Journal of Econometrics 194 (2016) 1–23

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Modeling covariance breakdowns in multivariate GARCH[☆]

Xin Jin[a,b], John M. Maheu[c,*]

[a] School of Economics, Shanghai University of Finance and Economics, Shanghai 200433, China
[b] Key Laboratory of Mathematical Economic (SUFE), Ministry of Education, Shanghai 200433, China
[c] DeGroote School of Business, McMaster University, 1280 Main Street West, Hamilton, ON, Canada, L8S4M4

ARTICLE INFO

Article history:
Received 15 April 2014
Received in revised form
4 February 2016
Accepted 28 March 2016
Available online 20 April 2016

JEL classification:
C32
C51
C58

Keywords:
Correlation breakdown
Marginal likelihood
Particle filter
Markov chain
Generalized variance

ABSTRACT

This paper proposes a flexible way of modeling dynamic heterogeneous covariance breakdowns in multivariate GARCH models through a 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 and show how to compute the marginal likelihood. Applied to daily stock market and bond market data, we identify a number of different covariance breakdowns which leads to a significant improvement in the marginal likelihood and gains in portfolio choice.

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 and 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 and 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 and Teräsvirta (2009) apply the smooth transition modeling approach to conditional correlations. Other regime-switching approaches include Ang and Bekaert (2004), Guidolin and Timmermann (2006) and Pelletier (2006).

In contrast to the literature, which has tended to focus on correlation breakdowns, we investigate breakdowns in each

☆ The paper has been improved by comments from an Associate Editor and two anonymous referees. 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. Jin is supported by Program for Changjiang Scholars and Innovative Research Team in SUFE. Maheu is grateful to the SSHRC (grant number 435-2014-2015) for financial support.

* Corresponding author.
E-mail addresses: jin.xin@mail.shufe.edu.cn (X. Jin), maheujm@mcmaster.ca (J.M. Maheu).

http://dx.doi.org/10.1016/j.jeconom.2016.03.003
0304-4076/

[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 and Lestrapes, 1990; Hillebrand, 2005).

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

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 and 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] This stochastic scale matrix can change several times over the course of a single breakdown. This approach is very flexible and allows a single breakdown to display different characteristics over time 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 as well as a version of that model subject to Markov switching. Bayes factors strongly support the inclusion of covariance breakdowns. The volatility dynamics during breakdown periods are very different for the 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.

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

Estimates indicate that covariance breakdowns occur 42% of the time and their expected duration is 1.5 months in our sample. 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 and Engle (2001) and the dynamic conditional correlation (DCC) by Engle (2002). See Bauwens et al. (2006) for a review.

The dynamics of Λ_t depend on a latent discrete state variable $s_t \in \{1, 2, 3\}$. s_t follows a Markov chain whose transition matrix is represented as

$$P = \begin{pmatrix} \pi_1 & 1-\pi_1 & 0 \\ (1-\pi_2)\pi_4 & \pi_2 & (1-\pi_2)(1-\pi_4) \\ (1-\pi_3)\pi_5 & (1-\pi_3)(1-\pi_5) & \pi_3 \end{pmatrix}, \tag{2}$$

with each $\pi_i \in [0, 1]$, $i = 1, \ldots, 5$ being a free parameter. According to (2), moving directly from state 1 to state 3 is prohibited but all other moves are possible. While in state 2, the probability of staying is π_2; conditional on leaving, the probability of moving into state 1 is π_4. Similarly, while in state 3, the probability of staying is π_3; conditional on leaving, the probability of moving into state 1 is π_5 and the probability of moving to state 2 is $(1-\pi_5)$.

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

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

Thus, when $s_t = 1$, Λ_t is the identity matrix. If s_t does not change, neither does Λ_t. Whenever s_t switches into 2 or 3, Λ_t is a new stochastic draw from G_0.

s_t divides the sample path into periods of *normal* states ($s_t = 1$) and periods of *covariance breakdown* states ($s_t = 2$ or $s_t = 3$). Switches out of state 1 and back into state 1 delineate a covariance breakdown. Therefore, a breakdown can be characterized by a sequence of states all equal to 2 with one associated Λ_t (1,2,2,2,2,1) or by alternating between states 2 and 3 (1,2,2,3,3,3,2,2,1) along with a different draw for Λ_t every time the state switches

[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.

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	↑	↑	↑	↑
1.23	0.08	1.74	2.46	2.21	5.19	0.62	↑	↑	↑	≈
6.74	−13.15	29.70	13.48	−15.35	25.09	−0.83	↑	⇓	↑	⇓
0.24	−0.06	0.30	0.48	0.25	0.66	0.44	↓	↓	↓	↓
0.70	−0.19	0.33	1.40	0.68	0.85	0.62	↓	↓	↓	≈
0.81	0.30	0.32	1.62	1.78	2.36	0.91	↓	↑	↓	↑
8.49	−4.26	2.69	16.98	4.49	2.23	0.73	↑	↑	↓	↑

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 $H = \begin{pmatrix} 2 & 1.5 \\ 1.5 & 3 \end{pmatrix}$. ↑ means increasing, ↓ means decreasing, ⇓ means decreasing from positive value to negative value, ≈ means approximately equal to.

between 2 and 3. In this way we can capture breakdowns that have one feature (e.g. increase in conditional variances) along with more complicated breakdowns that display several features (e.g. increase in conditional variances followed by a decrease in conditional variances and covariances). To identify states a breakdown always begins by moving from state 1 to state 2.

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 current (state-2) breakdown continues until either s_t switches back to 1, in which case we are back in the normal state where the volatility once again coincides with H_t, or, s_t changes from 2 to 3, in which case a new phase of the breakdown occurs with a new draw of Λ_t from G_0, so the volatility process will still be different from H_t, but under a different pattern from the previous phase of the breakdown. State 3 ends whenever the volatility dynamics return to the normal state ($s_t = 1$), or directly enter into another new phase of the breakdown ($s_t = 2$) which features another new Λ_t. And so on. Thus, by entertaining two breakdown states and permitting s_t to switch from one to another, the model allows a series of different Λ_t to affect the conditional covariance without a normal period in between.[6]

A single breakdown period may feature different Λ_t. On the other hand, different breakdown episodes will differ and have their unique break patterns, e.g. increased variability or decreased variability, etc., as each breakdown period is characterized by a set of unique Λ_t drawn from G_0 and a unique run of states 2 and 3. Even if two breakdown periods have an identical run of states 2 and 3 they will differ in their impact on the conditional second moments since the draws of Λ_t will differ.

Therefore, even though s_t is a three-state Markov chain, the breakdown periods are "heterogeneous" and there could be infinitely many "types" of covariance breakdowns. Hence this modeling structure distinguishes itself from a standard regime-switching approach which moves between 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 $s_t = 1$.

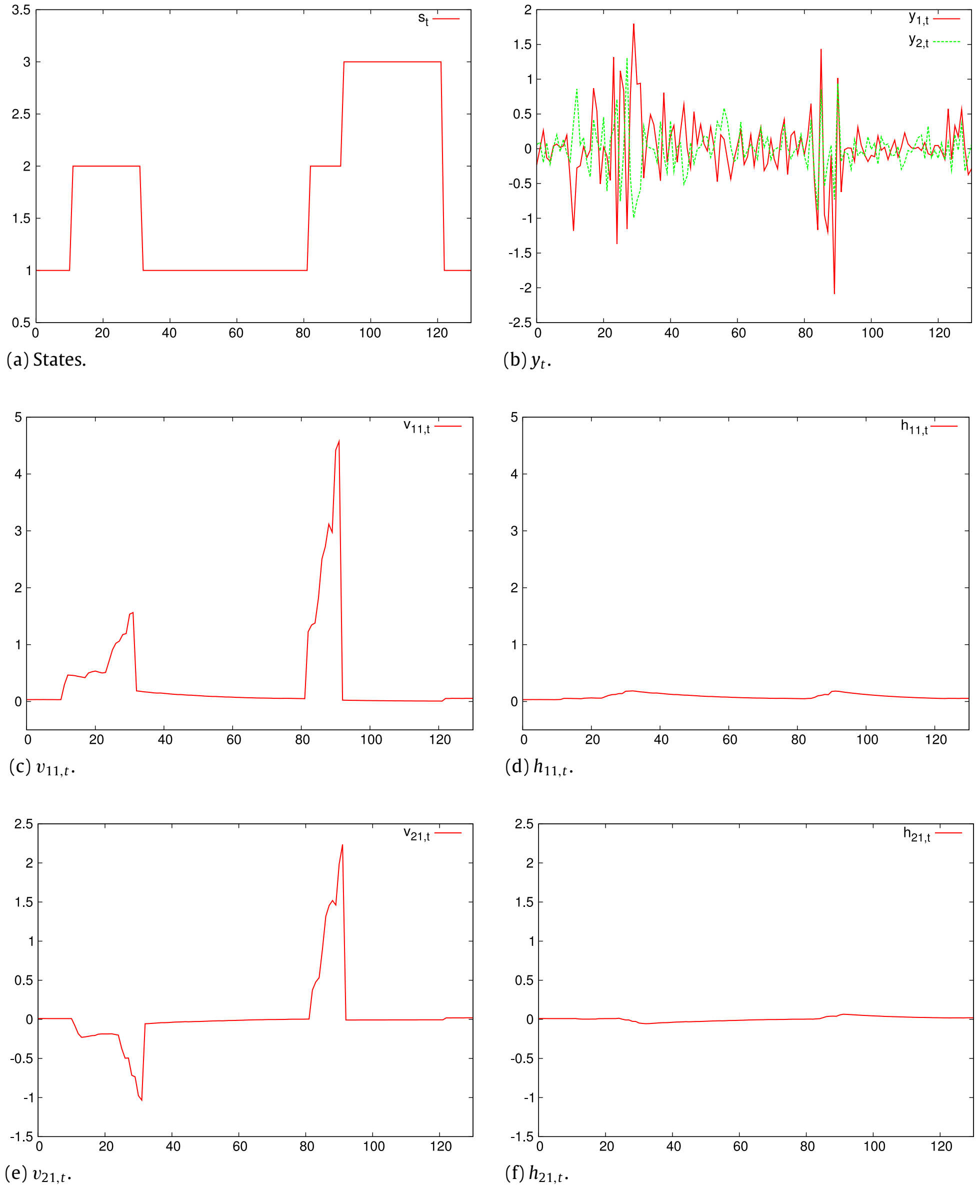
The functional form for covariance breakdowns $H_t^{1/2}\Lambda_t(H_t^{1/2})'$ is convenient in that it remains positive definite and a well defined conditional covariance matrix. It is also very flexible as the term can represent any positive definite matrix by choosing the right Λ_t.[7] As a result, various changes in the conditional covariance matrix can happen when a breakdown hits. Table 1 provides an example. We see cases where the variances, the covariance, and the correlation coefficient all increase (or decrease), or 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.

One way to have an idea of the general effect of Λ_t is to look at the overall variability reflected in the covariance matrix. One such measure 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

6 In the previous version of this paper the model had only one breakdown state, so different breakdown periods are separated by at least one normal period. We thank a referee for suggesting the idea of using a second breakdown state to fit consecutive breakdowns.

7 Given any symmetric positive definite matrices H and V of the same dimension, there always exists another symmetric positive-definite matrix Λ which satisfies $H^{1/2}\Lambda(H^{1/2})' = V$. Indeed, such Λ can be obtained by letting $\Lambda = (H_t^{1/2})^{-1}V((H^{1/2})')^{-1}$.



(a) States. (b) y_t. (c) $v_{11,t}$. (d) $h_{11,t}$. (e) $v_{21,t}$. (f) $h_{21,t}$.

Fig. 1. Simulation example.

$\Lambda_t \neq I$, 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.

Conditional on the previous state s_{t-1} the model implies a mixture of multivariate normal and Student-t distributions. For instance, if $s_{t-1} = 1$ and setting $Q_0 = I(\nu - k - 1)$ such that $E[\Lambda_t] = I$, then integrating out Λ_t in state 2 gives the following mixture,

$$y_t|s_{t-1} = 1 \sim \pi_1 N(\mu, H_t) + (1-\pi_1) \times t\left(\mu, \frac{\nu-k-1}{\nu-k+1}H_t, \nu-k+1\right), \quad (4)$$

where $t(\mu, \frac{\nu-k-1}{\nu-k+1}H_t, \nu-k+1)$ denotes a Student-t distribution with mean μ, scale matrix $\frac{\nu-k-1}{\nu-k+1}H_t$ and degree of freedom $\nu - k + 1$. Similar results hold conditional on being in state 2 or 3. The Appendix establishes that the last term is a Student-t distribution.

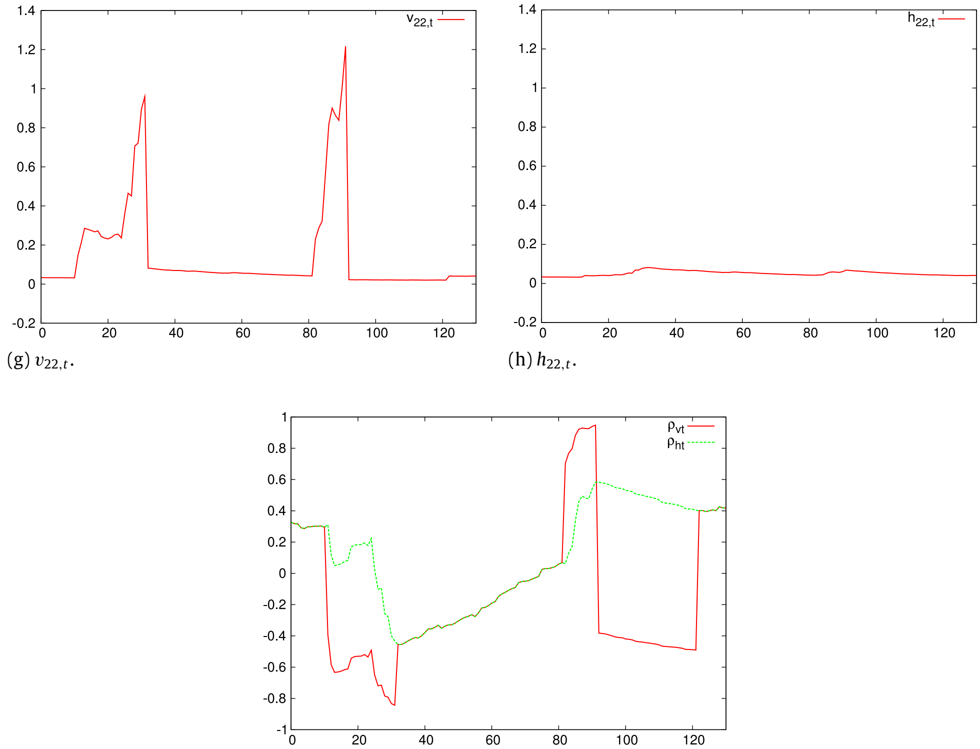
The possibility of covariance breakdowns leads to a fat-tailed predictive distribution. Later we extend the model to have Student-t innovations (z_t) which leads to a predictive distribution that is a more complex mixture.

Consider a 2×2 example. Assume the following VDGARCH specification[8]

$$H_t = CC' + aa' \odot (y_{t-1} - \mu)(y_{t-1} - \mu)' + bb' \odot H_{t-1} \quad (5)$$

where C is a 2×2 lower triangular matrix, a and b are $k \times 1$, $\odot$ denotes the element-by-element matrix product (Hadamard product). Let $G_0 = W^{-1}(7, 4I)$. Simulating from this model, Fig. 1 displays 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 $H_t = (h_{ij,t})$. The episode contains two breakdown periods with different break patterns. The first

[8] Covariance stationary conditions for this model in the absence of breakdowns are $a_i^2 + b_i^2 < 1, i = 1, \ldots, k$ (Ledoit et al., 2003). At present the stationary conditions for the full model with breakdowns are unknown.



(g) $v_{22,t}$.

(h) $h_{22,t}$.

(i) $\rho_{21,t}$.

Fig. 1. (*continued*)

breakdown is a move to state 2 and then back to state 1. The second breakdown is more complex with a move to state 2 followed by state 3 before returning to the normal state 1. This example shows that abrupt changes in the second moments can be accounted for by this model including jumps and decreases in conditional correlations.

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 π, Θ, μ, where $\pi = \{\pi_1, \ldots, \pi_5\}$, and Θ denotes the set of parameters in the GARCH specification that governs H_t. For example, if the VDGARCH specification in (5) 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 $\mathbf{S} = \{s_t\}_{t=1}^T$, $\mathbf{Y} = \{y_t\}_{t=1}^T$, $\boldsymbol{\Lambda} = \{\Lambda_t\}_{t=1}^T$. To sample from the joint posterior distribution $p(\pi, \Theta, \mu, \mathbf{S}, \boldsymbol{\Lambda}|\mathbf{Y})$, we propose an efficient MCMC algorithm that iteratively samples through the conditional posterior distribution of each block:

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

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 N draws $\{\pi^{(i)}, \Theta^{(i)}, \mu^{(i)}, \mathbf{S}^{(i)}, \boldsymbol{\Lambda}^{(i)}\}_{i=1}^N$ 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 $N^{-1}\sum_{i=1}^N \pi_1^{(i)}$. Next we discuss each block in more detail.

3.1. Sampling S

Conditional on H_t and μ, apply the transformation

$$\tilde{y}_t = H_t^{-1/2}(y_t - \mu), \tag{6}$$

so that

$$\tilde{y}_t|\Lambda_t \sim N(0, \Lambda_t). \tag{7}$$

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

We sequentially sample from the three-point discrete distributions $p(s_t|\mathbf{S}_{-t}, \pi, \tilde{\mathbf{Y}})$ for $t = 1, \ldots, T$, where $\mathbf{S}_{-t} = \mathbf{S} \setminus s_t$. It requires calculating $Pr(s_t = i|\mathbf{S}_{-t}, \pi, \tilde{\mathbf{Y}})$, $i = 1, 2, 3$ for each t. Since Λ_t randomly changes with states it must be integrated out. This leads to several cases depending on the values of $\mathbf{S}_{-t}$. Let t_1, t_2 be integers such that $T \geq t_1 \geq 1$, $T \geq t_2 \geq 1$. Suppress π 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:

- $s_{t-1} = 1, s_{t+1} = 1$
- $s_{t-1} = 1, s_{t+1} = \cdots = s_{t+t_2} = 2, s_{t+t_2+1} \neq 2$
- $s_{t-1} = 1, s_{t+1} = \cdots = s_{t+t_2} = 3, s_{t+t_2+1} \neq 3$
- $s_{t-t_1-1} \neq 2, s_{t-t_1} = \cdots = s_{t-1} = 2, s_{t+1} = 1$
- $s_{t-t_1-1} \neq 2, s_{t-t_1} = \cdots = s_{t-1} = 2, s_{t+1} = \cdots = s_{t+t_2} = 2, s_{t+t_2+1} \neq 2$
- $s_{t-t_1-1} \neq 2, s_{t-t_1} = \cdots = s_{t-1} = 2, s_{t+1} = \cdots = s_{t+t_2} = 3, s_{t+t_2+1} \neq 3$
- $s_{t-t_1-1} \neq 3, s_{t-t_1} = \cdots = s_{t-1} = 3, s_{t+1} = 1$

- $s_{t-t_1-1} \neq 3, s_{t-t_1} = \cdots = s_{t-1} = 3, s_{t+1} = \cdots = s_{t+t_2} = 2, s_{t+t_2+1} \neq 2$
- $s_{t-t_1-1} \neq 3, s_{t-t_1} = \cdots = s_{t-1} = 3, s_{t+1} = \cdots = s_{t+t_2} = 3, s_{t+t_2+1} \neq 3.$

To illustrate we consider the sixth case. According to the transition matrix it is impossible to move from state 2 to 1 to 3. The possible moves are to state 2 or 3 before a switch to state 3. The probabilities are

$$
\begin{aligned}
&Pr(s_t = 1|\mathbf{S}_{-t}, \tilde{\mathbf{Y}}) = 0\\
&Pr(s_t = 2|\mathbf{S}_{-t}, \tilde{\mathbf{Y}}) \propto \left[\int \left(\prod_{j=-t_1}^{0} p(\tilde{y}_{t+j}|\Lambda)\right) p(\Lambda)d\Lambda\right]\\
&\qquad \times \left[\int \left(\prod_{j=1}^{t_2} p(\tilde{y}_{t+j}|\Lambda)\right) p(\Lambda)d\Lambda\right]\\
&\qquad \times Pr(s_t = 2|s_{t-1} = 2)Pr(s_{t+1} = 3|s_t = 2)\\
&Pr(s_t = 3|\mathbf{S}_{-t}, \tilde{\mathbf{Y}}) \propto \left[\int \left(\prod_{j=-t_1}^{-1} p(\tilde{y}_{t+j}|\Lambda)\right) p(\Lambda)d\Lambda\right]\\
&\qquad \times \left[\int \left(\prod_{j=0}^{t_2} p(\tilde{y}_{t+j}|\Lambda)\right) p(\Lambda)d\Lambda\right]\\
&\qquad \times Pr(s_t = 3|s_{t-1} = 2)Pr(s_{t+1} = 3|s_t = 3).
\end{aligned}
$$

From these s_t can be sampled. Each of the other cases are fully detailed in the Appendix.

In all cases,[9] 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

$$
\int \left(\prod_{t=t_3}^{t_4} p(\tilde{y}_t|\Lambda)\right) p(\Lambda)d\Lambda = \frac{2^{\frac{nk}{2}} (2\pi)^{-\frac{nk}{2}} |Q_0|^{\frac{\nu}{2}} \prod_{j=1}^{k} \Gamma\left(\frac{n+\nu+1-j}{2}\right)}{|Q|^{\frac{n+\nu}{2}} \prod_{j=1}^{k} \Gamma\left(\frac{\nu+1-j}{2}\right)}, \tag{8}
$$

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

The conditional posterior distributions are:

$$p(\pi_1|\mathbf{S}) \propto p(\mathbf{S}|\pi_1)p(\pi_1) \propto p(\pi_1)\pi_1^{n_{11}}(1-\pi_1)^{n_{12}} \tag{9}$$

$$p(\pi_2|\mathbf{S}) \propto p(\mathbf{S}|\pi_2)p(\pi_2) \propto p(\pi_2)\pi_2^{n_{22}}(1-\pi_2)^{n_{21}+n_{23}} \tag{10}$$

$$p(\pi_3|\mathbf{S}) \propto p(\mathbf{S}|\pi_3)p(\pi_3) \propto p(\pi_3)\pi_3^{n_{33}}(1-\pi_3)^{n_{31}+n_{32}} \tag{11}$$

$$p(\pi_4|\mathbf{S}) \propto p(\mathbf{S}|\pi_4)p(\pi_4) \propto p(\pi_4)\pi_4^{n_{21}}(1-\pi_4)^{n_{23}} \tag{12}$$

$$p(\pi_5|\mathbf{S}) \propto p(\mathbf{S}|\pi_5)p(\pi_5) \propto p(\pi_5)\pi_5^{n_{31}}(1-\pi_5)^{n_{32}} \tag{13}$$

where

$$n_{ij} = \#\{t \in \{1, \ldots, T-1\}|s_t = i, s_{t+1} = j\}$$

and # denotes the number of elements in a set. $p(\pi_i)$ are prior distributions. In practice, we follow, among others, He and Maheu (2010) to reparameterize π_i as $\log(\frac{\pi_i}{1-\pi_i})$ and impose priors on $\log(\frac{\pi_i}{1-\pi_i})$. This makes it convenient to impose informative priors that favor infrequent covariance breakdowns and state persistence. We jointly sample π using a Metropolis–Hastings (M–H) step with a multivariate-normal random walk proposal.

3.3. Sampling Θ

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

$$
\begin{aligned}
p(\mathbf{Y}|\Theta, \mu, \mathbf{S}) &= \int \left(\prod_{t=1}^{T} p(y_t|\mu, H_t, \Lambda_t)\right) p(\boldsymbol{\Lambda}|\mathbf{S})d\boldsymbol{\Lambda}\\
&= \int \left(\prod_{t=1}^{T} (2\pi)^{-\frac{k}{2}} |H_t^{\frac{1}{2}} \Lambda_t (H_t^{\frac{1}{2}})'|^{-\frac{1}{2}}\right.\\
&\quad \times \exp\left(-\frac{1}{2}(y_t-\mu)'(H_t^{\frac{1}{2}} \Lambda_t (H_t^{\frac{1}{2}})')^{-1}\right.\\
&\quad \left.\left.\times (y_t-\mu)\right)\right) p(\boldsymbol{\Lambda}|\mathbf{S})d\boldsymbol{\Lambda}\\
&= \left[\prod_{t=1}^{T} |H_t|^{-\frac{1}{2}}\right] \int \left(\prod_{t=1}^{T} (2\pi)^{-\frac{k}{2}} |\Lambda_t|^{-\frac{1}{2}}\right.\\
&\quad \left.\times \exp\left(-\frac{1}{2}\tilde{y}_t'\Lambda_t^{-1}\tilde{y}_t\right)\right) p(\boldsymbol{\Lambda}|\mathbf{S})d\boldsymbol{\Lambda}\\
&= \left[\prod_{t=1}^{T} |H_t|^{-\frac{1}{2}}\right] p(\tilde{\mathbf{Y}}|\mathbf{S}).
\end{aligned} \tag{14}
$$

The last line in (14) requires the computation of $p(\tilde{\mathbf{Y}}|\mathbf{S})$. Let $\mathcal{A}_{t_1,t_2}$ denote a stretch of state 2 or a stretch of state 3 with endpoints t_1 and t_2. That is, $\mathcal{A}_{t_1,t_2} = \{t\}_{t=t_1}^{t_2}$ such that $s_{t_1} = s_{t_1+1} = \cdots = s_{t_2} = 2$ or 3, and $s_{t_1-1} \neq s_{t_1}, s_{t_2} \neq s_{t_2+1}$. Suppose, given $\mathbf{S}$, there are $B \geq 0$ such stretch(es) over the whole sample simply denoted as $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_B$. Also let $\mathcal{A}^c$ denote the union of all the normal periods, then $\mathcal{A}^c, \mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_B$ forms a partition of $\{1, 2, \ldots, T\}$.

Now

$$
\begin{aligned}
p(\tilde{\mathbf{Y}}|\mathbf{S}) &= \prod_{t\in\mathcal{A}^c} N(\tilde{y}_t|0, I) \prod_{q=1}^{B} \int \left(\prod_{t\in\mathcal{A}_q} p(\tilde{y}_t|\Lambda)\right) p(\Lambda)d\Lambda\\
&= \prod_{t\in\mathcal{A}^c} \left((2\pi)^{-\frac{k}{2}} \exp\left(-\frac{1}{2}\tilde{y}_t'\tilde{y}_t\right)\right)\\
&\quad \times \prod_{q=1}^{B} \frac{2^{\frac{N_qk}{2}} (2\pi)^{-\frac{N_qk}{2}} |Q_0|^{\frac{\nu}{2}} \prod_{j=1}^{k} \Gamma\left(\frac{N_q+\nu+1-j}{2}\right)}{|Q_q|^{\frac{N_b+\nu}{2}} \prod_{j=1}^{k} \Gamma\left(\frac{\nu+1-j}{2}\right)}
\end{aligned} \tag{15}
$$

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

$$
\begin{aligned}
p(\Theta|\mathbf{S}, \mu, \mathbf{Y}) &\propto p(\mathbf{Y}|\Theta, \mu, \mathbf{S})p(\Theta)\\
&\propto \left[\prod_{t=1}^{T} |H_t|^{-\frac{1}{2}}\right] p(\tilde{\mathbf{Y}}|\mathbf{S})p(\Theta).
\end{aligned} \tag{16}
$$

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.

[9] 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.

3.4. Sampling $\boldsymbol{\Lambda}$

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

$$\tilde{\Lambda}_q \sim W^{-1}(\nu_q, Q_q), \tag{17}$$

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

3.5. Sampling μ

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

$$p(\mathbf{Y}|\boldsymbol{\Lambda}, \Theta, \mu) = \prod_{t=1}^{T} N(y_t|\mu, H_t^{\frac{1}{2}}\Lambda_t(H_t^{\frac{1}{2}})'), \tag{18}$$

and the conditional posterior density is

$$p(\mu|\boldsymbol{\Lambda}, \Theta, \mathbf{Y}) \propto p(\mu)p(\mathbf{Y}|\boldsymbol{\Lambda}, \Theta, \mu). \tag{19}$$

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. Following Geweke (1993) 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.[10] More specifically, if y_t follows a Student-t distribution, $y_t \sim t(\mu, H_t^{1/2}\Lambda_t(H_t^{1/2})', d)$, then it can be written as

$$y_t = \mu + u_t^{-1/2}(y_t^* - \mu), \tag{20}$$

where

$$y_t^* \sim N(\mu, H_t^{1/2}\Lambda_t(H_t^{1/2})') \tag{21}$$

$$u_t \sim G(d/2, d/2). \tag{22}$$

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

$$\begin{aligned} p(u_t|y_t, \Theta, \boldsymbol{\Lambda}, \mu, d) &\propto p(u_t)p(y_t|u_t, \Theta, \boldsymbol{\Lambda}, \mu) \\ &\propto G(u_t|d/2, d/2)N(y_t|\mu, u_t^{-1}V_t) \\ &\propto u_t^{\frac{k+d}{2}-1} e^{-\frac{1}{2}u_t(d+(y_t-\mu)'V_t^{-1}(y_t-\mu))} \\ &\propto G\left(u_t \middle| \frac{k+d}{2}, \frac{1}{2}(d + (y_t-\mu)'V_t^{-1}(y_t-\mu))\right). \end{aligned} \tag{23}$$

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

$$\begin{aligned} p(d|\mathbf{U}) &\propto p(d)p(\mathbf{U}|d) \\ &\propto \text{Exp}(d|\lambda_0)\mathbb{I}_{d>2} \prod_{t=1}^{T} \frac{(\frac{d}{2})^{d/2}}{\Gamma(\frac{d}{2})} u_t^{d/2-1} e^{-du_t/2} \\ &\propto \lambda_0 e^{-\lambda_0 d}\mathbb{I}_{d>2} \left(\frac{(\frac{d}{2})^{d/2}}{\Gamma(\frac{d}{2})}\right)^T \exp\left(-\frac{d}{2}\sum_{t=1}^{T}(u_t - \log(u_t))\right). \end{aligned} \tag{24}$$

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[11]:

1. $p(\mathbf{S}|\Theta, \pi, \mu, \mathbf{Y}^*)$,
2. $p(\pi|\mathbf{S})$,
3. $p(\Theta|\mathbf{S}, \mu, \mathbf{Y}^*)$,
4. $p(\boldsymbol{\Lambda}|\mathbf{S}, \Theta, \mu, \mathbf{Y}^*)$,
5. $p(\mu|\boldsymbol{\Lambda}, \Theta, \mathbf{Y}^*)$,
6. $p(\mathbf{U}|\Theta, \boldsymbol{\Lambda}, \mu, d, \mathbf{Y})$,
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 parameters into the posterior sampling scheme, we can learn about the typical effect of Λ_t. As shown in Appendix A.4, sampling of ν and Q_0 can be included in the posterior sampling algorithm as an M–H step and a Gibbs step, respectively.

3.8. Predictive density and moments

A summary of the main specification of the model (VDGARCH-t-B) we focus on is the following.

$$y_t = \mu + H_t^{1/2}\Lambda_t^{1/2}v_t, \quad v_t \sim t(0, I, d) \tag{25}$$

$$s_t|s_{t-1} \sim P_{s_{t-1}}, \quad s_t \in \{1, 2, 3\} \tag{26}$$

$$\Lambda_t|s_{1:t} = \begin{cases} I & \text{if } s_t = 1 \\ \Lambda_{t-1} & \text{if } s_t = s_{t-1} \\ \sim W^{-1}(\nu, Q_0) & \text{if } s_t \neq s_{t-1} \text{ and } s_t \neq 1 \end{cases} \tag{27}$$

$$H_t = CC' + aa' \odot (y_{t-1} - \mu)(y_{t-1} - \mu)' + bb' \odot H_{t-1}, \tag{28}$$

where $P_{s_{t-1}}$ denotes row s_{t-1} of the transition matrix (2). The key features of the model have been discussed above as have the posterior simulation methods. After dropping a suitable number of burn-in iterations the next N MCMC iterations provide,

$$\{\mathbf{S}^{(i)}, \boldsymbol{\Lambda}^{(i)}, \pi^{(i)}, \Theta^{(i)}, \mu^{(i)}, \mathbf{U}^{(i)}, d^{(i)}, \nu^{(i)}, Q_0^{(i)}\}_{i=1}^N. \tag{29}$$

From this output the predictive density for y_t, conditional on $\mathcal{F}_{t-1}$ can be estimated as

$$p(y_t|\mathcal{F}_{t-1}) \approx \frac{1}{N}\sum_{i=1}^{N} t(y_t|\mu^{(i)}, H_t^{(i)1/2}\Lambda_t^{(i)}(H_t^{(i)1/2})', d^{(i)}), \tag{30}$$

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

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

Table 2
Estimates for simulated data I.

Parameter	True value	VDGARCH-B			VDGARCH		
		Mean	NSE	0.95 DI	Mean	NSE	0.95 DI
C_{11}	0.04	0.0391	0.0014	(0.0364, 0.0419)	0.0427	0.0019	(0.0391, 0.0466)
C_{21}	0.01	0.0112	0.0009	(0.0095, 0.0130)	0.0016	0.0021	(−0.0026, 0.0058)
C_{22}	0.03	0.0296	0.0011	(0.0276, 0.0319)	0.0437	0.0021	(0.0397, 0.0480)
a_1	0.12	0.1132	0.0057	(0.1039, 0.1237)	0.3627	0.0081	(0.3472, 0.3808)
a_2	0.10	0.0975	0.0051	(0.0892, 0.1070)	0.3886	0.0082	(0.3730, 0.4060)
b_1	0.97	0.9717	0.0017	(0.9680, 0.9748)	0.9194	0.0032	(0.9119, 0.9251)
b_2	0.98	0.9805	0.0012	(0.9781, 0.9825)	0.9186	0.0031	(0.9116, 0.9243)
π_1	0.99	0.9858	0.0027	(0.9798, 0.9906)			
π_2	0.97	0.9664	0.0064	(0.9531, 0.9783)			
π_3	0.98	0.9773	0.0053	(0.9654, 0.9857)			
π_4	0.70	0.7306	0.0820	(0.5581, 0.8697)			
π_5	0.90	0.9357	0.0519	(0.8001, 0.9904)			

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.

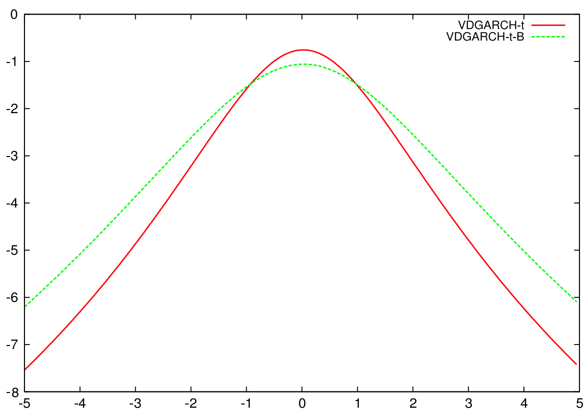


Fig. 2. Log-density functions for 2008/10/14.

where $H_t^{(i)}$ is constructed from the GARCH recursion (28) given parameter $\Theta^{(i)}$, $\Lambda_t^{(i)}$ is determined from (27) given s_t which is drawn as $s_t \sim P^{(i)}_{s^{(i)}_{t-1}}$. $t(y_t|\mu^{(i)}, H_t^{(i)1/2}\Lambda_t^{(i)}(H_t^{(i)1/2})', d^{(i)})$ denotes the Student-t probability density function with mean $\mu^{(i)}$, scale matrix $H_t^{(i)1/2}\Lambda_t^{(i)}(H_t^{(i)1/2})'$ and degree of freedom parameter $d^{(i)}$, evaluated at y_t.

In a similar way the predictive mean can be estimated as

$$E[y_t|\mathcal{F}_{t-1}] \approx \frac{1}{N}\sum_{i=1}^{N}\mu^{(i)} \tag{31}$$

and the predictive variance is

$$\mathrm{Var}(y_t|\mathcal{F}_{t-1}) = E[y_t y_t'|\mathcal{F}_{t-1}] - E[y_t|\mathcal{F}_{t-1}]E[y_t|\mathcal{F}_{t-1}]' \tag{32}$$

with

$$E[y_t y_t'|\mathcal{F}_{t-1}] \approx \frac{1}{N}\sum_{i=1}^{N}\left(\frac{d^{(i)}}{d^{(i)}-2}H_t^{(i)1/2}\Lambda_t^{(i)}(H_t^{(i)1/2})' + \mu^{(i)}\mu^{(i)'}\right). \tag{33}$$

As an example of the tail structure of this model Fig. 2 plots the log-density of an equally-weighted portfolio for the covariance breakdown model (VDGARCH-B-t) along with a model without breakdowns but Student-t innovations (VDGARCH-t). It is clear that the breakdown model has significantly thicker tails.

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

$$C = \begin{pmatrix} 0.04 & 0 \\ 0.01 & 0.03 \end{pmatrix}, \quad a = \begin{pmatrix} 0.12 \\ 0.1 \end{pmatrix}, \quad b = \begin{pmatrix} 0.97 \\ 0.98 \end{pmatrix},$$
$$\pi_1 = 0.99, \quad \pi_2 = 0.97, \quad \pi_3 = 0.97, \quad \pi_4 = 0.7,$$
$$\pi_5 = 0.9, \quad \mu = 0.$$

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: $\log(\frac{\pi_1}{1-\pi_1}) \sim N(4, 0.1)$, $\log(\frac{\pi_i}{1-\pi_i}) \sim N(3, 0.1)$, $i = 2, \ldots, 5$; $C_{ii} \sim TN_+(0, 100)$ 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. Overall the priors are uninformative except for the priors on the transition probabilities which favor infrequent covariance breakdowns, state persistence and more time spent in state 1. In the posterior sampling, the first 10 000 MCMC draws are discarded as burn-in and the next 10 000 draws are used for inference.

Parameter estimates are reported in Table 2. All parameter values are accurately recovered. 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.[12] 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.

Fig. 3 plots the smoothed states and smoothed variances from the models during several covariance breakdown episodes. 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.

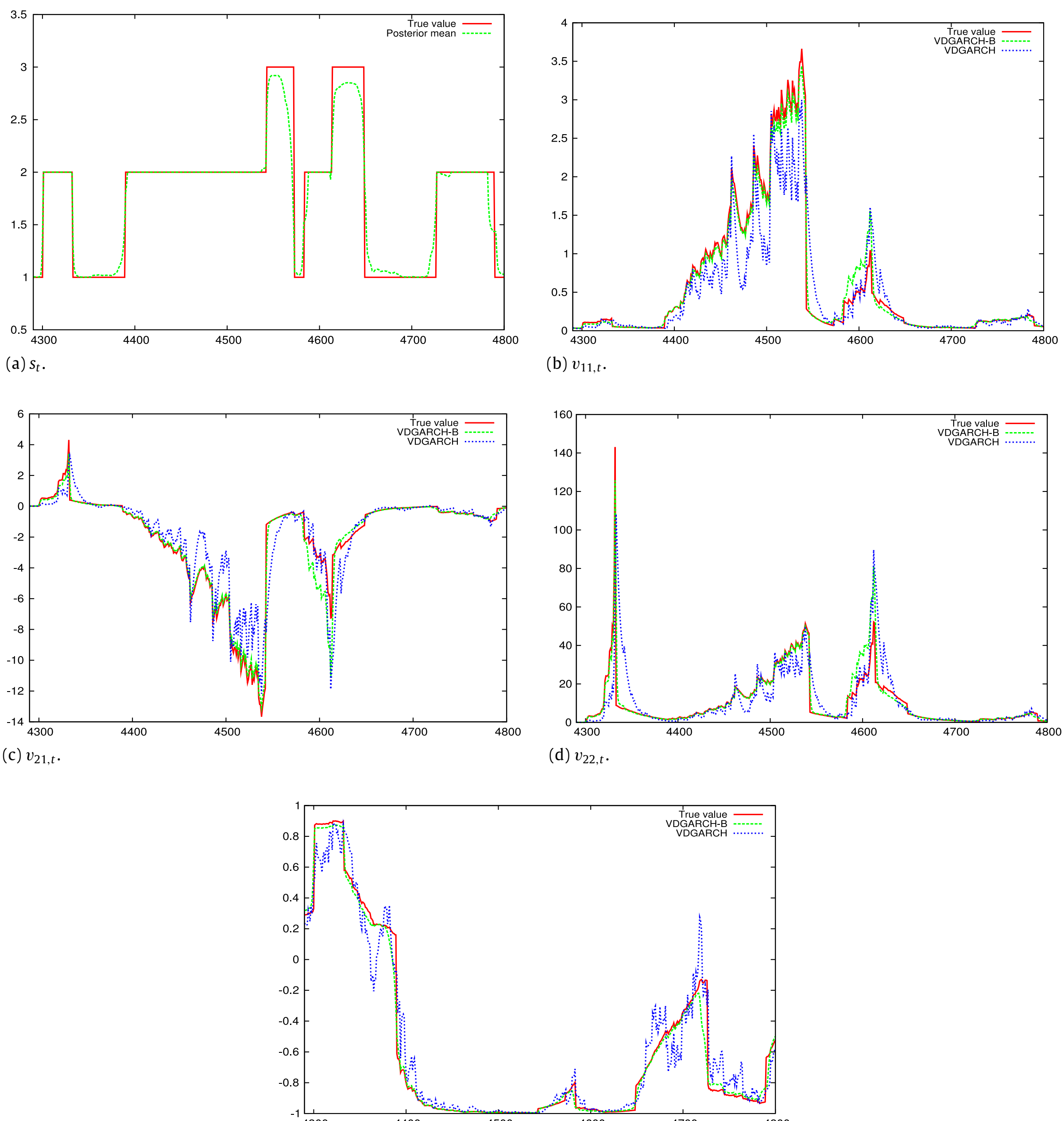
Table 3 repeats the exercise by estimating the models on simulated data with no covariance breakdowns. Here data are simulated from a plain VDGARCH model with parameters specified earlier and both specifications are estimated. The VDGARCH-B model does a good job in identifying no breakdowns and recovering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

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

Table 3
Estimates for simulated data II.

Parameter	True value	VDGARCH-B			VDGARCH		
		Mean	NSE	0.95 DI	Mean	NSE	0.95 DI
C_{11}	0.04	0.0382	0.0039	(0.0307, 0.0454)	0.0345	0.0042	(0.0275, 0.0451)
C_{21}	0.01	0.0116	0.0009	(0.0099, 0.0135)	0.0114	0.0019	(0.0080, 0.0151)
C_{22}	0.03	0.0323	0.0027	(0.0277, 0.0378)	0.0314	0.0054	(0.0206, 0.0405)
a_1	0.12	0.1061	0.0112	(0.0827, 0.1277)	0.1007	0.0138	(0.0752, 0.1292)
a_2	0.10	0.1016	0.0105	(0.0803, 0.1221)	0.0944	0.0121	(0.0709, 0.1175)
b_1	0.97	0.9732	0.0047	(0.9631, 0.9816)	0.9776	0.0053	(0.9641, 0.9860)
b_2	0.98	0.9768	0.0034	(0.9696, 0.9819)	0.9784	0.0061	(0.9669, 0.9894)
π_1	1	0.9983	0.0010	(0.9955, 0.9997)			

Data are simulated from the VDGARCH model for $T = 5000$ observations. The data exclude covariance breakdowns. This table reports the posterior mean, its numerical standard error (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.



(a) s_t.

(b) $v_{11,t}$.

(c) $v_{21,t}$.

(d) $v_{22,t}$.

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

Fig. 3. VDGARCH-B vs VDGARCH model using simulated data with breakdowns: smoothed volatility compared to the true values.

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	0.152	0.562	0.020	0.021
$RMSE_b$	0.254	0.658	–	–

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 $RMSE_b = \frac{1}{T_b}\sum_{s_t=2,3} \|V_t - \bar{V}_t\|/\|V_t\|$ for both model, where T_b is the number of data points in state 2 and state 3.

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	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.

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 0.152 and the VDGARCH model has a RMSE of 0.562. Modeling the covariance breakdowns is important for accurate volatility estimation. Focusing only on the RMSE from covariance breakdown periods ($\{t|s_t = 2, 3\}$) 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:

$$f(\mathbf{Y}) = \frac{f(\mathbf{Y}|\Psi)f(\Psi)}{f(\Psi|\mathbf{Y})}, \tag{34}$$

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. Eq. (34) 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

$$\log f(\mathbf{Y}) = \log f(\mathbf{Y}|\Psi^*) + \log f(\Psi^*) - \log f(\Psi^*|\mathbf{Y}). \tag{35}$$

To estimate $\log f(\mathbf{Y}|\Psi^*)$ the latent state variables $\mathbf{S}$ and $\boldsymbol{\Lambda}$ are integrated out of the likelihood function. We design a particle filter based on the auxiliary particle filter of Pitt and 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. The final term, $\log f(\Psi^*|\mathbf{Y})$, is the log-posterior ordinate. We follow Chib and 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 Appendix A.5.

6. Empirical application

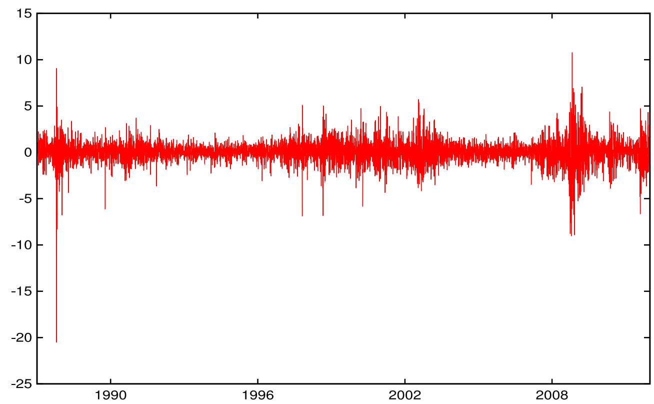
In this section we apply the model to study the volatility dynamics among the stock market and the bond market.[13] We use daily excess returns on the S&P 500 index ($y_{1,t}$), a ten-year Treasury bond ($y_{2,t}$), and a one-year Treasury bond ($y_{3,t}$). The return data are obtained from the Center 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. Fig. 4 plots the three excess return series and Table 5 provides summary statistics. Returns are in percentage.

We estimate the VDGARCH-t model and the VDGARCH-t-B model using the return data. We will focus on estimation of the preferred VDGARCH-t-B along with a specification with no covariance breakdowns. For π and Θ, the priors are identical to those listed in Section 4. 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 first 10 000 draws of the MCMC chain are discarded as burn-in and the next 10 000 draws are used for inference.

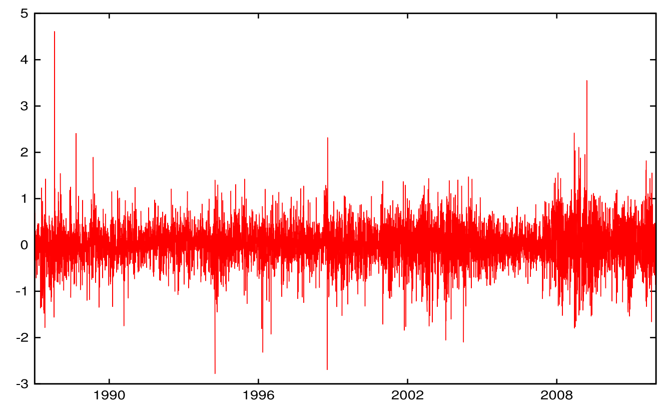
Parameter estimates are reported in Table 6. State persistence is high and equal to 0.9773, 0.9563 and 0.9665 for states 1, 2 and 3 respectively. On average, movement out of state 2 is more likely to go to state 1 as to state 3 (π_4 vs $(1-\pi_4)$). A similar result holds for moves out of state 3. Based on the posterior means, the unconditional state probabilities are $Pr(s_t = 1) = 0.63$, $Pr(s_t = 2) = 0.33$, $Pr(s_t = 3) = 0.04$ so that most time is spent in the normal state. Similar to the simulation exercise, the $a_i i = 1, 2, 3$ parameters in the VDGARCH-t-B model are smaller than the VDGARCH-t values, while the b_i parameters are larger. In the breakdown model $a_i^2 + b_i^2$, a measure of volatility persistence in conventional GARCH models, is about 0.99 for $i = 1, 2, 3$. The degree of freedom parameter in the Student-t distribution of the VDGARCH-t is 5.6 while it is 7.5 in the covariance breakdown model.

The empirical posterior distribution of the duration of the breakdown periods is shown in Table 7. The majority of breakdown (states 2 and 3) durations are in (2,20] but there is a significant proportion (46%) that have a duration greater than a month. On average, a breakdown episode lasts for 1.5 months and 42% of the sample is identified as a covariance breakdown. Note, in contrast to this, the breakdown model with Gaussian innovations (not reported) identifies many more short-term breakdowns as they are used to capture outliers in the data.

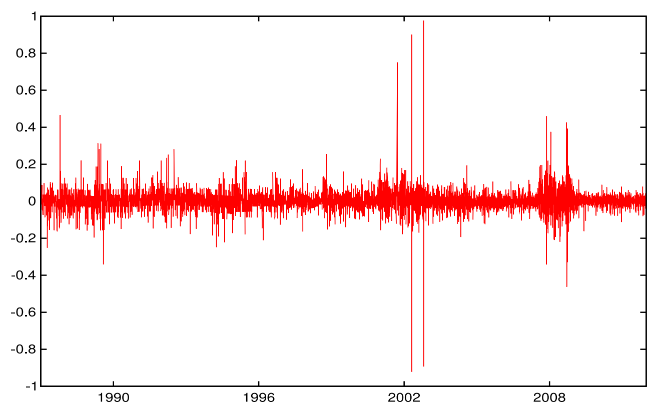
[13] For applications of MGARCH models to stock and bond markets without covariance breakdowns see Cappiello et al. (2006) and De Goeij and Marquering (2004).



(a) S&P 500.



(b) Ten-year bond.



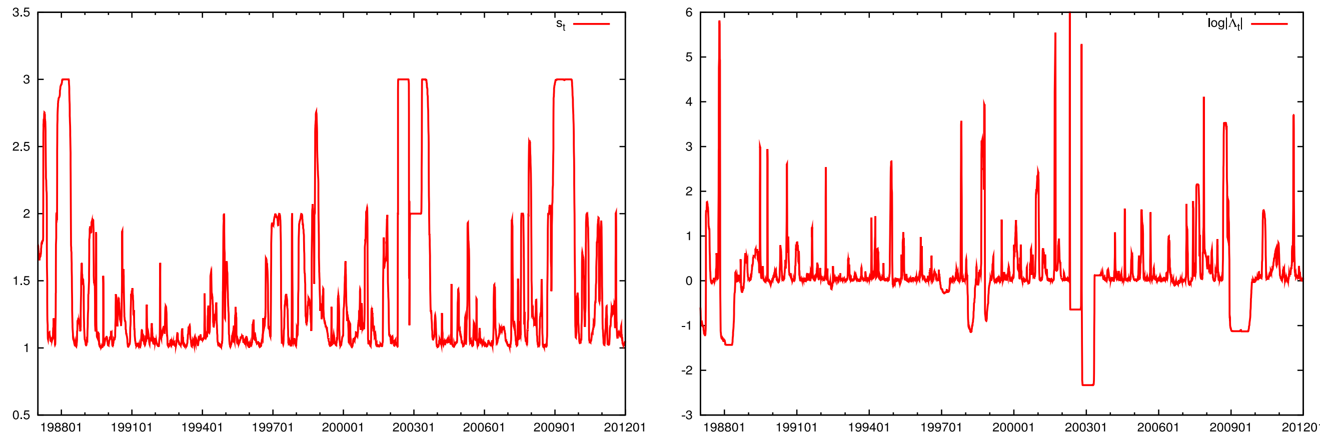
(c) One-year bond.

Fig. 4. Daily excess returns in percentage.

Table 6
Parameter estimates with student-t innovations.

Parameter	VDGARCH-t-B			VDGARCH-t		
	Mean	NSE	0.95 DI	Mean	NSE	0.95 DI
π_1	0.9773	0.0047	(0.9675, 0.9857)			
π_2	0.9563	0.0061	(0.9433, 0.9679)			
π_3	0.9665	0.0053	(0.9548, 0.9755)			
π_4	0.9296	0.0217	(0.8781, 0.9616)			
π_5	0.9474	0.0165	(0.9081, 0.9721)			
C_{11}	0.0287	0.0069	(0.0146, 0.0423)	0.0555	0.0062	(0.0428, 0.0674)
C_{21}	0.0065	0.0032	(0.0002, 0.0130)	0.0076	0.0024	(0.0031, 0.0123)
C_{31}	0.0004	0.0002	(−0.0001, 0.0009)	0.0005	0.0003	(−0.0001, 0.0011)
C_{22}	0.0226	0.0026	(0.0171, 0.0273)	0.0332	0.0026	(0.0280, 0.0384)
C_{32}	0.0017	0.0002	(0.0013, 0.0020)	0.0034	0.0003	(0.0027, 0.0035)
C_{33}	0.0014	0.0002	(0.0010, 0.0018)	0.0030	0.0002	(0.0025, 0.0035)
a_1	0.1082	0.0067	(0.0959, 0.1222)	0.1701	0.0070	(0.1577, 0.1843)
a_2	0.1031	0.0068	(0.0897, 0.1156)	0.1773	0.0065	(0.1643, 0.1901)
a_3	0.0860	0.0069	(0.0731, 0.0994)	0.1982	0.0090	(0.1823, 0.2178)
b_1	0.9903	0.0011	(0.9876, 0.9922)	0.9768	0.0018	(0.9732, 0.9800)
b_2	0.9890	0.0011	(0.9865, 0.9912)	0.9723	0.0019	(0.9684, 0.9761)
b_3	0.9913	0.0009	(0.9890, 0.9929)	0.9616	0.0033	(0.9519, 0.9668)
μ_1	0.0613	0.0088	(0.0433, 0.0786)	0.0607	0.0087	(0.0440, 0.0778)
μ_2	0.0154	0.0040	(0.0076, 0.0234)	0.0144	0.0042	(0.0060, 0.0224)
μ_3	0.0019	0.0004	(0.0010, 0.0026)	0.0018	0.0004	(0.0008, 0.0026)
d	7.5088	0.5125	(6.5698, 8.5628)	5.6075	0.2211	(5.1774, 6.0500)
$Q_{0,11}$	5.2289	0.9043	(3.6225, 7.0689)			
$Q_{0,21}$	−0.1397	0.4587	(−1.0782, 0.7177)			
$Q_{0,31}$	0.1630	0.3173	(−0.4486, 0.8012)			
$Q_{0,22}$	5.4544	0.8430	(3.9566, 7.2648)			
$Q_{0,32}$	0.1290	0.4368	(−0.6891, 1.0266)			
$Q_{0,33}$	5.4364	1.1650	(3.3528, 7.7767)			
ν	6.2434	0.6099	(5.1382, 7.4998)			

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.



(a) s_t.

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

Fig. 5. Posterior mean of s_t and $\log(|\Lambda_t|)$ with t-innovations (VDGARCH-t-B).

Table 7
Empirical posterior distribution of breakdown duration.

VDGARCH-t-B								
Duration	=1	(2, 20]	(20, 40]	(40, 60]	(60, 80]	(80, 100]	(100, 120]	>120
Frequency	3	42	18	8	4	2	3	4

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 frequency of covariance breakdowns for each interval is rounded to the nearest integer.

Table 8
Posterior distribution of $\log(|E[\Lambda_t]|) = \log|Q_0/(\nu - k - 1)|$, VDGARCH-t-B model.

Mean	Median	Stdev	0.95 DI
2.7084	2.6463	0.5826	(1.7674, 3.9926)

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

Fig. 5(a) displays the posterior estimates for the breakdown periods. Some feature moves between state 2 and 3 while others only go to state 2 and back to state 1. Some of these episodes are discussed later. Fig. 5(b) plots the posterior mean of $\log(|\Lambda_t|)$. When there is likely a normal state ($s_t = 1$) indicated by the posterior mean of s_t, $\log(|\Lambda_t|)$ is close to zero 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 8 reports posterior summary statistics for the expected impact of covariance breakdowns. The posterior mean of 2.7084 implies an average increase in variability when a breakdown occurs. The 0.95 density interval indicates that we can expect breakdowns to increase the generalized variance on average but this does not rule out cases with a decrease as seen in Fig. 5(b).

6.1. Model comparison

To formally assess whether covariance breakdowns admitted in our model are supported by the data, we compute the marginal and predictive 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 and 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.

For a given model $\mathcal{M}$, the predictive likelihood over the data $y_{t_s}, \ldots, y_{t_e}$, $t_s \leq t_e$ is defined as

$$p(y_{t_s}, \ldots, y_{t_e}|\mathcal{F}_{t_s-1}, \mathcal{M}) = \prod_{j=t_s}^{t_e} p(y_j|\mathcal{F}_{j-1}, \mathcal{M}) \tag{36}$$

where all parameter uncertainty has been integrated out of the one-period ahead predictive likelihood terms $p(y_j|\mathcal{F}_{j-1}, \mathcal{M})$. This makes it clear that the predictive likelihood is the out-of-sample forecast record of a model (Geweke, 2005). Models with larger log-predictive likelihoods are models that are better able to account for the data. In a similar way if $t_s = 1$ and $t_e = T$ we have the marginal likelihood. For the breakdown model the individual predictive likelihoods can be estimated by (30).

We also include in the comparison a Markov-switching VDGARCH model introduced by Haas and Mittnik (2008), again with Student-t innovations (MS-VDGARCH-t). This specification avoids the path dependence in GARCH models by making the GARCH recursion a function only of the current state. The model is,

$$y_t|s_t \sim t(\mu, H_t(s_t), d) \tag{37}$$

$$\begin{aligned} H_t(s_t) \;=\; & C_{s_t}C_{s_t}' + a_{s_t}a_{s_t}' \odot (y_{t-1} - \mu)(y_{t-1} - \mu)' \\ & + b_{s_t}b_{s_t}' \odot H_{t-1}(s_t), \end{aligned} \tag{38}$$

where C_{s_t} is a lower triangular 3×3 matrix and a_{s_t} and b_{s_t} are 3×1 vectors. The latent discrete state variable $s_t \in \{1, 2\}$ follows a two-state Markov chain with transition matrix

$$P = \begin{pmatrix} \pi_1 & 1-\pi_1 \\ 1-\pi_2 & \pi_2 \end{pmatrix}. \tag{39}$$

This model can be estimated by a straightforward extension of the methods detailed in Chib (1996) for Markov switching models. Priors are identical to those listed for the breakdown model in Section 4 with the addition of the identification restriction $\pi_1 > \pi_2$. Full sample parameter estimates are found in Table 9. The main difference in states is in $C_{s_t}C_{s_t}'$. In state 2 the elements of this matrix

Table 9
Parameter estimates for MS-VDGARCH-t.

Parameter	Mean	NSE	0.95 DI	Mean	NSE	0.95 DI
	State 1			State 2		
C_{11}	0.0536	0.0071	(0.0396, 0.0674)	0.0942	0.0438	(0.0097, 0.1765)
C_{21}	0.0086	0.0028	(0.0032, 0.0147)	0.0111	0.0155	(−0.0108, 0.0503)
C_{31}	0.0005	0.0003	(−0.0001, 0.0011)	0.0011	0.0017	(−0.0013, 0.0057)
C_{22}	0.0296	0.0032	(0.0231, 0.0358)	0.0478	0.0113	(0.0260, 0.0719)
C_{32}	0.0025	0.0003	(0.0018, 0.0031)	0.0051	0.0014	(0.0016, 0.0079)
C_{33}	0.0030	0.0003	(0.0025, 0.0036)	0.0011	0.0007	(0.0001, 0.0025)
a_1	0.1735	0.0089	(0.1569, 0.1919)	0.2247	0.0371	(0.1551, 0.2994)
a_2	0.1767	0.0075	(0.1624, 0.1924)	0.1108	0.0192	(0.0779, 0.1548)
a_3	0.1829	0.0083	(0.1670, 0.1998)	0.0494	0.0124	(0.0324, 0.0842)
b_1	0.9762	0.0023	(0.9713, 0.9804)	0.9579	0.0114	(0.9331, 0.9770)
b_2	0.9741	0.0021	(0.9694, 0.9780)	0.9735	0.0082	(0.9542, 0.9860)
b_3	0.9701	0.0024	(0.9648, 0.9745)	0.9657	0.0120	(0.9324, 0.9824)
μ_1	0.0612	0.0088	(0.0437, 0.0785)			
μ_2	0.0145	0.0041	(0.0065, 0.0226)			
μ_3	0.0018	0.0004	(0.0010, 0.0026)			
d	5.7944	0.2366	(5.3479, 6.2712)			
π_1	0.9923	0.0047	(0.9798, 0.9980)			
π_2	0.9497	0.0048	(0.9007, 0.9818)			

This table reports the posterior mean, the numerical standard error (NSE), a 0.95 density interval (DI) and the inefficiency factors for the parameters of the MS-VDGARCH-t model. For π_1 and π_2, the following priors are used: $\log(\frac{\pi_1}{1-\pi_1}) \sim N(4, 0.1)$, $\log(\frac{\pi_2}{1-\pi_2}) \sim N(3, 0.1)$, with the restriction $\pi_1 > \pi_2$; For other parameters, the priors are the same as those of the common parameters in the VDGARCH-t model.

Table 10
Model comparison.

Model	Log marginal likelihood	Log(BF)
VDGARCH	−81.01	1662.12
VDGARCH-t	1500.23	80.88
MS-VDGARCH-t	1535.78	45.33
VDGARCH-t-B-2state	1559.81	21.30
VDGARCH-t-B	1581.11	
Model	**Log predictive likelihood**	**Log(PBF)**
VDGARCH-t	467.09	17.59
MS-VDGARCH-t	474.82	9.86
VDGARCH-t-B	484.68	

Log(BF) is the Log-Bayes factor for Student-*t* Breakdown (VDGARCH-t-B) model against each alternative using the full sample while Log(PBF) is the Log-predictive Bayes factor for Student-*t* Breakdown (VDGARCH-t-B) model against the alternatives for the sample 2004/01/2–2011/12/30.

are between 1.8–3.8 times larger than the associated values in state 1. Therefore, state 2 is identified as a high volatility state. The other model parameters are fairly similar over states.

Table 10 reports the results for several MGARCH specifications with Gaussian and Student-*t* innovations, and with or without covariance breakdowns and Markov switching. The final column of the table reports the log-Bayes factor (the difference of the log marginal likelihoods of the two models) for the VDGARCH-t-B against each of its alternatives. This model is strongly favored against all alternatives. The model VDGARCH-t-B-2state is a model that only has one breakdown state ($s_t = 2$) and does not allow the Λ_t to change over the course of one breakdown period. Clearly, having the 2 breakdown states ($s_t = 2, 3$) leads to a significant improvement.

Adding covariance breakdowns improves each MGARCH specification. The log Bayes factor in favor of the VDGARCH-t-B model versus the VDGARCH-t is 80.88, which is overwhelming evidence.[14] Nor is the MS-VDGARCH-t competitive. In fact, the restricted VDGARCH-t-B-2state model is preferred.

The second portion of this table reports the log-predictive likelihoods for the more recent sample 2004/01/2–2011/12/30 and provides some robustness on the full sample marginal likelihoods. Here, only the competitive models are included. We see again that the breakdown model provides superior density forecasts over this sample period with predictive Bayes factors in excess of 9.

In summary, the Bayes factors provide strong support for the VDGARCH-t-B model.

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. 1987–1989

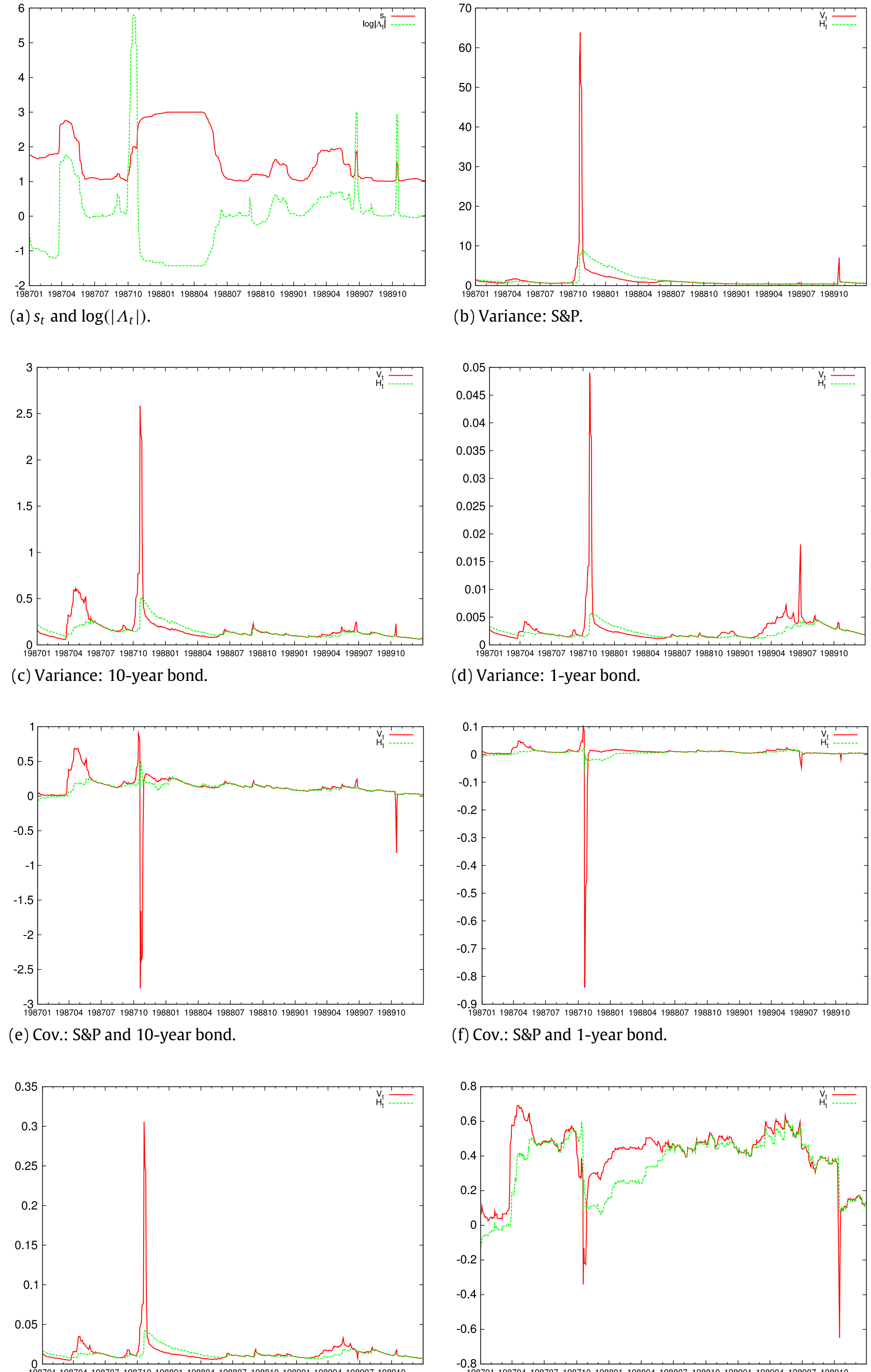
The first episode is from 1987/01/02 to 1989/12/30 and is plotted in Fig. 6. This features the stock market crash in October of 1987, which is identified as a breakdown period in addition to a previous breakdown before the 87 crash.

The main feature of this period is the large increase in conditional variances of all excess returns. The breakdown model implies at least a 4–6 times increase of the conditional variances during the crash compared to those from H_t. The second phase ($s_t = 3$) of the 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 evidence that the breakdown following the crash impacted all conditional correlations.

6.2.2. 2001–2005

The second episode is between 2001/01/02 and 2005/12/30 and found in Fig. 7. Two adjacent extended breakdown periods are identified (Fig. 7(a)). This period shows several short lived breakdowns that cause a sharp increase in the generalized variance along with more sustained breakdowns that result in a drop in

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



(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.

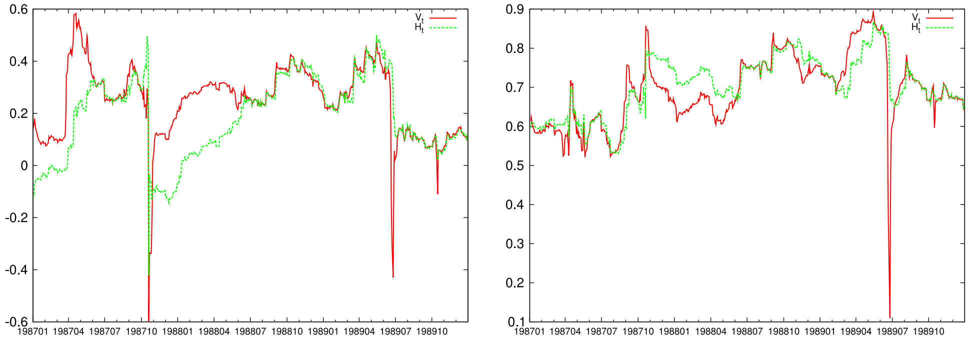
(h) Corr. Coef.: S&P and 10-year bond.

Fig. 6. Volatility components $V_t = H_t^{1/2}\Lambda_t(H_t^{1/2})'$ and H_t: (1987/01/02–1989/12/30).

the generalized variance. Breakdowns consist of both consecutive runs of states 2 and 3 which show the importance of this richer specification.

All three assets display very large increases in their conditional variances. The conditional correlations of the breakdown model show large deviations from those obtained from H_t. This time period was characterized by the bankruptcy case of WorldCom.[15]

15 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).



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

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

Fig. 6. (*continued*)

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

6.2.3. 2008–2011

The final episode is from 2008/01/02 to 2011/12/30 and is found in Fig. 8. This includes the recent financial crisis of 2008. Not surprisingly, two separate 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 two months. During this period the covariance breakdown implies an increase in overall variability of $\exp(3.5) \approx 33$ times.[16]

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. The breakdowns in the conditional variances and covariances tend to cancel out in the conditional correlations. Relative to H_t, there is no evidence of a correlation breakdown in V_t.

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 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:

$$\min_{w_{t+1}} w'_{t+1}\Sigma_{t+1}w_{t+1}, \tag{40}$$

$$\text{s.t.} \quad w'_{t+1}\mu = \mu_0. \tag{41}$$

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

$$w_{t+1} = \frac{\Sigma_{t+1}^{-1}\mu}{\mu'\Sigma_{t+1}^{-1}\mu}\mu_0. \tag{42}$$

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

$$R_{t+1} = w'_{t+1}y_{t+1}. \tag{43}$$

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 and 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.[17] Given a utility function $\mathcal{U}(.)$, we find a constant Δ that equates the total realized utility in

$$\sum_{t=T_0}^{T}\mathcal{U}(R_{1t}) = \sum_{t=T_0}^{T}\mathcal{U}(R_{2t}-\Delta). \tag{44}$$

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 quadratic utility function in Fleming et al. (2001, 2003)

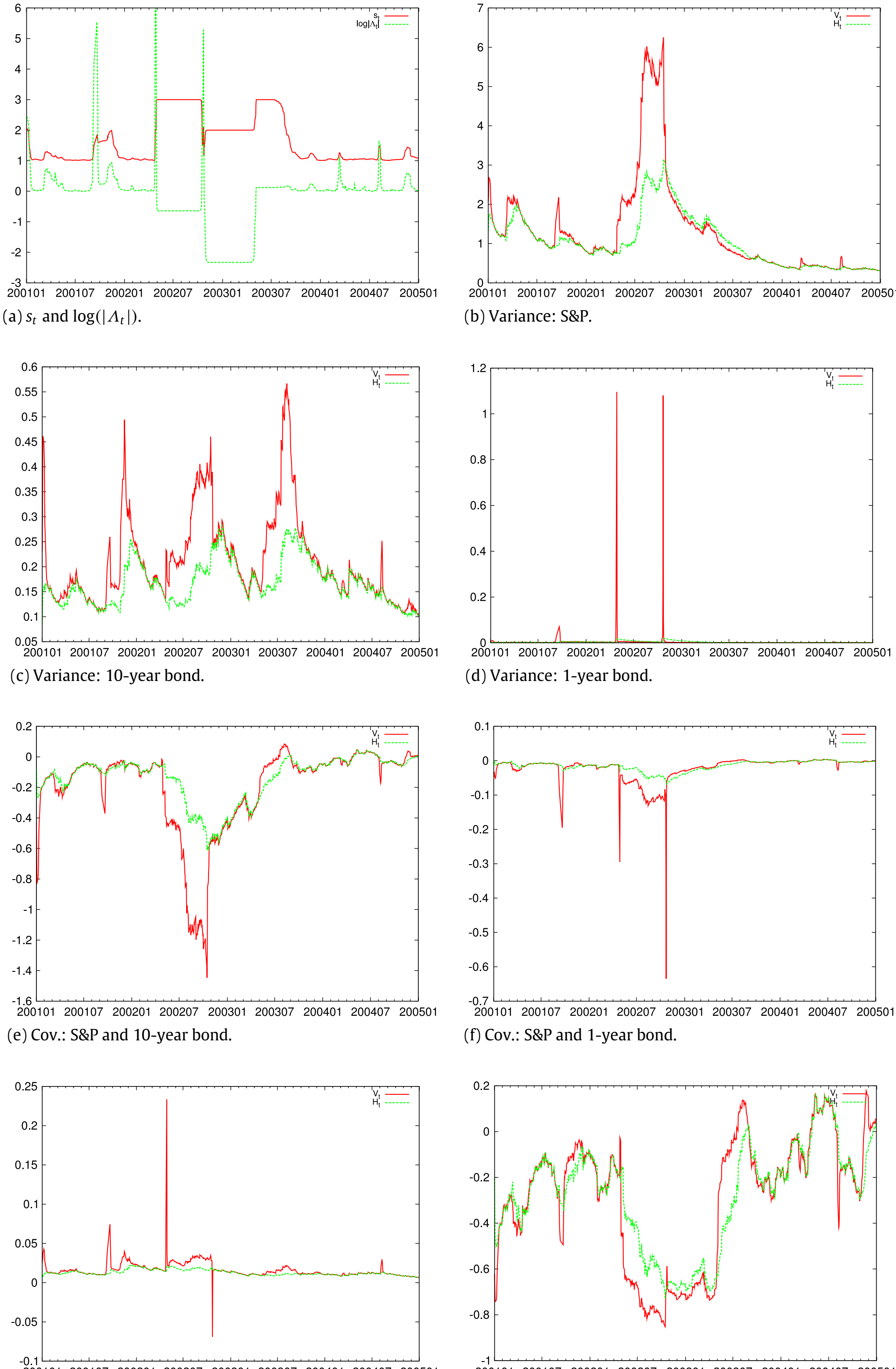
$$\mathcal{U}_q(R_t) = (1+r_{ft}+R_t) - \frac{\tau}{2(1+\tau)}(1+r_{ft}+R_t)^2, \tag{45}$$

where r_{ft} is the risk-free return and τ is the investor's relative 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 (41) to be the sample mean for both models. Any differences in portfolio choice are directly related to the

16 $\log(|\Lambda_t|) \approx 3.5$ during this time period (Fig. 8(a)).

17 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.



(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.

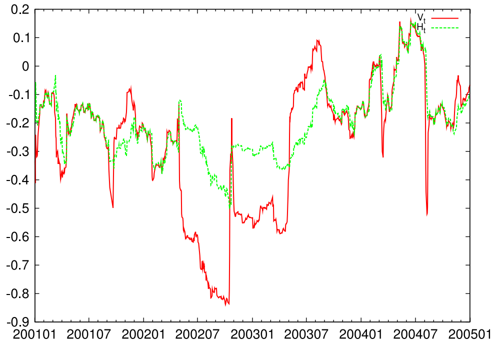
(f) Cov.: S&P and 1-year bond.

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

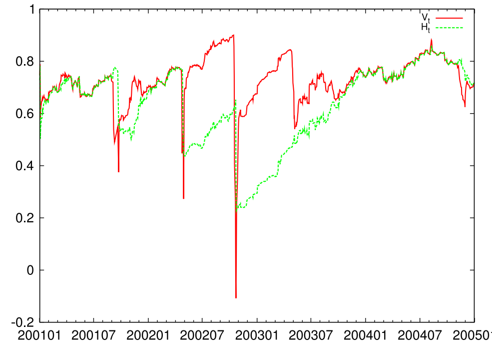
(h) Corr. Coef.: S&P and 10-year bond.

Fig. 7. Volatility components $V_t = H_t^{1/2}\Lambda_t(H_t^{1/2})'$ and H_t: (2001/01/02–2004/12/30).

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.



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



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

Fig. 7. (*continued*)

Table 11
Portfolio performance covariance timing strategies.

μ_0	VDGARCH-t-B			VDGARCH-t			Δ_q	
	$\hat{\mu}_p$	$\hat{\sigma}_p$	SR	$\hat{\mu}_p$	$\hat{\sigma}_p$	SR	$\tau = 1$	$\tau = 10$
Panel A: out-of-sample 1 (20070103–20111230)								
0.01	0.0107	0.1265	0.0840	0.0093	0.1175	0.0791	15.0	14.4
0.03							44.5	39.3
0.05							73.5	59.0
0.07							102.0	73.6
0.09							130.0	82.9
Panel B: out-of-sample 2 (20040102–20111230)								
0.01	0.0035	0.1106	0.0311	0.0027	0.1034	0.0261	13.4	13.0
0.03							39.7	36.4
0.05							65.7	56.7
0.07							91.4	73.7
0.09							116.9	87.5

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). A positive value of Δ_q 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 relative 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. (42)), 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.

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.

Appendix

A.1. Derivation of Eq. (4)

Here we establish that the last term in (4) is a Student-*t* distribution. It is sufficient to show that marginalizing out Λ_t in the distribution of $\Lambda_t^{1/2} z_t$ results in a Student-*t* distribution. The rest of (4) follows from this. Dropping unnecessary subscripts we have

$$\begin{aligned} p(x) &= \int N(x|0, \Lambda) W^{-1}(\Lambda|\nu, Q_0) d\Lambda \\ &\propto \int |\Sigma|^{-1/2} \exp\left(-\frac{1}{2} x' \Lambda^{-1} x\right) |\Lambda|^{-(\nu+k+1)/2} |V|^{\nu/2} \\ &\quad \times \exp\left(-\frac{1}{2} tr(\Lambda^{-1} Q_0)\right) d\Lambda \\ &\propto |xx' + Q_0|^{-(\nu+1)/2} \int |\Lambda|^{-(\nu+k+2)/2} |xx' + Q_0|^{(\nu+1)/2} \\ &\quad \times \exp\left(-\frac{1}{2} tr(\Lambda^{-1}[xx' + Q_0])\right) d\Lambda. \end{aligned} \tag{46}$$

The integrand is the kernel of an inverse-Wishart distribution, so that

$$p(x) \propto |xx' + Q_0|^{-(\nu+1)/2} \tag{47}$$

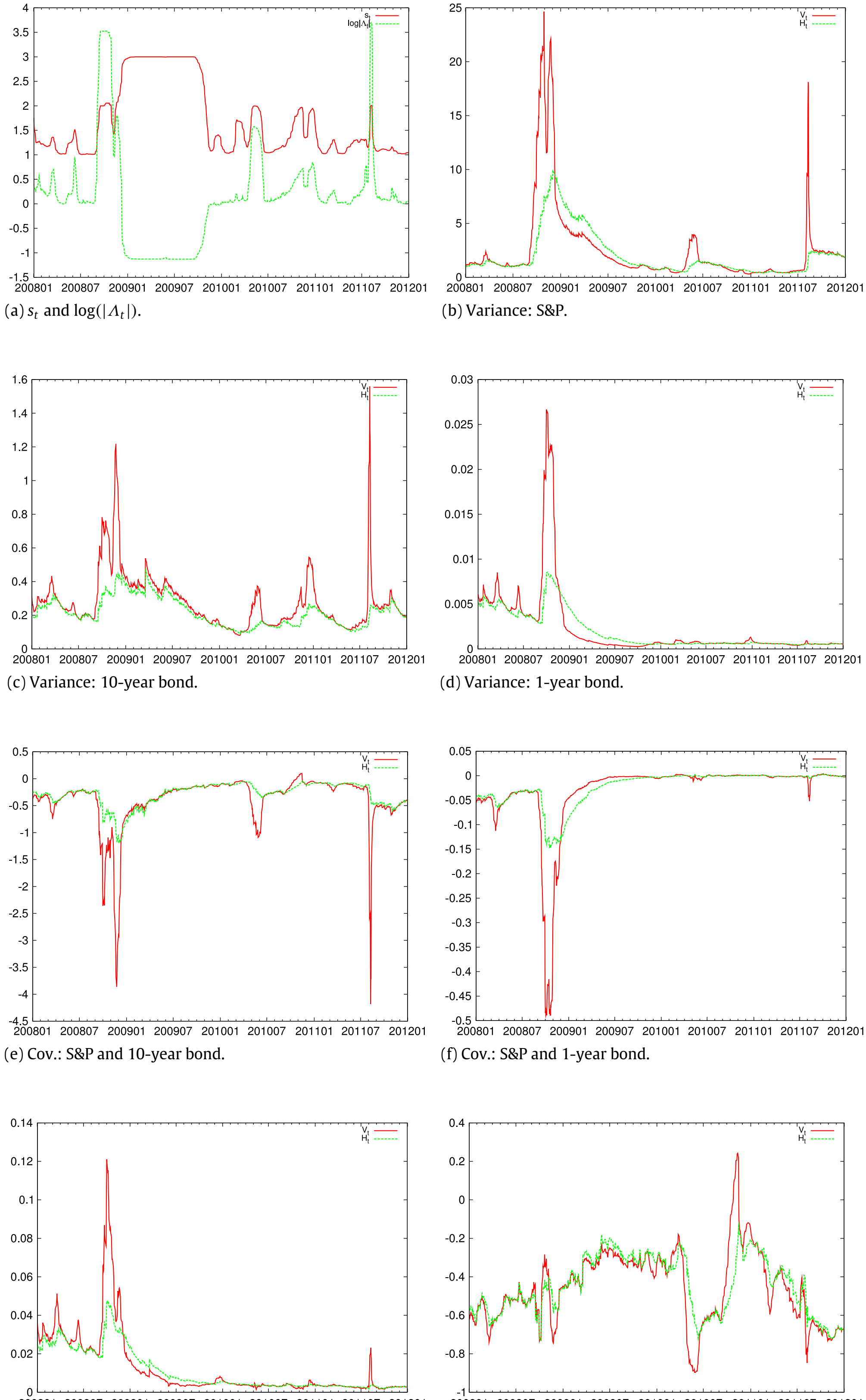
$$\propto (d + x' d Q_0^{-1} x)^{-(d+k)/2} \tag{48}$$

$$\propto t\,(x|0, Q_0/d, d) \tag{49}$$

which is a Student-*t* density with degree of freedom $d = \nu - k + 1$.

A.2. Sampling steps for **S**

Below are each of the cases to consider sampling states.



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

(b) Variance: S&P.

(c) Variance: 10-year bond.

(d) Variance: 1-year bond.

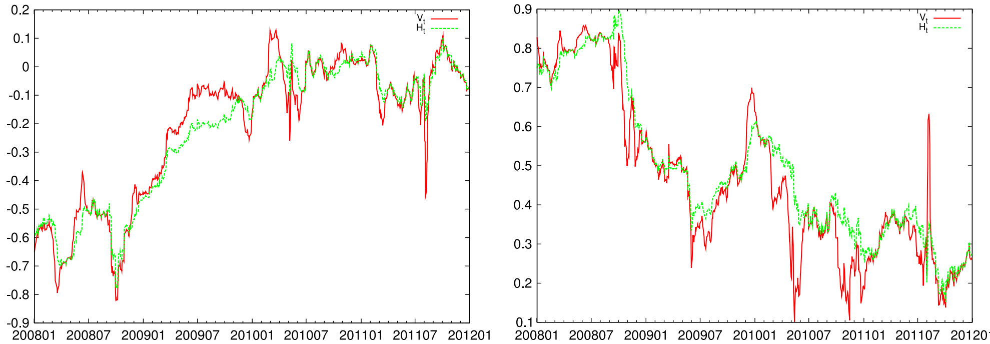
(e) Cov.: S&P and 10-year bond.

(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.

Fig. 8. Volatility components $V_t = H_t^{1/2}\Lambda_t(H_t^{1/2})'$ and H_t: (2008/01/02–2011/12/30).



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

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

Fig. 8. (*continued*)

• $s_{t-1}=1, s_{t+1}=1$

$$Pr(s_t=1|\mathbf{S}_{-t},\tilde{\mathbf{Y}}) \propto p(\tilde{y}_t|s_t=1)Pr(s_t=1|s_{t-1}=1) \times Pr(s_{t+1}=1|s_t=1)$$

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

$$Pr(s_t=3|\mathbf{S}_{-t},\tilde{\mathbf{Y}})=0.$$

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

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

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

$$Pr(s_t=3|\mathbf{S}_{-t},\tilde{\mathbf{Y}})=0.$$

• $s_{t-1}=1, s_{t+1}=\cdots=s_{t+t_2}=3, s_{t+t_2+1}\neq 3$

$$Pr(s_t=1|\mathbf{S}_{-t},\tilde{\mathbf{Y}})=0$$

$$Pr(s_t=2|\mathbf{S}_{-t},\tilde{\mathbf{Y}})=1$$

$$Pr(s_t=3|\mathbf{S}_{-t},\tilde{\mathbf{Y}})=0.$$

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

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

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

$$Pr(s_t=3|\mathbf{S}_{-t},\tilde{\mathbf{Y}}) \propto \left[\int\left(\prod_{j=-t_1}^{-1}p(\tilde{y}_{t+j}|\Lambda)\right)p(\Lambda)d\Lambda\right] \times\left[\int p(\tilde{y}_t|\Lambda)p(\Lambda)d\Lambda\right] \times Pr(s_t=3|s_{t-1}=2)Pr(s_{t+1}=1|s_t=3).$$

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

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

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

$$Pr(s_t=3|\mathbf{S}_{-t},\tilde{\mathbf{Y}}) \propto \left[\int\left(\prod_{j=-t_1}^{-1}p(\tilde{y}_{t+j}|\Lambda)\right)p(\Lambda)d\Lambda\right] \times\left[\int p(\tilde{y}_t|\Lambda)p(\Lambda)d\Lambda\right] \times\left[\int\left(\prod_{j=1}^{t_2}p(\tilde{y}_{t+j}|\Lambda)\right)p(\Lambda)d\Lambda\right] \times Pr(s_t=3|s_{t-1}=2)Pr(s_{t+1}=2|s_t=3).$$

• $s_{t-t_1-1}\neq 2, s_{t-t_1}=\cdots=s_{t-1}=2, s_{t+1}=\cdots=s_{t+t_2}=3, s_{t+t_2+1}\neq 3$

$$Pr(s_t=1|\mathbf{S}_{-t},\tilde{\mathbf{Y}})=0$$

$$Pr(s_t=2|\mathbf{S}_{-t},\tilde{\mathbf{Y}}) \propto \left[\int\left(\prod_{j=-t_1}^{0}p(\tilde{y}_{t+j}|\Lambda)\right)p(\Lambda)d\Lambda\right] \times\left[\int\left(\prod_{j=1}^{t_2}p(\tilde{y}_{t+j}|\Lambda)\right)p(\Lambda)d\Lambda\right] \times Pr(s_t=2|s_{t-1}=2)Pr(s_{t+1}=3|s_t=2)$$

$$Pr(s_t=3|\mathbf{S}_{-t},\tilde{\mathbf{Y}}) \propto \left[\int\left(\prod_{j=-t_1}^{-1}p(\tilde{y}_{t+j}|\Lambda)\right)p(\Lambda)d\Lambda\right] \times\left[\int\left(\prod_{j=0}^{t_2}p(\tilde{y}_{t+j}|\Lambda)\right)p(\Lambda)d\Lambda\right] \times Pr(s_t=3|s_{t-1}=2)Pr(s_{t+1}=3|s_t=3).$$

• $s_{t-t_1-1}\neq 3, s_{t-t_1}=\cdots=s_{t-1}=3, s_{t+1}=1$

$$Pr(s_t=1|\mathbf{S}_{-t},\tilde{\mathbf{Y}}) \propto \left[\int\left(\prod_{j=-t_1}^{-1}p(\tilde{y}_{t+j}|\Lambda)\right)p(\Lambda)d\Lambda\right] \times p(\tilde{y}_t|s_t=1)\times Pr(s_t=1|s_{t-1}=3)Pr(s_{t+1}=1|s_t=1)$$

$$Pr(s_t=2|\mathbf{S}_{-t},\tilde{\mathbf{Y}}) \propto \left[\int\left(\prod_{j=-t_1}^{-1}p(\tilde{y}_{t+j}|\Lambda)\right)p(\Lambda)d\Lambda\right] \times\left[\int p(\tilde{y}_t|\Lambda)p(\Lambda)d\Lambda\right] \times Pr(s_t=2|s_{t-1}=3)Pr(s_{t+1}=1|s_t=2)$$

$$Pr(s_t = 3|\mathbf{S}_{-t}, \tilde{\mathbf{Y}}) \propto \left[\int \left(\prod_{j=-t_1}^{0} p(\tilde{y}_{t+j}|\Lambda)\right) p(\Lambda)d\Lambda\right] \times Pr(s_t = 3|s_{t-1} = 3)Pr(s_{t+1} = 1|s_t = 3).$$

• $s_{t-t_1-1} \neq 3, s_{t-t_1} = \cdots = s_{t-1} = 3, s_{t+1} = \cdots = s_{t+t_2} = 2, s_{t+t_2+1} \neq 2$

$$Pr(s_t = 1|\mathbf{S}_{-t}, \tilde{\mathbf{Y}}) \propto \left[\int \left(\prod_{j=-t_1}^{-1} p(\tilde{y}_{t+j}|\Lambda)\right) p(\Lambda)d\Lambda\right] \times p(\tilde{y}_t|s_t = 1)\left[\int \left(\prod_{j=1}^{t_2} p(\tilde{y}_{t+j}|\Lambda)\right) p(\Lambda)d\Lambda\right] \times Pr(s_t = 1|s_{t-1} = 3)Pr(s_{t+1} = 2|s_t = 1)$$

$$Pr(s_t = 2|\mathbf{S}_{-t}, \tilde{\mathbf{Y}}) \propto \left[\int \left(\prod_{j=-t_1}^{-1} p(\tilde{y}_{t+j}|\Lambda)\right) p(\Lambda)d\Lambda\right] \times \left[\int \left(\prod_{j=0}^{t_2} p(\tilde{y}_{t+j}|\Lambda)\right) p(\Lambda)d\Lambda\right] \times Pr(s_t = 2|s_{t-1} = 3)Pr(s_{t+1} = 2|s_t = 2)$$

$$Pr(s_t = 3|\mathbf{S}_{-t}, \tilde{\mathbf{Y}}) \propto \left[\int \left(\prod_{j=-t_1}^{0} p(\tilde{y}_{t+j}|\Lambda)\right) p(\Lambda)d\Lambda\right] \times \left[\int \left(\prod_{j=1}^{t_2} p(\tilde{y}_{t+j}|\Lambda)\right) p(\Lambda)d\Lambda\right] \times Pr(s_t = 3|s_{t-1} = 3)Pr(s_{t+1} = 2|s_t = 3).$$

• $s_{t-t_1-1} \neq 3, s_{t-t_1} = \cdots = s_{t-1} = 3, s_{t+1} = \cdots = s_{t+t_2} = 3, s_{t+t_2+1} \neq 3$

$$Pr(s_t = 1|\mathbf{S}_{-t}, \tilde{\mathbf{Y}}) = 0$$

$$Pr(s_t = 2|\mathbf{S}_{-t}, \tilde{\mathbf{Y}}) \propto \left[\int \left(\prod_{j=-t_1}^{-1} p(\tilde{y}_{t+j}|\Lambda)\right) p(\Lambda)d\Lambda\right] \times \left[\int p(\tilde{y}_t|\Lambda)p(\Lambda)d\Lambda\right] \times \left[\int \left(\prod_{j=1}^{t_2} p(\tilde{y}_{t+j}|\Lambda)\right) p(\Lambda)d\Lambda\right] \times Pr(s_t = 2|s_{t-1} = 3)Pr(s_{t+1} = 3|s_t = 2)$$

$$Pr(s_t = 3|\mathbf{S}_{-t}, \tilde{\mathbf{Y}}) \propto \left[\int \left(\prod_{j=-t_1}^{t_2} p(\tilde{y}_{t+j}|\Lambda)\right) p(\Lambda)d\Lambda\right] \times Pr(s_t = 3|s_{t-1} = 3)Pr(s_{t+1} = 3|s_t = 3).$$

A.3. Derivation of Eq. (8)

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

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

Therefore,

$$\begin{aligned}&\left(\prod_{t=1}^n p(y_t|\Lambda)\right) p(\Lambda)\\ &= (2\pi)^{-\frac{nk}{2}}|\Lambda|^{-\frac{n}{2}} \exp\left(-\frac{1}{2}\sum_{t=1}^n Tr(\Lambda^{-1}y_ty_t')\right) \times \frac{|Q_0|^{\frac{\nu}{2}}|\Lambda|^{-\frac{\nu+k+1}{2}}}{2^{\frac{\nu k}{2}}\prod_{j=1}^k \Gamma\left(\frac{\nu+1-j}{2}\right)} \exp\left(-\frac{1}{2}Tr(\Lambda^{-1}Q_0)\right)\\ &= \frac{(2\pi)^{-\frac{nk}{2}}|\Lambda|^{-\frac{n+\nu+k+1}{2}}|Q_0|^{\frac{\nu}{2}}}{2^{\frac{\nu k}{2}}\prod_{j=1}^k \Gamma\left(\frac{\nu+1-j}{2}\right)} \times \exp\left(-\frac{1}{2}Tr\left[\Lambda^{-1}\left(\sum_{t=1}^n y_ty_t' + Q_0\right)\right]\right)\\ &= \frac{|Q|^{\frac{n+\nu}{2}}|\Lambda|^{-\frac{n+\nu+k+1}{2}}}{2^{\frac{(n+\nu)k}{2}}\prod_{j=1}^k \Gamma\left(\frac{n+\nu+1-j}{2}\right)} \exp\left(-\frac{1}{2}Tr(\Lambda^{-1}Q)\right) \times \frac{2^{\frac{nk}{2}}(2\pi)^{-\frac{nk}{2}}|Q_0|^{\frac{\nu}{2}}\prod_{j=1}^k \Gamma\left(\frac{n+\nu+1-j}{2}\right)}{|Q|^{\frac{n+\nu}{2}}\prod_{j=1}^k \Gamma\left(\frac{\nu+1-j}{2}\right)}\\ &= W^{-1}(\Lambda|n+\nu, Q) \times \frac{2^{\frac{nk}{2}}(2\pi)^{-\frac{nk}{2}}|Q_0|^{\frac{\nu}{2}}\prod_{j=1}^k \Gamma\left(\frac{n+\nu+1-j}{2}\right)}{|Q|^{\frac{n+\nu}{2}}\prod_{j=1}^k \Gamma\left(\frac{\nu+1-j}{2}\right)} \qquad (50)\end{aligned}$$

where $Q = \sum_{t=1}^n y_ty_t' + Q_0$. So integrating this final result with respect to Λ leaves the second term on the right hand side of (50) since the first term is a well defined inverse-Wishart density that integrates to 1. That is,

$$\int \left(\prod_{t=1}^n p(y_t|\Lambda)\right) p(\Lambda)d\Lambda = \frac{2^{\frac{nk}{2}}(2\pi)^{-\frac{nk}{2}}|Q_0|^{\frac{\nu}{2}}\prod_{j=1}^k \Gamma\left(\frac{n+\nu+1-j}{2}\right)}{|Q|^{\frac{n+\nu}{2}}\prod_{j=1}^k \Gamma\left(\frac{\nu+1-j}{2}\right)}. \qquad (51)$$

A.4. Sampling ν and Q_0

For ν, the conditional posterior distribution is

$$\begin{aligned}p(\nu|\mathbf{Y}, Q_0, \Theta, \mu, \mathbf{S}) &\propto p(\mathbf{Y}|\Theta, \mu, \nu, \mathbf{S}, Q_0)p(\nu)\\ &\propto p(\tilde{\mathbf{Y}}|\mathbf{S}, Q_0, \nu)p(\nu)\\ &\propto \left(\prod_{q=1}^B \frac{|Q_0|^{\frac{\nu}{2}}\prod_{j=1}^k \Gamma\left(\frac{N_q+\nu+1-j}{2}\right)}{|Q_q|^{\frac{\nu}{2}}\prod_{j=1}^k \Gamma\left(\frac{\nu+1-j}{2}\right)}\right) p(\nu). \qquad (52)\end{aligned}$$

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

$$\begin{aligned} p(Q_0|\nu, \{\tilde{\Lambda}_q\}_{q=1}^B) &\propto p(\{\tilde{\Lambda}_q\}_{q=1}^B|Q_0, \nu)p(Q_0) \\ &\propto \left(\prod_{q=1}^B |Q_0|^{\frac{\nu}{2}} \exp\left(-\frac{1}{2}Tr(\tilde{\Lambda}_q^{-1}Q_0)\right)\right) |Q_0|^{\frac{\gamma_0-k-1}{2}} \\ &\quad \times \exp\left(-\frac{1}{2}Tr(A^{-1}Q_0)\right) \\ &\propto |Q_0|^{\frac{B\nu}{2}} \exp\left(-\frac{1}{2}Tr\left(\sum_{q=1}^B \tilde{\Lambda}_q^{-1}Q_0\right)\right) \\ &\quad \times |Q_0|^{\frac{\gamma_0-k-1}{2}} \exp\left(-\frac{1}{2}Tr(A^{-1}Q_0)\right) \\ &\propto |Q_0|^{\frac{B\nu+\gamma_0-k-1}{2}} \exp\left(-\frac{1}{2}Tr\left(\left(\sum_{q=1}^B \tilde{\Lambda}_q^{-1} + A^{-1}\right)Q_0\right)\right) \\ &\propto W_k(Q_0|\gamma, \widetilde{A}) \end{aligned} \tag{53}$$

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

A.5. Marginal likelihood estimation

Below the estimation of each of the components of (35) is provided.

A.5.1. Estimating $f(\mathbf{Y}|\Psi)$

We assume Student-t innovation for the data[18]:

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

The parameter set Ψ includes $\Theta, \mu, \pi, d, \nu, Q_0$, where $\pi = \{\pi_1, \pi_2, \pi_3, \pi_4, \pi_5\}$. Write $\tilde{Y}_t = \{\tilde{y}_t\}_{t=1}^T$ and $\tilde{y}_t = H_t^{-1/2}(y_t - \mu)$. Note that

$$f(\mathbf{Y}|\Psi) = \left[\prod_{t=1}^T |H_t|^{-1/2}\right] f(\tilde{\mathbf{Y}}|\pi, d, \nu, Q_0). \tag{54}$$

The likelihood function can be obtained by computing $\prod_{t=1}^T |H_t|^{-1/2}$ and $f(\tilde{\mathbf{Y}}|\Psi_1)$, where $\Psi_1 = \{\pi, d, \nu, Q_0\}$. 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:

$$\tilde{y}_t|\Lambda_t \sim t(0, \Lambda_t, d). \tag{55}$$

The transformed likelihood function can be written as

$$\begin{aligned} f(\tilde{\mathbf{Y}}|\Psi_1) &= \prod_{t=1}^T f(\tilde{y}_t|\tilde{\mathbf{Y}}_{t-1}, \Psi_1) \\ &= \prod_{t=1}^T \int f(\tilde{y}_t|\Lambda_t)f(\Lambda_t|\tilde{\mathbf{Y}}_{t-1}, \Psi_1)d\Lambda_t \\ &= \prod_{t=1}^T \int t(\tilde{y}_t|0, \Lambda_t, d)f(\Lambda_t|\tilde{\mathbf{Y}}_{t-1}, \Psi_1)d\Lambda_t, \end{aligned} \tag{56}$$

where $\tilde{\mathbf{Y}}_t = \{\tilde{y}_i\}_{i=1}^t$, and $t(.|., ., .)$ is the density of the Student-t distribution. To approximate the likelihood function, we design an Auxiliary Particle Filter (APF) (Pitt and Shephard, 1999) to sequentially sample from the filtering distribution $f(\Lambda_t, s_t|\tilde{\mathbf{Y}}_t, \Psi_1)$, $t = 1, \ldots, T$. In the following Ψ_1 is suppressed from the conditioning set.

Given M particles $\{(\Lambda_t^{(j)}, s_t^{(j)})\}_{j=1}^M$ from $f(\Lambda_t, s_t|\tilde{\mathbf{Y}}_t)$, each with the same discrete probability mass $1/M$, the predictive density is

$$\begin{aligned} f(\Lambda_{t+1}, s_{t+1}|\tilde{\mathbf{Y}}_t) &= \int f(\Lambda_{t+1}, s_{t+1}|\Lambda_t, s_t)f(\Lambda_t, s_t|\tilde{\mathbf{Y}}_t)d(\Lambda_t, s_t) \\ &\approx \sum_{j=1}^M f(\Lambda_{t+1}, s_{t+1}|\Lambda_t^{(j)}, s_t^{(j)})\frac{1}{M}. \end{aligned} \tag{57}$$

Therefore

$$\begin{aligned} &f(\Lambda_{t+1}, s_{t+1}|\tilde{\mathbf{Y}}_{t+1}) \\ &\quad \propto f(\tilde{y}_{t+1}|\Lambda_{t+1}, s_{t+1}) \sum_{j=1}^M f(\Lambda_{t+1}, s_{t+1}|\Lambda_t^{(j)}, s_t^{(j)}) \\ &\quad = f(\tilde{y}_{t+1}|\Lambda_{t+1}) \sum_{j=1}^M f(\Lambda_{t+1}, s_{t+1}|\Lambda_t^{(j)}, s_t^{(j)}). \end{aligned} \tag{58}$$

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

$$\begin{aligned} &f(\Lambda_{t+1}, s_{t+1}, m|\tilde{\mathbf{Y}}_{t+1}) \\ &\quad \propto f(\tilde{y}_{t+1}|\Lambda_{t+1})f(\Lambda_{t+1}, s_{t+1}|\Lambda_t^{(m)}, s_t^{(m)}). \end{aligned} \tag{59}$$

If we draw from this joint distribution in (59) and then discard the index m, we will produce a sample from the distribution in (58).

We deploy a Sampling/Importance Resampling (SIR) algorithm to sample from $f(\Lambda_{t+1}, s_{t+1}, m|\tilde{\mathbf{Y}}_{t+1})$. That is, in order to produce M draws from $f(\Lambda_{t+1}, s_{t+1}, m|\tilde{\mathbf{Y}}_{t+1})$, instead of sampling from $f(\Lambda_{t+1}, s_{t+1}, m|\tilde{\mathbf{Y}}_{t+1})$ directly, we first sample $R \gg M$ draws from a proposal distribution with density $h(\Lambda_{t+1}, s_{t+1}, m|\tilde{\mathbf{Y}}_{t+1})$, and reweight each of these draws with the "correct" weight w_i, and then sample M draws from the reweighted sample. The detail is as follows.

Given the M particles $\{(\Lambda_t^{(j)}, s_t^{(j)})\}_{j=1}^M$ from $f(\Lambda_t, s_t|\tilde{\mathbf{Y}}_t)$, for each $j \in \{1, \ldots, M\}$, draw

$$(\tilde{\Lambda}_{t+1}^{(j)}, \tilde{s}_{t+1}^{(j)}) \sim f(\Lambda_{t+1}, s_{t+1}|\Lambda_t^{(j)}, s_t^{(j)}). \tag{60}$$

We approximate (59) by

$$h(\Lambda_{t+1}, s_{t+1}, m|\tilde{\mathbf{Y}}_{t+1}) \propto f(\tilde{y}_{t+1}|\tilde{\Lambda}_{t+1}^{(m)})f(\Lambda_{t+1}, s_{t+1}|\Lambda_t^{(m)}, s_t^{(m)}). \tag{61}$$

Marginalizing out (Λ_{t+1}, s_{t+1}) we get

$$\begin{aligned} &h(m|\tilde{\mathbf{Y}}_{t+1}) \\ &\quad \propto \int f(\tilde{y}_{t+1}|\tilde{\Lambda}_{t+1}^{(m)})f(\Lambda_{t+1}, s_{t+1}|\Lambda_t^{(m)}, s_t^{(m)})d(\Lambda_{t+1}, s_{t+1}) \\ &\quad = f(\tilde{y}_{t+1}|\tilde{\Lambda}_{t+1}^{(m)}). \end{aligned} \tag{62}$$

Thus we sample from $h(\Lambda_{t+1}, s_{t+1}, m|\tilde{\mathbf{Y}}_{t+1})$ by first sampling m with probability $\lambda_m \propto h(m|\tilde{\mathbf{Y}}_{t+1})$, and then sampling $(\Lambda_{t+1}, s_{t+1}) \sim f(\Lambda_{t+1}, s_{t+1}|\Lambda_t^{(m)}, s_t^{(m)})$. λ_m are called the first-stage weights.

Having sampled the joint density of $h(\Lambda_{t+1}, s_{t+1}, m|\tilde{\mathbf{Y}}_{t+1})$ R times, a reweighting is performed by putting on the draw $(\Lambda_{t+1}^{(i)}, s_{t+1}^{(i)}, m^{(i)})$ the weights in proportion to the so-called second-stage weights

$$w_i = \frac{f(\tilde{y}_{t+1}|\Lambda_{t+1}^{(i)})}{f(\tilde{y}_{t+1}|\tilde{\Lambda}_{t+1}^{(m^{(i)})})}, \quad i = 1, \ldots, R.$$

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

We then resample from this discrete distribution $\{(\Lambda_{t+1}^{(i)}, s_{t+1}^{(i)}, m^{(i)})\}_{i=1}^{R}$ to produce a sample of size M, which gives the desired particles from $f(\Lambda_{t+1}, s_{t+1}|\tilde{\mathbf{Y}}_{t+1})$.

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 $(\Lambda_{t-1}^{(j)}, s_{t-1}^{(j)}), j = 1, \ldots, M$.
2. For each $(\Lambda_{t-1}^{(j)}, s_{t-1}^{(j)})$, sample $(\Lambda_t^{(j)}, s_t^{(j)})$ from $f(\Lambda_t, s_t|\Lambda_{t-1}^{(j)}, s_{t-1}^{(j)})$ according to the transition distribution defined in (2) and (3).
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 $\sum_{t=1}^{T} \log \hat{f}(\tilde{y}_t|\tilde{\mathbf{Y}}_{t-1}, \Psi_1)$.

A.5.2. Estimating $f(\Psi|\mathbf{Y})$

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

$$\begin{aligned} f(\Psi^*|\mathbf{Y}) &= f(\Theta^*|\mathbf{Y}) \times f(\mu^*|\mathbf{Y}, \Theta^*) \times f(\tilde{\pi}^*|\mathbf{Y}, \Theta^*, \mu^*) \\ &\quad \times f(d^*|\mathbf{Y}, \Theta^*, \mu^*, \pi^*) \\ &\quad \times f(\nu^*|\mathbf{Y}, \Theta^*, \mu^*, \pi^*, d^*) \times f(Q_0^*|\mathbf{Y}, \Theta^*, \mu^*, \pi^*, d^*, \nu^*), \end{aligned} \quad (63)$$

where $\tilde{\pi} = \{\log \frac{\pi_i}{1-\pi_i}\}_{i=1}^{5}$. 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 and Jeliazkov (2001), the ordinate $f(\Theta^*|\mathbf{Y})$ can be expressed as in Box I.

The numerator can be estimated as

$$\frac{1}{M}\sum_{j=1}^{M} \alpha(\Theta^{(j)}, \Theta^*|\mathbf{Y}, \mu^{(j)}, \nu^{(j)}, Q_0^{(j)}, \mathbf{S}^{(j)}, \mathbf{U}^{(j)}) q(\Theta^{(j)}, \Theta^*), \quad (65)$$

where $(\Theta^{(j)}, \mu^{(j)}, \nu^{(j)}, Q_0^{(j)}, \mathbf{S}^{(j)}, \mathbf{U}^{(j)})$ is from the jth draw of the full MCMC run of the posterior distribution $f(\Psi, \mathbf{S}, \mathbf{U}, \boldsymbol{\Lambda}|\mathbf{Y})$, which consists of $\{\Theta^{(j)}, \mu^{(j)}, \pi^{(j)}, d^{(j)}, \nu^{(j)}, Q_0^{(j)}, \mathbf{S}^{(j)}, \mathbf{U}^{(j)}, \boldsymbol{\Lambda}^{(j)}\}$. 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 $(\mathbf{S}, \mathbf{U}, \boldsymbol{\Lambda})$ 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

$$\frac{1}{M}\sum_{l=1}^{M} \alpha(\Theta^*, \Theta^{(l)}|\mathbf{Y}, \mu^{(l)}, \nu^{(l)}, Q_0^{(l)}, \mathbf{S}^{(l)}, \mathbf{U}^{(l)}). \quad (66)$$

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

$$f(\mu^*|\mathbf{Y}, \Theta^*) = \frac{\int \alpha(\mu, \mu^*|\mathbf{Y}, \Theta^*, d, \boldsymbol{\Lambda}) q(\mu, \mu^*) f(\mu, d, \boldsymbol{\Lambda}|\mathbf{Y}, \Theta^*) d(\mu, \boldsymbol{\Lambda}, d)}{\int \alpha(\mu^*, \mu|\mathbf{Y}, \Theta^*, d, \boldsymbol{\Lambda}) q(\mu^*, \mu) f(d, \boldsymbol{\Lambda}|\mathbf{Y}, \Theta^*, \mu^*) d(\mu, \boldsymbol{\Lambda}, d)}. \quad (67)$$

The numerator can be estimated as

$$\frac{1}{M}\sum_{j=1}^{M} \alpha(\mu^{(j)}, \mu^*|\mathbf{Y}, \Theta^*, d^{(j)}, \boldsymbol{\Lambda}^{(j)}) q(\mu^{(j)}, \mu^*), \quad (68)$$

where $(\mu^{(j)}, d^{(j)}, \boldsymbol{\Lambda}^{(j)})$ is from the jth 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 (64). For the denominator in (67), the integral is with respect to $f(d, \boldsymbol{\Lambda}|\mathbf{Y}, \Theta^*, \mu^*) \times q(\mu^*, \mu)$. 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 $(\mu^{(l)}, d^{(l)}, \boldsymbol{\Lambda}^{(l)})$ from the distribution $f(d, \boldsymbol{\Lambda}|\mathbf{Y}, \Theta^*, \mu^*)q(\mu^*, \mu)$. The denominator is estimated as

$$\frac{1}{M}\sum_{l=1}^{M} \alpha(\mu^*, \mu^{(l)}|\mathbf{Y}, \Theta^*, d^{(l)}, \boldsymbol{\Lambda}^{(l)}). \quad (69)$$

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

$$f(\tilde{\pi}^*|\mathbf{Y}, \Theta^*, \mu^*) = \frac{\int \alpha(\tilde{\pi}, \tilde{\pi}^*|\mathbf{S}) q(\tilde{\pi}, \tilde{\pi}^*) f(\tilde{\pi}, \mathbf{S}|\mathbf{Y}, \Theta^*, \mu^*) d(\mathbf{S}, \tilde{\pi})}{\int \alpha(\tilde{\pi}^*, \tilde{\pi}|\mathbf{S}) q(\tilde{\pi}^*, \tilde{\pi}) f(\mathbf{S}|\mathbf{Y}, \Theta^*, \mu^*, \tilde{\pi}^*) d(\mathbf{S}, \tilde{\pi})}. \quad (70)$$

The numerator can be estimated as

$$\frac{1}{M}\sum_{j=1}^{M} \alpha(\pi^{(j)}, \pi^*|\mathbf{S}^{(j)}) q(\tilde{\pi}^{(j)}, \tilde{\pi}^*), \quad (71)$$

where $(\tilde{\pi}^{(j)}, \mathbf{S}^{(j)})$ is from the jth 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 (67). To estimate the integral in the denominator of (70), fix $\Theta, \mu, \tilde{\pi}$ at $\Theta^*, \mu^*, \tilde{\pi}^*$, 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 $\tilde{\pi}$. At each iteration l, draw $\tilde{\pi}^{(l)}$ from the proposal density $q(\tilde{\pi}^*, \tilde{\pi})$. The resulting $(\mathbf{S}^{(l)}, \tilde{\pi}^{(l)}), l = 1, \ldots, M$ are draws from the distribution $f(\mathbf{S}|\mathbf{Y}, \Theta^*, \mu^*, \tilde{\pi}^*)q(\tilde{\pi}^*, \tilde{\pi})$. The denominator is estimated as

$$\frac{1}{M}\sum_{l=1}^{M} \alpha(\pi^*, \pi^{(l)}|\mathbf{S}^{(l)}). \quad (72)$$

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

$$f(d^*|\mathbf{Y}, \Theta^*, \mu^*, \pi^*) = \frac{\int \alpha(d, d^*|\mathbf{U}) q(d, d^*) f(d, \mathbf{U}|\mathbf{Y}, \Theta^*, \mu^*, \pi^*) d(\mathbf{U}, d)}{\int \alpha(d^*, d|\mathbf{U}) q(d^*, d) f(\mathbf{U}|\mathbf{Y}, \Theta^*, \mu^*, \pi^*, d^*) d(\mathbf{U}, d)}. \quad (73)$$

The numerator can be estimated as

$$\frac{1}{M}\sum_{j=1}^{M} \alpha(d^{(j)}, d^*|\mathbf{U}^{(j)}) q(d^{(j)}, d^*), \quad (74)$$

where $(d^{(j)}, \mathbf{U}^{(j)})$ is from the jth draw of the third reduced MCMC run of the posterior distribution used in the previous step to estimate the denominator of (70). To estimate the integral in the denominator of (73), fix Θ, μ, π, d at $\Theta^*, \mu^*, \pi^*, d^*$, 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

$$\frac{1}{M}\sum_{l=1}^{M} \alpha(d^*, d^{(l)}|\mathbf{U}^{(l)}). \quad (75)$$

The ordinate $f(\nu^*|\mathbf{Y}, \Theta^*, \mu^*, \pi^*, d^*)$ can be expressed as in Box II.

$$f(\Theta^*|\mathbf{Y}) = \frac{\int \alpha(\Theta, \Theta^*|\mathbf{Y}, \mu, \nu, Q_0, \mathbf{S}, \mathbf{U}) q(\Theta, \Theta^*) f(\Theta, \mu, \nu, Q_0, \mathbf{S}, \mathbf{U}|\mathbf{Y}) d(\Theta, \mu, \nu, Q_0, \mathbf{S}, \mathbf{U})}{\int \alpha(\Theta^*, \Theta|\mathbf{Y}, \mu, \nu, Q_0, \mathbf{S}, \mathbf{U}) q(\Theta^*, \Theta) f(\mu, \nu, Q_0, \mathbf{S}, \mathbf{U}|\mathbf{Y}, \Theta^*) d(\Theta, \mu, \nu, Q_0, \mathbf{S}, \mathbf{U})}. \tag{64}$$

Box I.

$$f(\nu^*|\mathbf{Y}, \Theta^*, \mu^*, \pi^*, d^*) = \frac{\int \alpha(\nu, \nu^*|\mathbf{Y}, \Theta^*, \mu^*, Q_0, \mathbf{S}, \mathbf{U}) q(\nu, \nu^*) f(\nu, Q_0, \mathbf{S}, \mathbf{U}|\mathbf{Y}, \Theta^*, \mu^*, \pi^*, d^*) d(\nu, Q_0, \mathbf{S}, \mathbf{U})}{\int \alpha(\nu^*, \nu|\mathbf{Y}, \Theta^*, \mu^*, Q_0, \mathbf{S}, \mathbf{U}) q(\nu^*, \nu) f(Q_0, \mathbf{S}, \mathbf{U}|\mathbf{Y}, \Theta^*, \mu^*, \pi^*, d^*, \nu^*) d(\nu, Q_0, \mathbf{S}, \mathbf{U})}. \tag{76}$$

Box II.

The numerator can be estimated as

$$\frac{1}{M} \sum_{j=1}^{M} \alpha(\nu^{(j)}, \nu^*|\mathbf{Y}, \Theta^*, \mu^*, Q_0^{(j)}, \mathbf{S}^{(j)}, \mathbf{U}^{(j)}) q(\nu^{(j)}, \nu^*), \tag{77}$$

where $(\nu^{(j)}, Q_0^{(j)}, \mathbf{S}^{(j)}, \mathbf{U}^{(j)})$ is from the jth draw of the fourth reduced MCMC run of the posterior distribution used in the previous step to estimate the denominator of (73). To estimate the integral in the denominator of (76), 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

$$\frac{1}{M} \sum_{l=1}^{M} \alpha(\nu^*, \nu^{(l)}|\mathbf{Y}, \Theta^*, \mu^*, Q_0^{(l)}, \mathbf{S}^{(l)}, \mathbf{U}^{(l)}). \tag{78}$$

Since the conditional posterior distribution of Q_0 is sampled using a Gibbs step,

$f(Q_0^*|\mathbf{Y}, \Theta^*, \mu^*, \pi^*, d^*, \nu^*)$ can be estimated by

$$f(Q_0^*|\mathbf{Y}, \Theta^*, \mu^*, \pi^*, , d^*, \nu^*) \approx \frac{1}{M} \sum_{j=1}^{M} f(Q_0^*|\boldsymbol{\Lambda}^{(j)}, \nu^*), \tag{79}$$

where $f(Q_0|\boldsymbol{\Lambda}, \nu) = W_k(Q_0|\gamma, \widetilde{A})$ (both γ and $\widetilde{A}$ depend on ν and $\boldsymbol{\Lambda}$, see Appendix A.4), and $\boldsymbol{\Lambda}^{(j)}$ is from the jth draw of the fifth reduced MCMC run of the posterior distribution used in the previous step to estimate the denominator of (76).

A.5.3. Estimating $f(\Psi)$

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

$f(\Psi^*) = f(\Theta^*) f(\mu^*) f(\tilde{\pi}^*) f(d^*) f(\nu^*) f(Q_0^*)$.

Evaluating the prior ordinates is straightforward for all blocks, as they are standard distributions such as (truncated) normal or exponential.

References

Akhigbe, A., Martin, A.D., Whyte, A.M., 2005. Contagion effects of the world's largest bankruptcy: the case of worldcom. Q. Rev. Econ. Finance 45 (1), 48–64.
Ang, A., Bekaert, G., 2004. How regimes affect asset allocation. Financ. Anal. J. 60 (2), 86–99.
Bauwens, L., Dufays, A., Rombouts, J.V., 2014. Marginal likelihood for Markov-switching and change-point GARCH models. J. Econometrics 178, 508–522.
Bauwens, L., Laurent, S., Rombouts, J., 2006. Multivariate GARCH models: a survey. J. Appl. Econometrics 21 (1), 79–109.
Cappiello, L., Engle, R., Sheppard, K., 2006. Asymmetric dynamics in the correlations of global equity and bond returns. J. Financ. Econ. 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. J. Amer. Statist. Assoc. 90 (432), 1313–1321.
Chib, S., 1996. Calculating posterior distributions and modal estimates in Markov mixture models. J. Econometrics 75, 79–97.
Chib, S., Jeliazkov, I., 2001. Marginal likelihood from the Metropolis-Hasting output. J. Amer. Statist. Assoc. 96, 270–281.
Christoffersen, P., Errunza, V., Jacobs, K., Langlois, H., 2012. Is the potential for international diversification disappearing? A dynamic copula approach. Rev. Financ. Stud. 25 (12), 3711–3751.
Clements, A., Silvennoinen, A., 2013. Volatility timing: How best to forecast portfolio exposures. J. Empir. Finance 24 (0), 108–115.
De Goeij, P., Marquering, W., 2004. Modeling the conditional covariance between stock and bond returns: A multivariate GARCH approach. J. Financ. Econ. 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. J. Bus. Econom. Statist. 20 (3), 339–350.
Fleming, J., Kirby, C., Ostdiek, B., 2001. The economic value of volatility timing. J. Finance 56, 329–352.
Fleming, J., Kirby, C., Ostdiek, B., 2003. The economic value of volatility timing using realized volatility. J. Financ. Econ. 67 (3), 473–509.
Forbes, K.J., Rigobon, R., 2002. No contagion, only interdependence: Measuring stock market comovements. J. Finance 57 (5), 2223–2261.
Geweke, J., 1993. Bayesian treatment of the independent Student-t linear model. J. Appl. Econometrics 8, S19–S40.
Geweke, J., 2005. Contemporary Bayesian Econometrics and Statistics. Wiley.
Guidolin, M., Timmermann, A., 2006. An econometric model of nonlinear dynamics in the joint distribution of stock and bond returns. J. Appl. Econometrics 21 (1), 1–22.
Haas, M., Mittnik, S., 2008. Multivariate Regime Switching GARCH with an Application to International Stock Markets, CFS Working Paper Series 2008/08, Center for Financial Studies (CFS).
Haas, M., Mittnik, S., Paolella, M., 2004. A new approach to Markov-switching GARCH models. J. Financ. Econ. 2 (4), 493–530.
He, Z., Maheu, J., 2010. Real time detection of structural breaks in GARCH models. Comput. Stat. Data Anal. 54 (11), 2628–2640.
Hillebrand, E., 2005. Neglecting parameter changes in GARCH models. J. Econometrics 129 (1–2), 121–138.
Kass, R.E., Raftery, A.E., 1995. Bayes factors. J. Amer. Statist. Assoc. 90 (420), 773–795.
Kenourgios, D., Samitas, A., Paltalidis, N., 2011. Financial crises and stock market contagion in a multivariate time-varying asymmetric framework. J. Int. Financ. Mark. Inst. Money 21 (1), 92–106.
Klaassen, F., 2002. Improving GARCH volatility forecasts with regime-switching GARCH. Empir. Econom. 27 (2), 363–394.
Lamoureux, C.G., Lestrapes, W.D., 1990. Persistence in variance, structural change, and the GARCH model. J. Bus. Econom. Statist. 8 (2), 225–234.
Ledoit, O., Santa-Clara, P., Wolf, M., 2003. Flexible multivariate GARCH modeling with an application to international stock markets. Rev. Econ. Stat. 85 (3), 735–747.
Muirhead, R., 1982. Aspects of Multivariate Statistical Theory, Vol. 42. Wiley Online Library.
Pelletier, D., 2006. Regime switching for dynamic correlations. J. Econometrics 131 (1), 445–473.
Pitt, M., Shephard, N., 1999. Filtering via simulation: Auxiliary particle filters. J. Amer. Statist. Assoc. 94 (446), 590–599.
Rangel, J.G., Engle, R.F., 2012. The Factor-Spline-GARCH model for high and low frequency correlations. J. Bus. Econom. Statist. 30 (1), 109–124.
Silvennoinen, A., Teräsvirta, T., 2009. Modeling multivariate autoregressive conditional heteroskedasticity with the double smooth transition conditional correlation GARCH model. J. Financ. Econ. 7 (4), 373–411.