), 'Improving garch volatility forecasts with regime-switching garch', *Empirical Economics* **27**(2), 363–394.

Lamoureux, C. G. & Lestrapes, W. D. (1990), 'Persistence in variance, structural change, and the garch model', *Journal of Business & Economic Statistics* **8**(2), 225–234.

Muirhead, R. (1982), *Aspects of multivariate statistical theory*, Vol. 42, Wiley Online Library.

Pelletier, D. (2006), 'Regime switching for dynamic correlations', *Journal of Econometrics* **131**(1), 445–473.

Pitt, M. & Shephard, N. (1999), 'Filtering via simulation: Auxiliary particle filters', *Journal of the American statistical association* **94**(446), 590–599.

Rangel, J. G. & Engle, R. F. (2012), 'The Factor-Spline-GARCH model for high and low frequency correlations', *Journal of Business & Economic Statistics* **30**(1), 109–124.

Silvennoinen, A. & Tersvirta, T. (2009), 'Modeling multivariate autoregressive conditional heteroskedasticity with the double smooth transition conditional correlation garch model', *Journal of Financial Econometrics* **7**(4), 373–411.

Skouras, S. (2007), 'Decisionmetrics: A decision-based approach to econometric modelling', *Journal of Econometrics* **137**(2), 414 – 440.

Table 1: Examples of Effects of Λ on $V = H^{1/2}\Lambda(H^{1/2})'$

Λ			V				increase/decrease			
λ_{11}	λ_{21}	λ_{22}	v_{11}	v_{21}	v_{22}	$\frac{v_{21}}{\sqrt{v_{11}v_{22}}}$	Var_{11}	Cov_{21}	Var_{22}	ρ_{21}
1.00	0.00	1.00	2.00	1.50	3.00	0.61	=	=	=	=
2.76	3.73	7.82	5.52	11.36	28.60	0.90	$\uparrow$	$\uparrow$	$\uparrow$	$\uparrow$
1.23	0.08	1.74	2.46	2.21	5.19	0.62	$\uparrow$	$\uparrow$	$\uparrow$	$\approx$
6.74	−13.15	29.70	13.48	−15.35	25.09	−0.83	$\uparrow$	$\Downarrow$	$\uparrow$	$\Downarrow$
0.24	−0.06	0.30	0.48	0.25	0.66	0.44	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$
0.70	−0.19	0.33	1.40	0.68	0.85	0.62	$\downarrow$	$\downarrow$	$\downarrow$	$\approx$
0.81	0.30	0.32	1.62	1.78	2.36	0.91	$\downarrow$	$\uparrow$	$\downarrow$	$\uparrow$
8.49	−4.26	2.69	16.98	4.49	2.23	0.73	$\uparrow$	$\uparrow$	$\downarrow$	$\uparrow$

This table provides numerical examples of effects of $\Lambda = (\lambda_{ij})$ on $H^{1/2}\Lambda(H^{1/2})'$ in a 2×2 dimension. In all cases $H = \begin{pmatrix} 2 & 1.5 \\ 1.5 & 3 \end{pmatrix}$. $\uparrow$ means increasing, $\downarrow$ means decreasing, $\Downarrow$ means decreasing from positive value to negative value, $\approx$ means approximately equal to.

Table 2: Estimates for Simulated Data I

		VDGARCH-B			VDGARCH		
Parameter	True value	Mean	NSE	0.95 DI	Mean	NSE	0.95 DI
C_{11}	0.04	0.0431	0.0062	(0.0316, 0.0555)	0.0229	0.0028	(0.0171, 0.0282)
C_{21}	0.01	0.0097	0.0011	(0.0077, 0.0119)	−0.0029	0.0027	(−0.0081, 0.0025)
C_{22}	0.03	0.0321	0.0021	(0.0281, 0.0367)	0.0338	0.0027	(0.0283, 0.0391)
a_1	0.12	0.1288	0.0177	(0.0914, 0.1643)	0.2473	0.0107	(0.2281, 0.2707)
a_2	0.10	0.1266	0.0111	(0.1057, 0.1492)	0.3309	0.0092	(0.3141, 0.3499)
b_1	0.97	0.9733	0.0065	(0.9592, 0.9850)	0.9643	0.0031	(0.9572, 0.9700)
b_2	0.98	0.9806	0.0023	(0.9753, 0.9847)	0.9405	0.0030	(0.9341, 0.9458)
π_1	0.99	0.9917	0.0021	(0.9869, 0.9954)			
π_2	0.98	0.9823	0.0041	(0.9735, 0.9897)			

Data are simulated from the VDGARCH-B model for $T = 5000$ observations. This table reports the posterior mean, its numerical standard error (NSE) and a 0.95 density interval (DI) for the parameters of both the VDGARCH-B model and the VDGARCH model using the same simulation data. True values of the parameters are also listed.

Table 3: Estimates for Simulated Data II

		VDGARCH-B			VDGARCH		
Parameter	True value	Mean	NSE	0.95 DI	Mean	NSE	0.95 DI
C_{11}	0.04	0.0356	0.0075	$(0.0231, 0.0524)$	0.0345	0.0042	$(0.0275, 0.0451)$
C_{21}	0.01	0.0122	0.0024	$(0.0089, 0.0191)$	0.0114	0.0019	$(0.0080, 0.0151)$
C_{22}	0.03	0.0332	0.0060	$(0.0236, 0.0485)$	0.0314	0.0054	$(0.0206, 0.0405)$
a_1	0.12	0.1049	0.0139	$(0.0804, 0.1337)$	0.1007	0.0138	$(0.0752, 0.1292)$
a_2	0.10	0.1059	0.0119	$(0.0833, 0.1308)$	0.0944	0.0121	$(0.0709, 0.1175)$
b_1	0.97	0.9754	0.0093	$(0.9526, 0.9882)$	0.9776	0.0053	$(0.9641, 0.9860)$
b_2	0.98	0.9746	0.0084	$(0.9513, 0.9857)$	0.9784	0.0061	$(0.9669, 0.9894)$
π_1	1.	0.9999	0.0001	$(0.9998, 1.0000)$			

Data are simulated from the VDGARCH model for $T = 5000$ observations. The data exclude covariance breakdowns. This table reports the posterior mean, its NSE and a 0.95 DI for the parameters of both the VDGARCH-B model and the VDGARCH model using the same simulation data. True values of the parameters are also listed.

Table 4: Root mean squared error of volatility fit for simulated data

	Data with Breakdowns		Data without Breakdowns	
	VDGARCH-B	VDGARCH	VDGARCH-B	VDGARCH
$RMSE$	0.230	0.630	0.017	0.021
$RMSE_b$	0.375	1.199	-	-

This table reports root mean squared error, $RMSE = \frac{1}{T}\sum_{t=1}^{T}\|V_t - \bar{V}_t\|/\|V_t\|$ for both models, where $\|V\|= \sqrt{\sum_i \sum_j v_{ij}^2}$. V_t is the true volatility matrix at time t and $\bar{V}_t$ is the smoothed value from either model. The root mean squared error over covariance breakdown periods only is $RMSE_b = \frac{1}{T_b}\sum_{s_t=2}\|V_t - \bar{V}_t\|/\|V_t\|$ for both model, where T_b is the number of data points in state 2.

Table 5: Summary statistics for return data

	S&P 500	ten-year bond	one-year bond
Mean	0.0167	0.0131	0.0033
Std. Dev.	1.2050	0.4602	0.0568
Skewness	-0.9856	0.0477	0.7806
Kurtosis	23.2578	7.0883	57.1060
		Sample covariance	
	S&P 500	ten-year bond	one-year bond
S&P 500	1.4521	-0.0704	-0.0087
ten-year bond		0.2118	0.0156
one-year bond			0.0032

This table reports summary statistics for the daily excess return on the S&P 500 index, the ten-year bond and the one-year bond. All returns are in percentage. Total number of observations is 6244.

Table 6: Covariance Breakdown Model Estimates with Gaussian innovations

Parameter	Mean	NSE	0.95 DI
π_1	0.9241	0.0088	$(0.9058, 0.9400)$
π_2	0.9145	0.0095	$(0.8938, 0.9317)$
C_{11}	0.0405	0.0056	$(0.0292, 0.0515)$
C_{21}	0.0041	0.0025	$(-0.0008, 0.0092)$
C_{31}	0.0003	0.0004	$(-0.0002, 0.0008)$
C_{22}	0.0221	0.0029	$(0.0165, 0.0276)$
C_{32}	0.0019	0.0002	$(0.0015, 0.0025)$
C_{33}	0.0015	0.0002	$(0.0012, 0.0019)$
a_1	0.1149	0.0067	$(0.1023, 0.1282)$
a_2	0.1071	0.0076	$(0.0929, 0.1246)$
a_3	0.090	0.0057	$(0.0793, 0.1008)$
b_1	0.9902	0.0009	$(0.9882, 0.9920)$
b_2	0.9901	0.0013	$(0.9868, 0.9924)$
b_3	0.9896	0.0012	$(0.9873, 0.9918)$
μ_1	0.0552	0.0091	$(0.0370, 0.0730)$
μ_2	0.0139	0.0041	$(0.0057, 0.0218)$
μ_3	0.0017	0.0004	$(0.0009, 0.0025)$
$Q_{0,11}$	5.7962	0.6241	$(4.6433, 7.0921)$
$Q_{0,21}$	0.4587	0.3204	$(-0.1740, 0.1088)$
$Q_{0,31}$	0.5177	0.3489	$(-0.1428, -0.1229)$
$Q_{0,22}$	6.7930	0.7765	$(5.3410, 8.3719)$
$Q_{0,32}$	1.5321	0.4710	$(0.6388, 2.4738)$
$Q_{0,33}$	10.2318	1.3692	$(7.7761, 13.1195)$
ν	6.7510	0.3630	$(6.0623, 7.4781)$

This table reports the posterior mean, the numerical standard error (NSE), a 0.95 density interval (DI) for the VDGARCH-B model.

Table 7: Empirical Posterior Distribution of State 2 Duration

VDGARCH-B							
duration	$=5$	$(5, 30]$	$(30, 60]$	$(60, 90]$	$(90, 120]$	$(120, 150]$	>150
frequency	29	147	18	2	3	0	0
VDGARCH-t-B							
duration	$=5$	$(5, 30]$	$(30, 60]$	$(60, 90]$	$(90, 120]$	$(120, 150]$	>150
frequency	2	16	12	4	5	3	2

This table reports the empirical posterior distribution of the duration of the breakdown periods identified by the covariance breakdown models for the return data. The top panel shows the results using Gaussian assumption (VDGARCH-B) for the return innovation and the bottom panel shows the results using a Student-t (VDGARCH-t-B) assumption. The frequency of covariance breakdowns for each interval is rounded to the nearest integer.

Table 8: Parameter Estimates with Student-t Innovations

	VDGARCH-t-B			VDGARCH-t		
Parameter	Mean	NSE	0.95 DI	Mean	NSE	0.95 DI
π_1	0.9858	0.0039	$(0.9775, 0.9920)$			
π_2	0.9722	0.0043	$(0.9629, 0.9798)$			
C_{11}	0.0365	0.0067	$(0.0234, 0.0492)$	0.0555	0.0062	$(0.0428, 0.0674)$
C_{21}	0.0114	0.0034	$(0.0053, 0.0182)$	0.0076	0.0024	$(0.0031, 0.0123)$
C_{31}	0.0005	0.0003	$(0.0000, 0.0011)$	0.0005	0.0003	$(-0.0001, 0.0011)$
C_{22}	0.0273	0.0035	$(0.0200, 0.0345)$	0.0332	0.0026	$(0.0280, 0.0384)$
C_{32}	0.0019	0.0003	$(0.0014, 0.0025)$	0.0034	0.0003	$(0.0027, 0.0035)$
C_{33}	0.0022	0.0003	$(0.0013, 0.0028)$	0.0030	0.0002	$(0.0025, 0.0035)$
a_1	0.1174	0.0093	$(0.1025, 0.1379)$	0.1701	0.0070	$(0.1577, 0.1843)$
a_2	0.1153	0.0099	$(0.0972, 0.1343)$	0.1773	0.0065	$(0.1643, 0.1901)$
a_3	0.0998	0.0078	$(0.0856, 0.1139)$	0.1982	0.0090	$(0.1823, 0.2178)$
b_1	0.9881	0.0016	$(0.9844, 0.9909)$	0.9768	0.0018	$(0.9732, 0.9800)$
b_2	0.9860	0.0022	$(0.9812, 0.9899)$	0.9723	0.0019	$(0.9684, 0.9761)$
b_3	0.9889	0.0015	$(0.9862, 0.9918)$	0.9616	0.0033	$(0.9519, 0.9668)$
μ_1	0.0602	0.0091	$(0.0427, 0.0778)$	0.0607	0.0087	$(0.0440, 0.0778)$
μ_2	0.0161	0.0042	$(0.0080, 0.0243)$	0.0144	0.0042	$(0.0060, 0.0224)$
μ_3	0.0019	0.0004	$(0.0011, 0.0027)$	0.0018	0.0004	$(0.0008, 0.0026)$
d	6.2469	0.2883	$(5.7121, 6.8523)$	5.6075	0.2211	$(5.1774, 6.0500)$
$Q_{0,11}$	8.6365	1.4402	$(5.9913, 11.5989)$			
$Q_{0,21}$	-1.1972	0.5884	$(-2.3989, -0.0708)$			
$Q_{0,31}$	0.3503	0.3916	$(-0.4015, 1.1265)$			
$Q_{0,22}$	6.9727	1.0908	$(5.0202, 9.2310)$			
$Q_{0,32}$	-0.1558	0.4170	$(-0.9862, 0.6448)$			
$Q_{0,33}$	4.1472	0.6480	$(2.9667, 5.5206)$			
ν	8.8174	1.2311	$(6.7103, 11.3088)$			

This table reports the posterior mean, the numerical standard error (NSE), a 0.95 density interval (DI) for the parameters of the VDGARCH-t-B and the VDGARCH-t models. Both models assume Student-t innovations.

Table 9: Posterior Distribution of $\log(|E[\Lambda_t]|) = \log|Q_0/(\nu-k-1)|$, VDGARCH-t-B model

Mean	Median	Stdev	0.95 DI
0.5808	0.5996	0.43547	(-0.2960,1.3800)

This table reports posterior summary statistics on the impact of expected covariance breakdowns as measured by $\log(|E[\Lambda_t]|)$.

Table 10: Model Comparison for VDGARCH

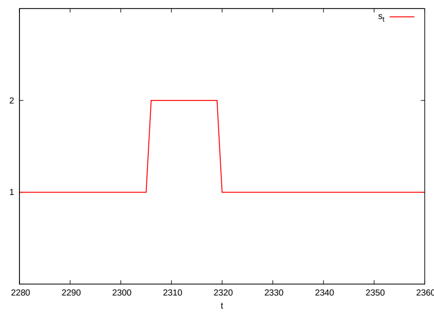
Model	Log Marginal Likelihood	Log(BF)
VDGARCH	−81.01	1640.82
VDGARCH-B	1251.78	308.03
VDGARCH-t	1500.23	59.58
VDGARCH-t-B	1559.81	

Log(BF) is the Log-Bayes factor for Student-t Breakdown (VDGARCH-t-B) model against each alternative.

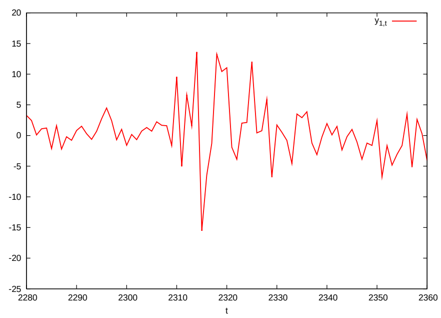
Table 11: Portfolio Performance Covariance Timing Strategies

	VDGARCH-t-B			VDGARCH-t			Δ_q		Δ_e	
μ_0	$\hat{\mu}_p$	$\hat{\sigma}_p$	SR	$\hat{\mu}_p$	$\hat{\sigma}_p$	SR	$\tau = 1$	$\tau = 10$	$\tau = 1$	$\tau = 10$
Panel A: out-of-sample 1 (20070103 to 20111230)										
0.01	0.0100	0.1217	0.0821	0.0093	0.1175	0.0791	17.6	16.4	17.6	16.5
0.03							52.0	41.6	52.0	41.9
0.05							85.4	56.3	85.4	57.8
0.07							117.8	60.6	117.8	64.3
0.09							149.0	54.3	149.9	61.1
Panel B: out-of-sample 2 (20040102 to 20111230)										
0.01	0.0035	0.1065	0.0328	0.0027	0.1034	0.0261	17.8	17.1	17.8	17.1
0.03							53.0	46.3	53.0	46.5
0.05							87.4	69.0	87.5	69.9
0.07							121.3	85.2	121.4	87.4
0.09							154.5	94.8	154.5	98.6

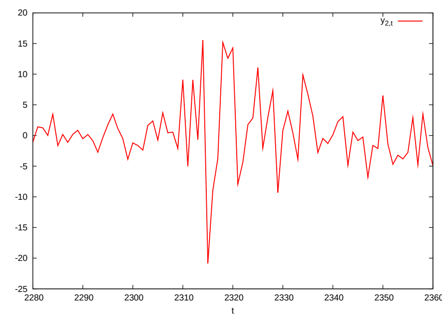
μ_0 is the required daily portfolio excess return in percentage. $\hat{\mu}_p$ and $\hat{\sigma}_p$ are the sample mean and the standard deviation of the realized portfolio returns (in percentage) over the out-of-samples, respectively. SR is the Sharpe ratio defined as $\hat{\mu}_p/\hat{\sigma}_p$. Δ_q is the annualized (assuming 252 trading days a year) basis point fee an investor with quadratic utility would be willing to pay to switch from the VDGARCH-t model to the covariance breakdown model (VDGARCH-t-B). Δ_e is the annualized basis point fee for the same switching associated with negative exponential utility. A positive value of Δ_q or Δ_e means that switching from the VDGARCH-t model to the VDGARCH-t-B model will generate extra economic benefit for the investor. τ is the coefficient of risk aversion. We only report $\hat{\mu}_p$, $\hat{\sigma}_p$ and SR for the case $\mu_0 = 0.01\%$, this is because since the weights are linear in μ_0 (see Eq. (26)), so are the portfolio returns and hence their sample mean and standard deviation. The Sharpe ratio will be the same for portfolio returns with different μ_0 but with the same volatility matrix forecasts.



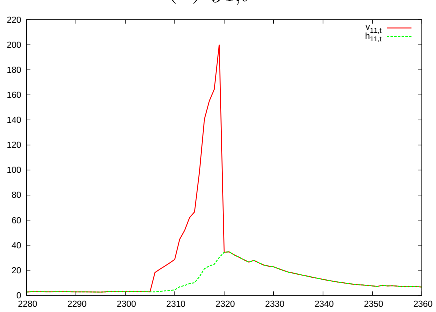
(a) States



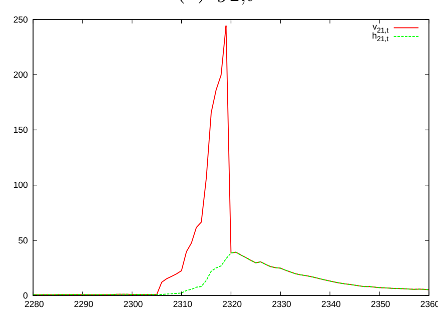
(b) $y_{1,t}$



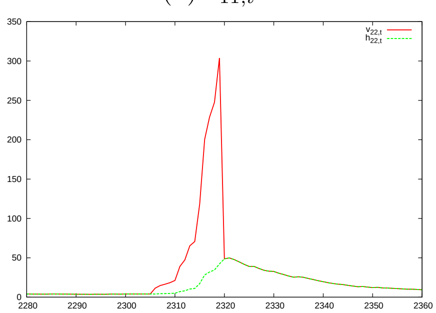
(c) $y_{2,t}$



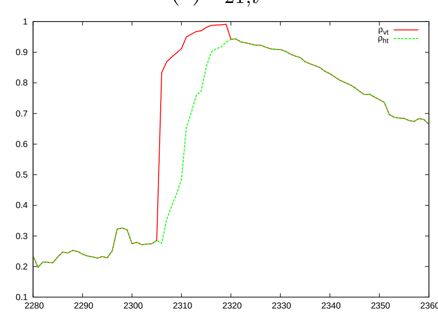
(d) $v_{11,t}$



(e) $v_{21,t}$

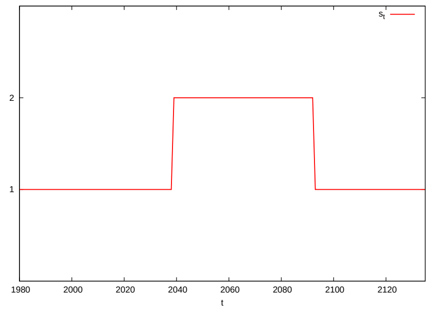


(f) $v_{22,t}$

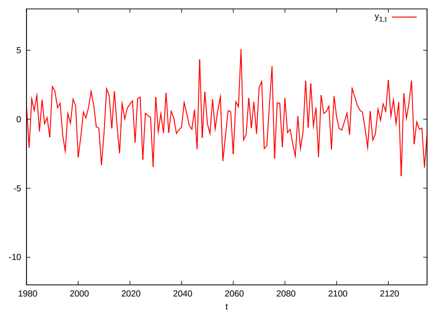


(g) $\rho_{21,t}$

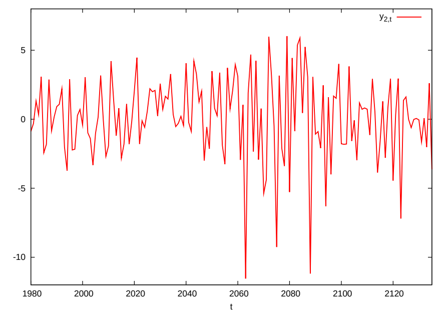
Figure 1: Simulation example: episode 1



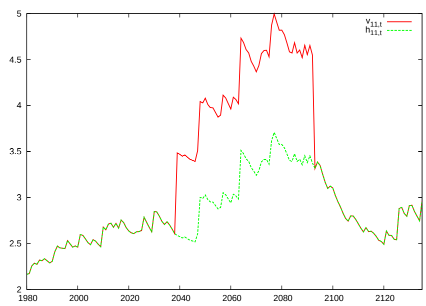
(a) States



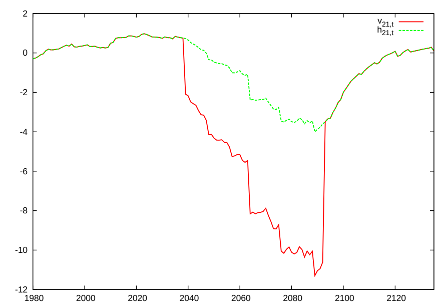
(b) $y_{1,t}$



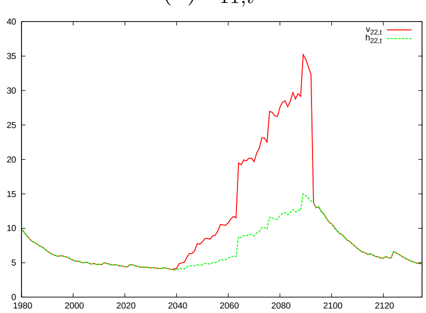
(c) $y_{2,t}$



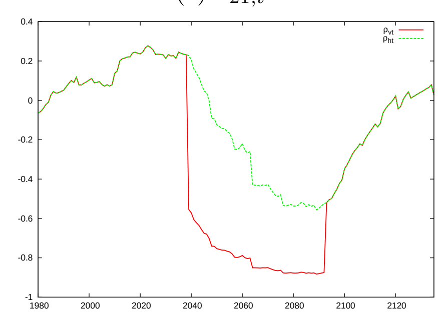
(d) $v_{11,t}$



(e) $v_{21,t}$

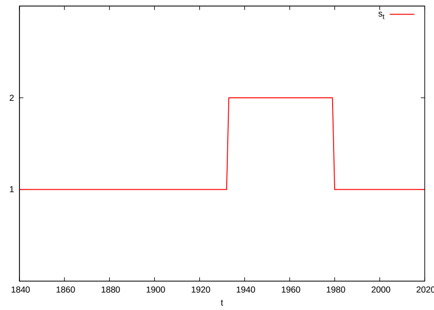


(f) $v_{22,t}$

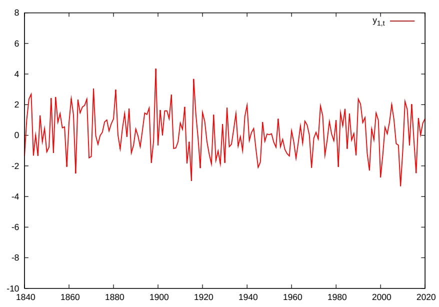


(g) $\rho_{21,t}$

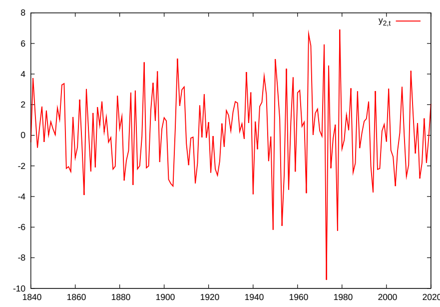
Figure 2: Simulation example: episode 2



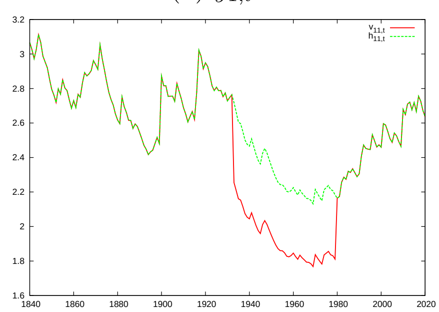
(a) States



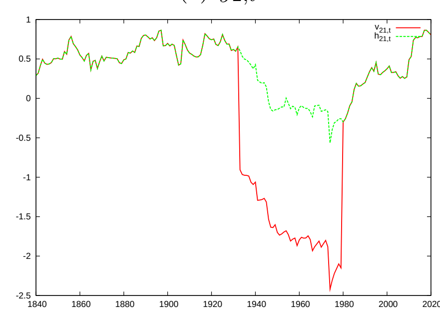
(b) $y_{1,t}$



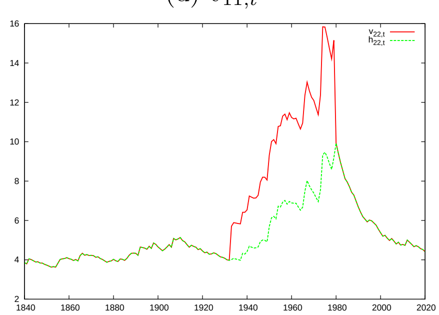
(c) $y_{2,t}$



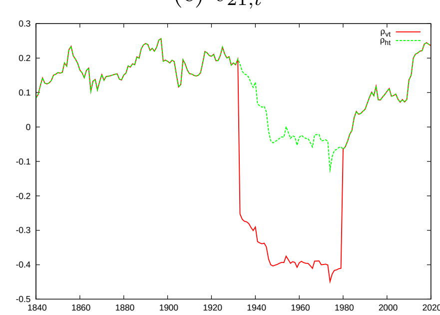
(d) $v_{11,t}$



(e) $v_{21,t}$



(f) $v_{22,t}$



(g) $\rho_{21,t}$

Figure 3: Simulation example: episode 3

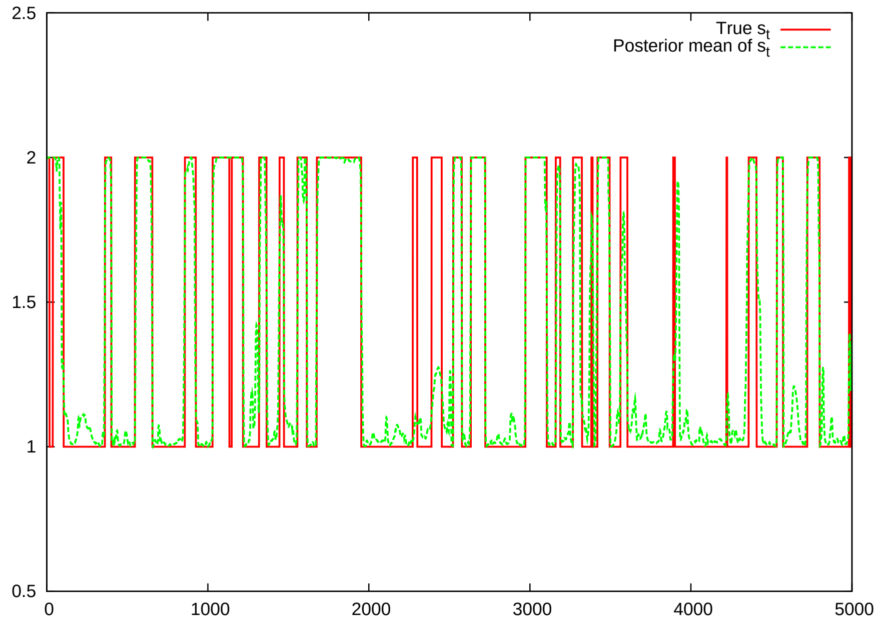
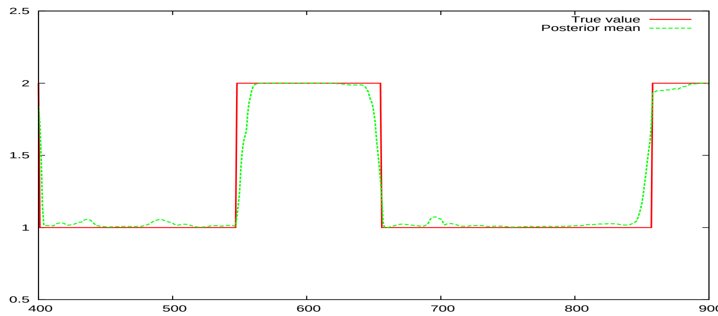
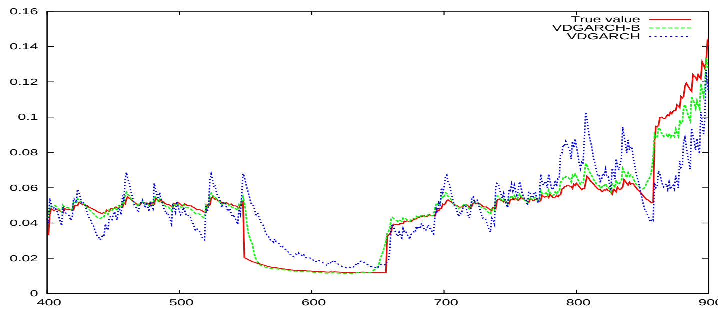


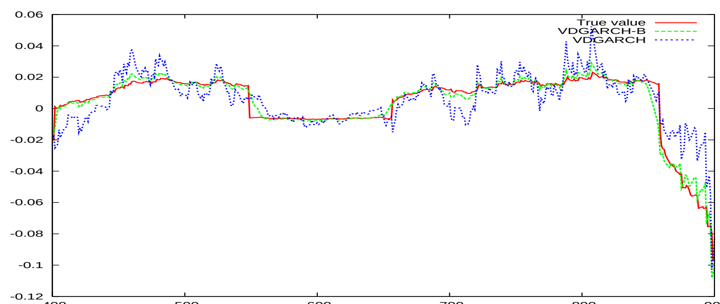
Figure 4: Model inference using simulated data: posterior mean of s_t compared to the true values



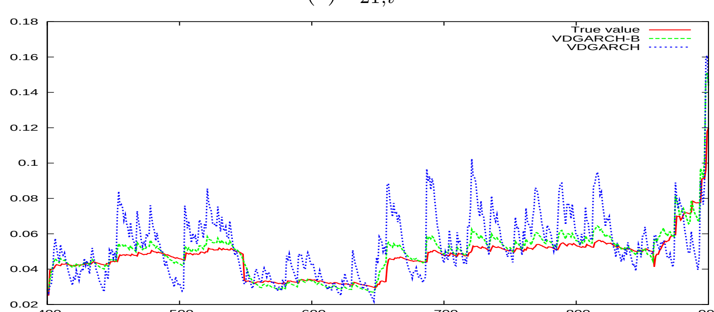
(a) s_t



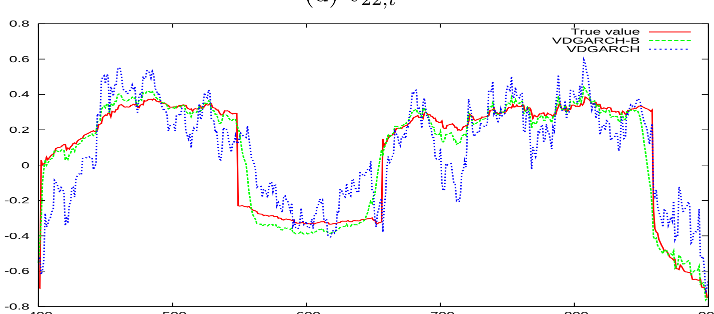
(b) $v_{11,t}$



(c) $v_{21,t}$

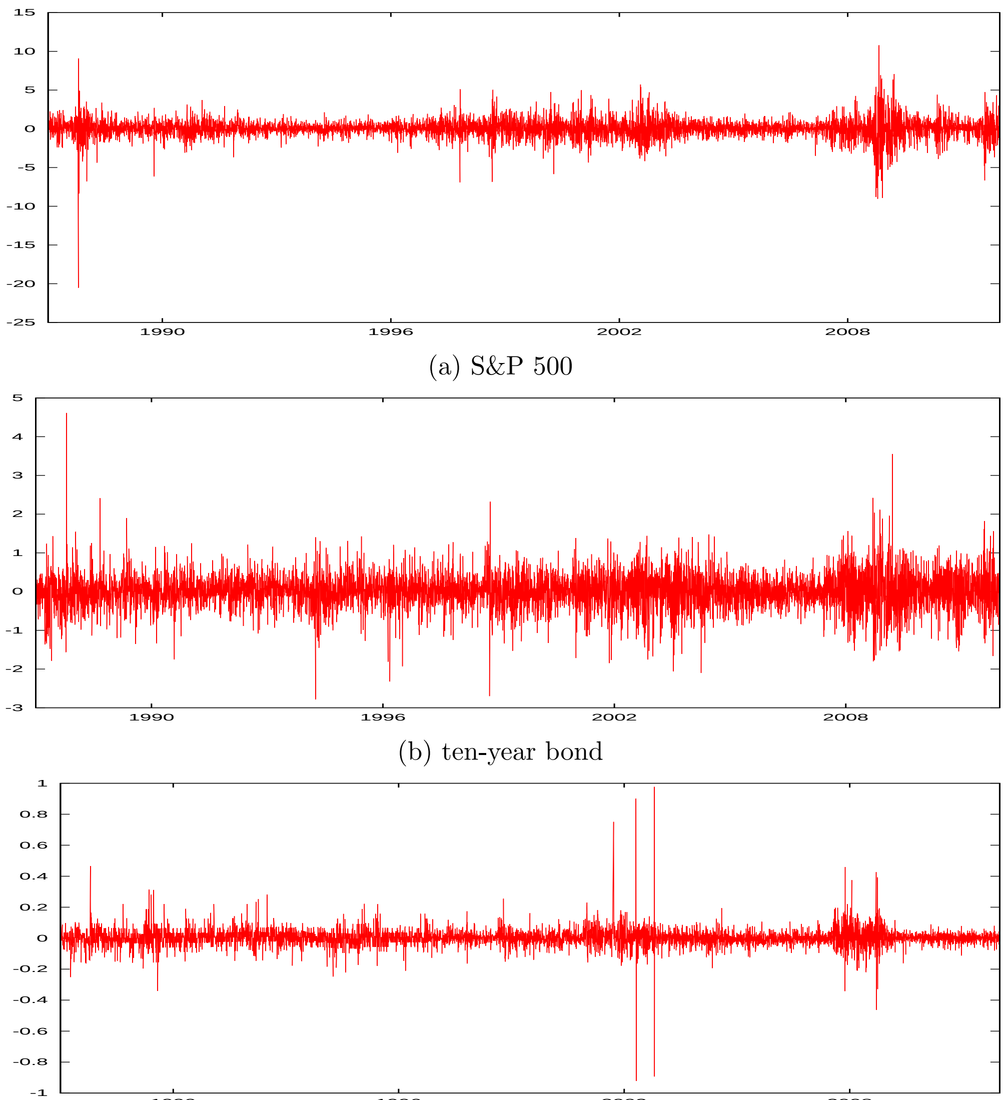


(d) $v_{22,t}$



(e) $\rho_{21,t}$

Figure 5: VDGARCH-B vs VDGARCH model using simulated data with breakdowns: smoothed volatility compared to the true values

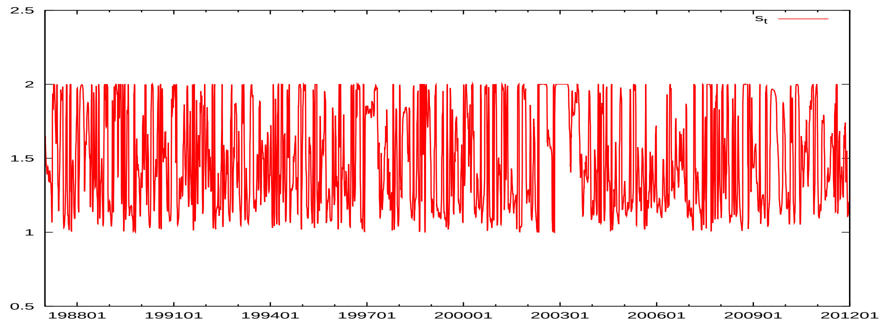


(a) S&P 500

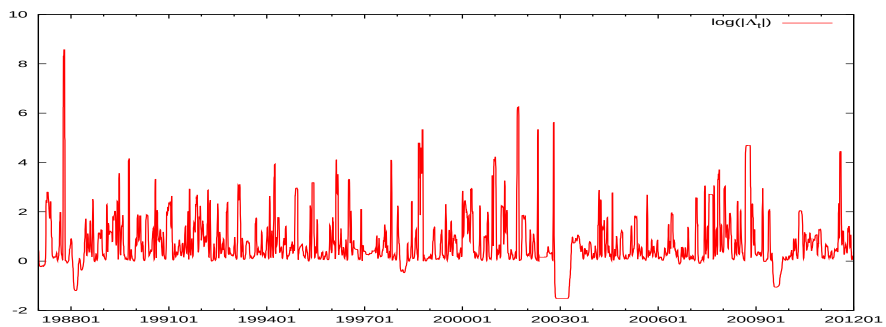
(b) ten-year bond

(c) one-year bond

Figure 6: Daily excess returns in percentage

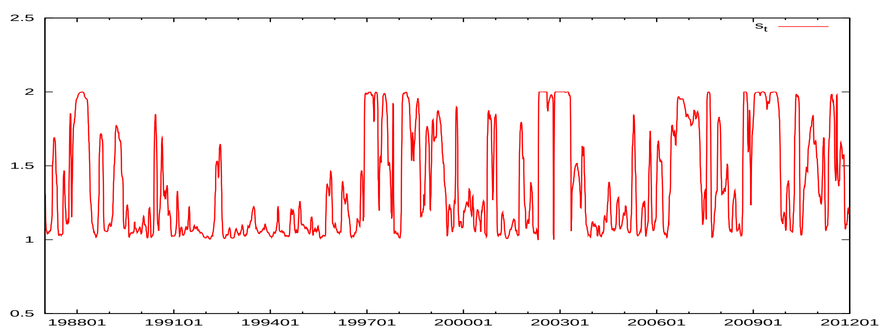


(a) s_t

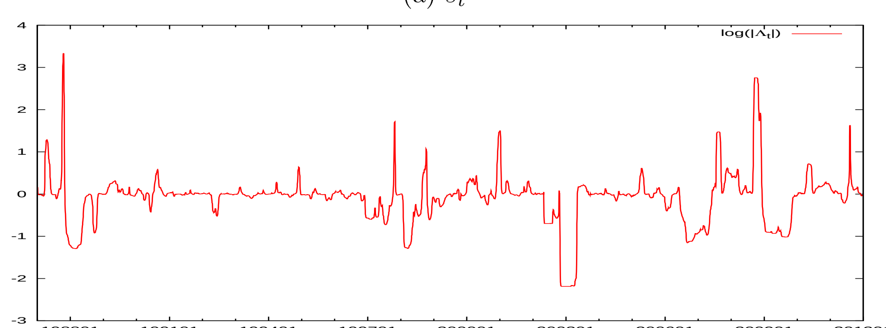


(b) $\log(|\Lambda_t|)$

Figure 7: Posterior mean of s_t and $\log(|\Lambda_t|)$ with normal innovations(VDGARCH-B)

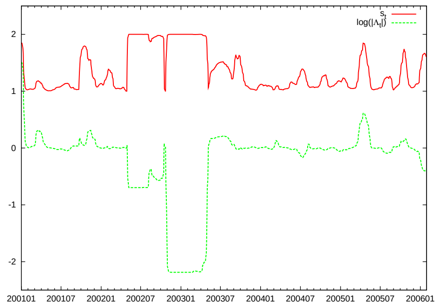


(a) s_t

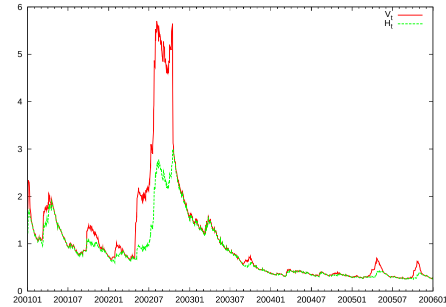


(b) $\log(|\Lambda_t|)$

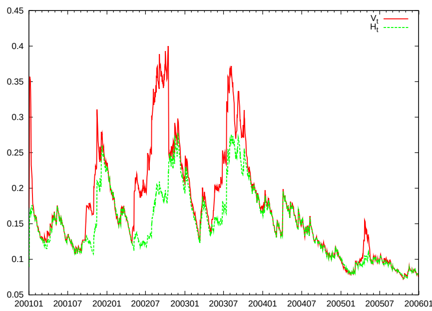
Figure 8: Posterior mean of s_t and $\log(|\Lambda_t|)$ with t-innovations (VDGARCH-t-B)



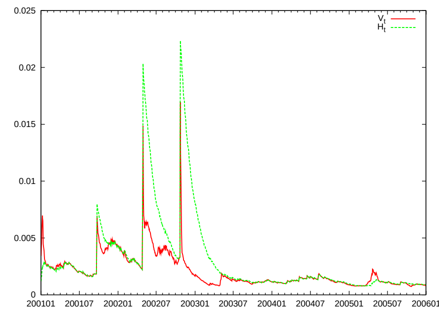
(a) s_t and $\log(|\Lambda_t|)$



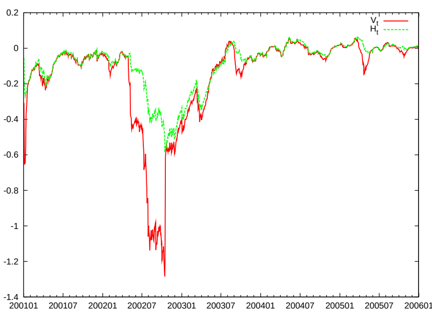
(b) Variance: S&P



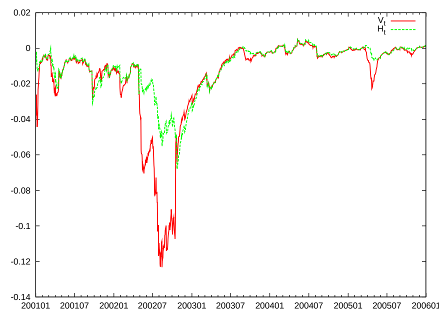
(c) Variance: 10-year bond



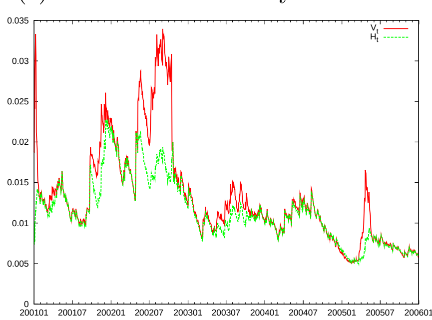
(d) Variance: 1-year bond



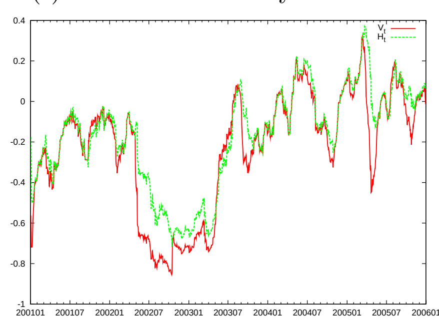
(e) Cov.: S&P and 10-year bond



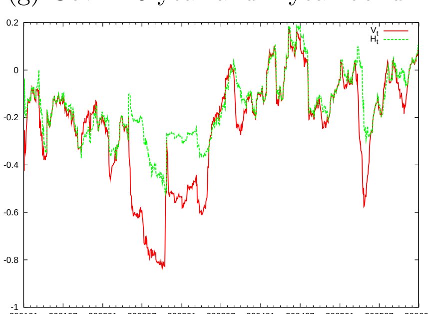
(f) Cov.: S&P and 1-year bond



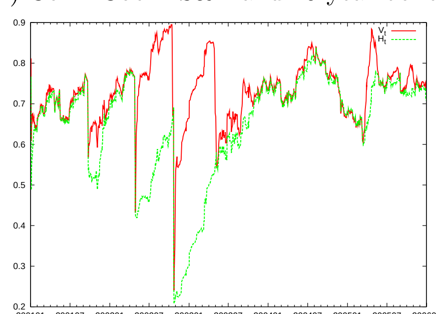
(g) Cov.: 10-year and 1-year bond



(h) Corr. Coef.: S&P and 10-year bond

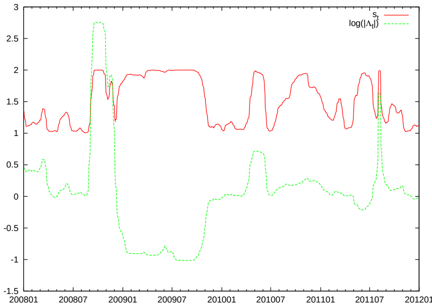


(i) Corr. Coef.: S&P and 1-year bond

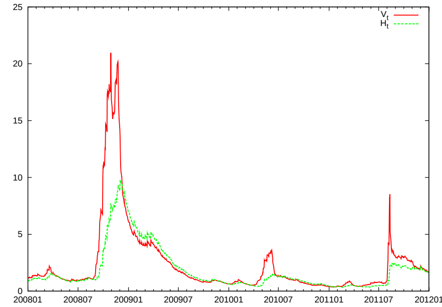


(j) Corr. Coef.: 10-year and 1-year bond

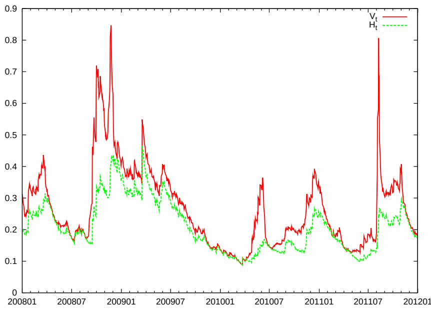
Figure 9: Volatility components $V_t = H_t^{1/2}\Lambda_t(H_t^{1/2})'$ and H_t: VDGARCH-t-B model (2001/01/02-2005/12/30)



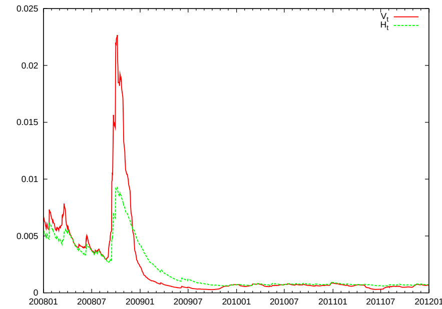
(a) s_t and $\log(|\Lambda_t|)$



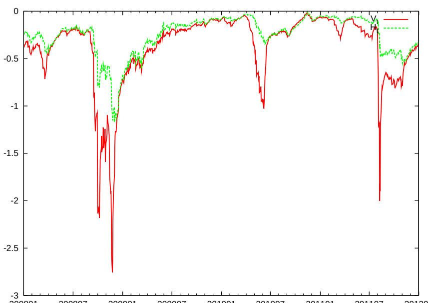
(b) Variance: S&P



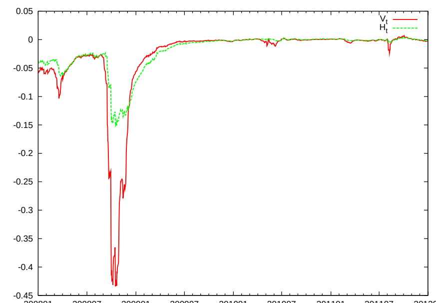
(c) Variance: 10-year bond



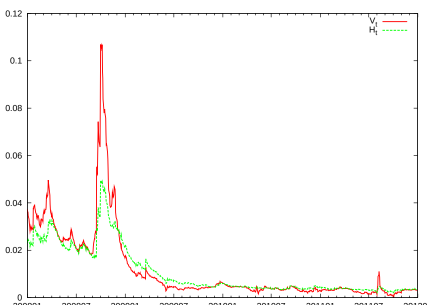
(d) Variance: 1-year bond



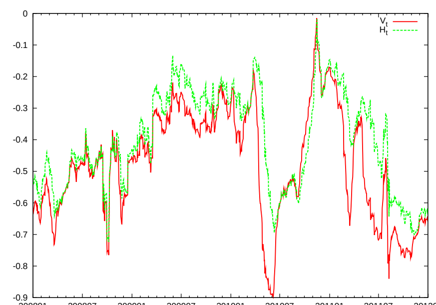
(e) Cov.: S&P and 10-year bond



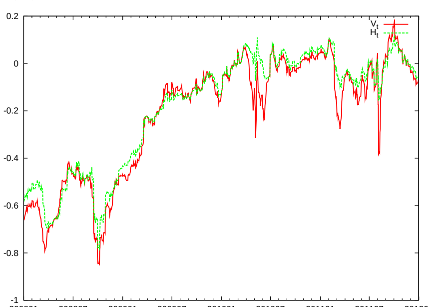
(f) Cov.: S&P and 1-year bond



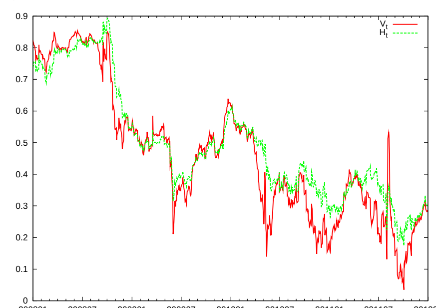
(g) Cov.: 10-year and 1-year bond



(h) Corr. Coef.: S&P and 10-year bond

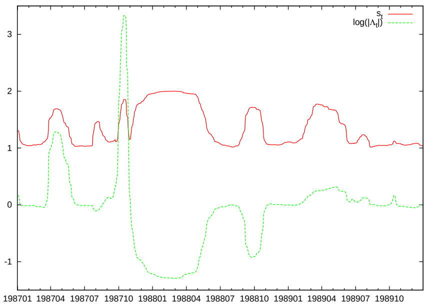


(i) Corr. Coef.: S&P and 1-year bond

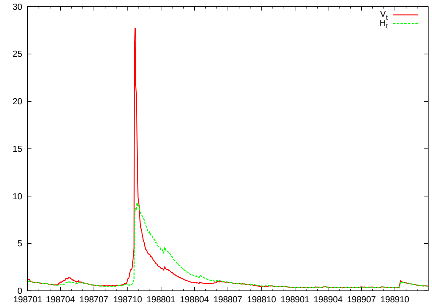


(j) Corr. Coef.: 10-year and 1-year bond

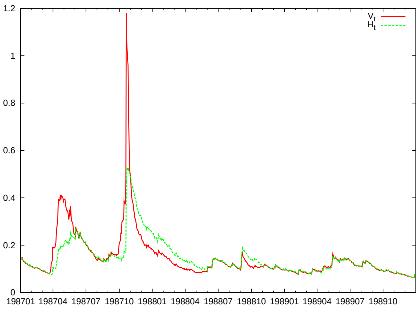
Figure 10: Volatility components $V_t = H_t^{1/2}\Lambda_t(H_t^{1/2})'$ and H_t: VDGARCH-t-B model (2008/01/02-2011/12/30)



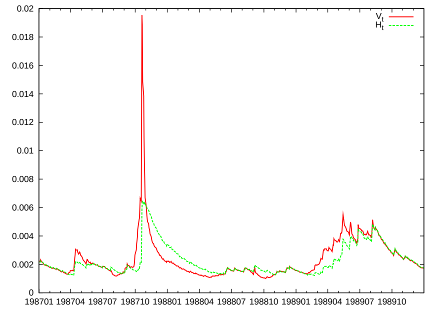
(a) s_t and $\log(|\Lambda_t|)$



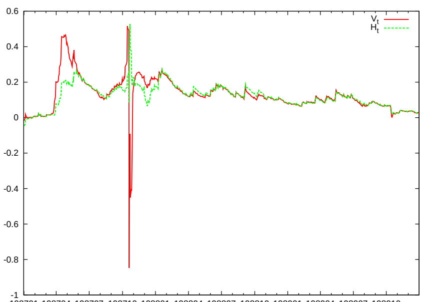
(b) Variance: S&P



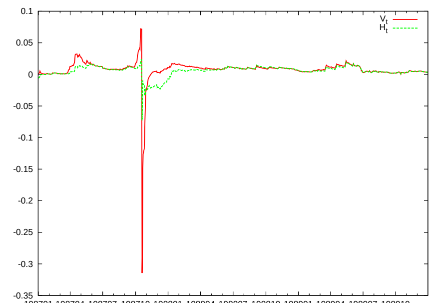
(c) Variance: 10-year bond



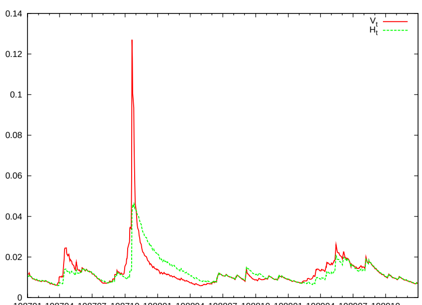
(d) Variance: 1-year bond



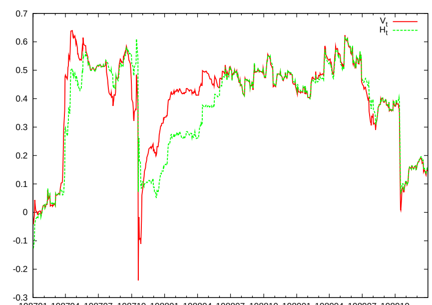
(e) Cov.: S&P and 10-year bond



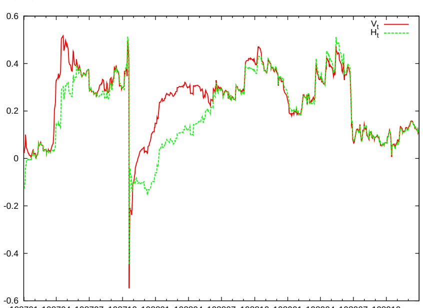
(f) Cov.: S&P and 1-year bond



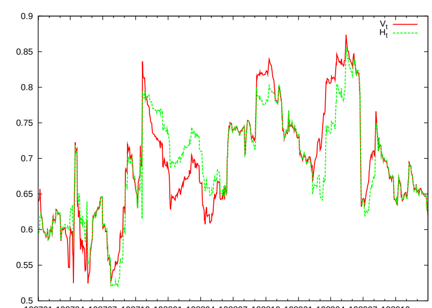
(g) Cov.: 10-year and 1-year bond



(h) Corr. Coef.: S&P and 10-year bond



(i) Corr. Coef.: S&P and 1-year bond



(j) Corr. Coef.: 10-year and 1-year bond

Figure 11: Volatility components $V_t = H_t^{1/2}\Lambda_t(H_t^{1/2})'$ and H_t: VDGARCH-t-B model (1987/01/02-1989/12/30)