University of Toronto
Department of Economics

Copula Multivariate GARCH Model with Constrained Hamiltonian Monte Carlo

By Martin Burda and Louis Belisle

April 29, 2019

Copula Multivariate GARCH Model with Constrained Hamiltonian Monte Carlo

Martin Burda[†]   Louis Bélisle[‡]

April 17, 2019

Abstract

The Copula Multivariate GARCH (CMGARCH) model is based on a dynamic copula function with time-varying parameters. It is particularly suited for modelling dynamic dependence of non-elliptically distributed financial returns series. The model allows for capturing more flexible dependence patterns than a multivariate GARCH model and also generalizes static copula dependence models. Nonetheless, the model is subject to a number of parameter constraints that ensure positivity of variances and covariance stationarity of the modeled stochastic processes. As such, the resulting distribution of parameters of interest is highly irregular, characterized by skewness, asymmetry, and truncation, hindering the applicability and accuracy of asymptotic inference. In this paper, we propose Bayesian analysis of the CMGARCH model based on Constrained Hamiltonian Monte Carlo (CHMC), which has been shown in other contexts to yield efficient inference on complicated constrained dependence structures. In the CMGARCH context, we contrast CHMC with traditional random-walk sampling used in the previous literature and highlight the benefits of CHMC for applied researchers. We estimate the posterior mean, median and Bayesian confidence intervals for the coefficients of tail dependence. The analysis is performed in an application to a recent portfolio of S&P500 financial asset returns.

[†]Department of Economics, University of Toronto, 150 St. George St., Toronto, ON M5S 3G7, Canada; Phone: (416) 978-4479; Email: martin.burda@utoronto.ca

[‡]Department of Economics, University of Toronto, 150 St. George St., Toronto, ON M5S 3G7, Canada; Email: louis.salmon.belisle@mail.utoronto.ca

1 Introduction

Modeling, predicting, and understanding the volatility of financial time series lies in the core of financial econometrics. Volatility prediction has a very wide range of applications, such as in portfolio optimization, risk management, asset allocation, and asset pricing. Since the seminal papers of Engle (1982) and Bollerslev (1987) introducing the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) modeling framework, voluminous research has been conducted on the topic, encompassing various model extensions and generalizations. The usual assumption in multivariate GARCH (MGARCH) models is that the conditional joint distribution of the returns follows a multivariate normal or multivariate t-distribution (see e.g. the Bauwens et al. (2006) survey). However, these elliptical distribution models require a very strong symmetry of the data and may not be appropriate in many circumstances.

Researchers have been addressing two key aspects of volatility analysis: looking for the most appropriate model specification and selecting the most efficient approach for inference and prediction. In this paper, for the modeling aspect we adopt the copula multivariate GARCH (CMGARCH) model (Lee and Long, 2009), featuring very appealing dynamic dependence structure for financial time series. We then focus on the latter aspect of inference and prediction by proposing Bayesian inference via Constrained Hamiltonian Monte Carlo (CHMC), which we show is particularly suited for practical inference relative to traditional random-walk based Markov Chain Monte Carlo (RW-MCMC) methods utilized for the CMGARCH in the previous literature (Ausín and Lopes, 2010). In addition to parameter inference, we show that CHMC facilitates computationally efficient estimation of the posterior mean, median and Bayesian confidence intervals for the coefficients of tail dependence.

Traditional approaches measure dependence by linear correlation, quantified by the Pearson's correlation coefficient. However, linear correlation expresses a symmetric linear notion of dependence. Modern asset returns present nonlinear, non-Gaussian, and dynamic features. Copula functions, a powerful tool originating from mathematical analysis, can capture the dependence structures among financial time series without the standard linear dependence and multivariate i.i.d return distribution restrictions. Since the 1990s, copula-based analysis have entered the core of dependence modeling in finance (Nelsen, 1999; Embrechts et al., 2001).

A recent stream of literature joins the dynamic MGARCH modeling framework with the cop-

ula dependence structure. The key advantage of this approach is that the individual marginal densities of the returns can be defined separately from their dependence structure. Using this approach, many non-elliptical and flexible multivariate distributions can be obtained, capturing the corresponding features of financial portfolio series. In this context, most researchers have specified the marginal series following univariate GARCH processes and captured the dependence structure between them by a copula function to form the CMGARCH model (Dias and Embrechts, 2004; Rodríguez, 2007; Hu, 2006; Jondeau and Rockinger, 2006; Liu and Luger, 2009). However, the CMGARCH model requires a set of parameter constraints maintaining positivity of variances and covariance stationarity of the modeled stochastic processes. The model likelihood is then highly irregular, skewed, asymmetric, and truncated in regions of relatively high posterior density, hindering the applicability and accuracy of asymptotic inference (Liu and Luger, 2009; Jondeau and Rockinger, 2006). Perhaps for this reason the number of application of CMGARCH has so far remained relatively limited. We contribute to this stream of literature by focusing on the practical aspects of computationally efficient model implementation.

The remainder of the paper is organized as follows. In Section 2 we lay out the CMGARCH model, including background on GARCH modeling and copula functions. In Section 3 we detail Constrained Hamiltonian Monte Carlo. In Section 4 we contrast CHMC with traditional RW-MCMC in the context of CMGARCH in an empirical application to a portfolio of S&P500 equity returns, quantifying the coefficients of tail dependence. Section 5 concludes.

2   Copula MGARCH

We will first lay out the background of dynamic GARCH modeling and copula functions, and then describe the copula multivariate GARCH model considered in this paper. Bayesian analysis of the model with CHMC will follow next.

2.1 GARCH Framework

Denote by y_{it} the marginal financial series for asset $i = 1, ..., N$ and time $t = 1, ..., T$ in an N-dimensional portfolio. Then y_{it} follows a GARCH(1,1) model if

$$y_{it} = \mu_i + \sqrt{h_{it}}\varepsilon_{it} \tag{1}$$

$$h_{it} = \omega_i + \alpha_i(y_{i,t-1} - \mu_i)^2 + \beta_i h_{i,t-1} \tag{2}$$

where h_{it} is the conditional variance of y_{it} given the information set $I_{i,t-1} = \{y_{i,t-1}, y_{i,t-2}, ...\}$, and ε_{it} are independent and identically distributed random variables with zero mean. The parameter constraints

$$\omega_i, \alpha_i, \beta_i > 0 \tag{3}$$

$$\alpha_i + \beta_i < 1 \tag{4}$$

are required to ensure positivity of h_{it} and covariance stationarity of y_{it}, respectively.

We assume that ε_{it} follow the Student t-distribution, $\varepsilon_{it} \sim t_{\nu_i}$, with ν_i degrees of freedom, zero mean and variance $\nu_i/(\nu_i - 2)$. The density of ε_{it}, $f(\varepsilon_{it})$ is thus given by

$$f(\varepsilon_{it}) = \frac{\Gamma\left((\nu_i + 1)/2\right)}{\Gamma\left(\nu_i/2\right)(\nu_i\pi)^{1/2}} \left(1 + \varepsilon_{it}^2/\nu_i\right)^{-(\nu_i+1)/2}$$

This distributional assumption is typical for modelling fat tails in univariate time series (Bollerslev, 1987). The conditional distribution function of each marginal series is then

$$F_i(y_{it}|\mu_i, h_{it}) = t_{\nu_i}\left((y_{it} - \mu_i)h_{it}^{-1/2}\right)$$

for $i = 1, ..., N$.

2.2 Copulas

A N-dimensional copula $C(u_1, ..., u_N)$ is a multivariate distribution function in the unit hypercube $[0, 1]^N$, with uniform $U(0, 1)$ marginal distributions. By the Sklar's Theorem, every joint distribution $F(x_1, ..., x_N)$, whose marginals are given by $F_1(x_1), ..., F_N(x_N)$, can

be written as

$$F(x_1, ..., x_N) = C\left(F_1(x_1), ..., F_N(x_N)\right). \tag{5}$$

Furthermore, if the marginal distributions are continuous, then there is a unique copula associated to the joint distribution, F, that can be obtained from

$$C(u_1, ..., u_N) = F\left(F_1^{-1}(x_1), ..., F_N^{-1}(x_N)\right). \tag{6}$$

Conversely, given an N-dimensional copula, $C(u_1, ..., u_N)$, and N univariate distributions, $F_1(x_1), ..., F_N(x_N)$, the function (5) is an N-variate distribution function with margins

$$F_1(x_1), ..., F_N(x_N)$$

whose corresponding density function is given by

$$f(x_1, ..., x_N) = c\left(F_1(x_1), ..., F_N(x_N)\right) \prod_{i=1}^{N} f_i(x_i) \tag{7}$$

where f_i denotes the marginal density functions and c is the copula density function derived from (6) as

$$c(u_1, ..., u_N) = \frac{f\left(F_1^{-1}(x_1), ..., F_N^{-1}(x_N)\right)}{\prod\limits_{i=1}^{N} f_i(F_i^{-1}(x_i))}.$$

The principal characteristic of a copula function is its ability to decompose the joint distribution into two parts: marginal distributions and dependence structure. Different dependence structures can combine the same marginal distributions into different joint distributions. Similarly, different marginal distributions under the same dependence structure can also lead to different joint distributions. Many parsimonious parametric families of copula functions have been proposed in the literature (Nelsen, 2006). A fundamental case is the Gaussian copula, obtained from the multivariate normal distribution with correlation matrix R, given by

$$C_R^G(u_1, ..., u_N) = \int_{-\infty}^{\Phi^{-1}(u_1)} \cdots \int_{-\infty}^{\Phi^{-1}(u_N)} (2\pi)^{-N/2} |R|^{-1/2} \exp\left(-\mathbf{u}R^{-1}\mathbf{u}/2\right) d\mathbf{u}$$

where $\mathbf{u} = (u_1, ..., u_N)'$ and Φ^{-1} is the inverse of the cumulative distribution function of

4

the univariate standard normal distribution. The Gaussian copula assumes that there is no dependence in the tails of the distribution. Tail dependence reflects the limiting proportion of one variable reaching extreme values given that the other related variable has reached an extreme as well. Formally, it is defined by Salvadori et al. (2007) as follows.

Definition 1 (Tail dependence (2-dimensional case)) *Let* $Z = (X, Y)$. *The random vector* Z *is upper tail dependent if*

$$\lambda_U = \lim_{t \to 1^-} \mathbb{P}\left\{ X > F_X^{-1}(t) | Y > F_Y^{-1}(t) \right\} > 0,$$

provided that the limit exist. If $\lambda_U = 0$, *the* Z *is upper tail independent.* λ_U *is called the upper tail dependence coefficient.*

Similarly, the lower tail dependence coefficient λ_L *is defined as*

$$\lambda_L = \lim_{t \to 0^+} \mathbb{P}\left\{ X \leq F_X^{-1}(t) | Y \leq F_Y^{-1}(t) \right\},$$

provided that the limit exist. If $\lambda_L = 0$, *the* Z *is lower tail independent, and is lower tail dependent if* $\lambda_L > 0$.

Such dependencies are of interest in financial economics since financial markets exhibit higher correlation in times of distress (Embrechts et al., 1997), hence using the Gaussian copula would not be appropriate since it assumes this tail dependency to be null. It is therefore often more useful to consider the t-copula, which is obtained from the multivariate t-distribution with η degrees of freedom and correlation matrix, R, given by

$$C_{\eta,R}^t(u_1, ..., u_N) = \int_{-\infty}^{t_\eta^{-1}(u_1)} \cdots \int_{-\infty}^{t_\eta^{-1}(u_N)} \frac{\Gamma\left((\eta + N)/2\right)}{\Gamma\left(\eta/2\right)} (\eta\pi)^{-N/2} |R|^{-1/2} \left(1 + \mathbf{u}R^{-1}\mathbf{u}/\eta\right)^{-(\eta+N)/2} d\mathbf{u}$$

(8)

where t_η^{-1} denotes the inverse of the cumulative distribution function of the standard univariate Student-t distribution with η degrees of freedom. The Gaussian copula results as a special case of the t-copula when η goes to infinity. As the t-copula is of an elliptically symmetric distribution, the two measures λ_U and λ_L coincide and can be denoted simply by λ. While the Gaussian copula is characterized by zero tail dependence, $\lambda = 0$, Embrechts et al. (2001) provide the following simple formula to represent the tail dependence coefficient

λ of the t-copula:

$$\lambda = 2t_{\eta+1}\left(\frac{-(\eta+1)^{1/2}(1-\rho)^{1/2}}{(1+\rho)^{1/2}}\right) \tag{9}$$

for a pair of random variables with correlation coefficient ρ.

2.3 Copula Multivariate GARCH Model

We consider the CMGARCH model setup of Ausín and Lopes (2010). The dependence structure between the marginal series is described by a time-varying t-copula[1] function with η degrees of freedom, as defined in (8), with density at each time t given by

$$c_{\eta,R_t}^t(u_{1t}, ..., u_{Nt}) = \frac{f_{\eta,R_t}^t\left(t_\eta^{-1}(u_{1t}), ..., t_\eta^{-1}(u_{Nt})\right)}{\displaystyle\prod_{i=1}^{N} f_\eta^t(t_\eta^{-1}(u_{it}))}$$

where $u_{it} = F_i(y_{it}|\mu_i, h_{it})$ for $i = 1, ..., N$, f_{η,R_t}^t is the joint density of the standard multivariate Student-t distribution with η degrees of freedom and correlation matrix R_t and f_η^t is the density of the standard univariate t-distribution with η degrees of freedom. The parameter matrix R_t of the t-copula varies through time according to

$$R_t = (1 - a - b)R + a\Psi_{t-1} + bR_{t-1} \tag{10}$$

where a and b are nonnegative parameters, R is a time-invariant $N \times N$ positive definite parameter matrix with unit diagonal elements and Ψ_{t-1} is an $N \times N$ matrix whose (i,j)-th element is given by

$$\Psi_{ij,t-1} = \frac{\sum_{h=1}^{m} x_{i,t-h}x_{j,t-h}}{\left(\sum_{h=1}^{m} x_{i,t-h}^2 \sum_{h=1}^{m} x_{j,t-h}^2\right)^{1/2}}$$

which gives the sample correlation of the copula-transformed error term ε_{it} from equation (1), $\{\mathbf{x}_{t-1}, ..., \mathbf{x}_{t-m}\}$, with $m \geq 2$, where

$$\mathbf{x}_t = (x_{1t}, ..., x_{Nt})$$
$$= (t_\eta^{-1}(t_{\nu_1}(\varepsilon_{1t})), ..., t_\eta^{-1}(t_{\nu_N}(\varepsilon_{Nt}))).$$

---

[1]The superscript t indicates that the distribution is a Student-t distribution, while the subscript t corresponds to the time index.

The dynamics in (10) is based on the correlation matrix evolution proposed by Tse and Tsui (2002) in a multivariate GARCH model. The advantage of this specification is that the parameter matrix R_t is a well-defined positive definite correlation matrix with unit diagonal elements, without the need to use any transformation such as the logistic function considered e.g. in Patton (2006) and Dias and Embrechts (2004).

In order to maintain stationarity, the model is constrained with

$$0 \le a, b \le 1 \tag{11}$$

$$a + b \le 1 \tag{12}$$

$$-1 \le r_{ij} \le 1 \tag{13}$$

where r_{ij} is the (i,j)-th element of the parameter matrix R.

Using (7), the log-likelihood function of the CMGARCH model is given by

$$
\begin{aligned}
\ln L(\varphi|\mathbf{Y}) =& \sum_{t=1}^{T} \ln f(\mathbf{y}_t|\boldsymbol{\mu}, \mathbf{h}_t) \\
=& \sum_{t=1}^{T} \left(\ln \Gamma\left((\eta + N)/2\right) + (N-1)\ln\Gamma\left(\eta/2\right) - N\ln\Gamma((\eta+1)/2) \right) \\
& -\frac{1}{2}\sum_{t=1}^{T}\left((\eta+N)\ln\left(1 + \mathbf{x}_t R_t^{-1}\mathbf{x}_t/\eta\right) - \ln(|R_t|) \right) + (\eta+1)\sum_{t=1}^{T}\sum_{i=1}^{N}\ln\left(1 + x_{it}^2/\eta\right) \\
& +\frac{1}{2}\sum_{t=1}^{T}\sum_{i=1}^{N}\left(2\ln\Gamma\left((\nu_i+1)/2\right) - 2\ln\Gamma\left(\nu_i/2\right) - \ln(\pi\nu_i h_{it}) \right) \\
& -\frac{1}{2}\sum_{t=1}^{T}\sum_{i=1}^{N}(\nu_i+1)\ln(1 + (y_{it}-\mu_i)^2/(\nu_i h_{it}))
\end{aligned}
$$

where $\varphi = \{\{\mu_i, \omega_i, \alpha_i, \beta_i, \nu_i\}_{i=1}^{N}, a, b, R, \eta\}$ and $\mathbf{x}_t = (x_{1t}, ..., x_{Nt})$.

Figure 1 shows the model likelihood for (α_1, β_1) with the constraint (4), $\alpha_1 + \beta_1 < 1$, evidenced by the lack of likelihood contours in the upper right hand section of the graph. Such constraint would result in redundancies for traditional MCMC methods as any proposal falling into the region where the constraint is violated would be rejected. In contrast, CHMC proposals are reflected off the constraint barrier and the proposal is made every time in the

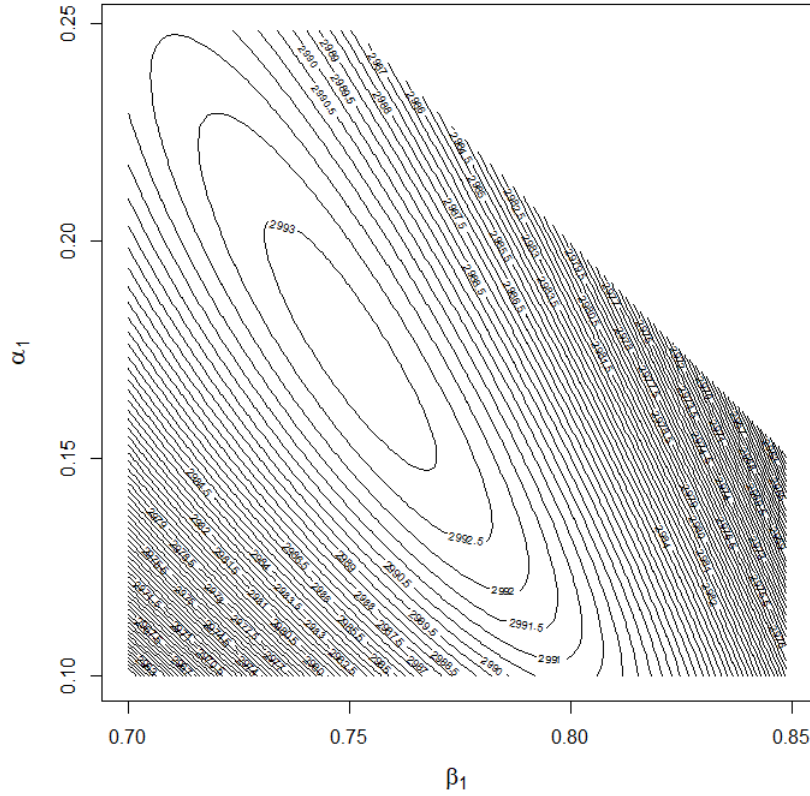region where the constraint is not violated, as detailed below.

Figure 1: Constrained likelihood

As in Ausín and Lopes (2010), for the parameters $\mu_i, \omega_i, \alpha_i, \beta_i$ we assume a uniform prior over their respective domains of stationarity given by the constraints (3)-(4). For the copula parameters a, b, R we also assume a uniform prior distribution restricted to satisfy the constraints (11)-(13). For the degrees of freedom of the t-distribution ν_i we assume a half-right side Cauchy prior

$$f(\nu_i) \sim (1 + \nu_i^2)^{-1}$$

for $\nu_i > 0$, and we assume the same prior for the degrees of freedom parameter η. All parameters are sampled jointly in each iteration.

3 Constrained Hamiltonian Monte Carlo

The purpose of traditional Markov Chain Monte Carlo (MCMC) methods is to formulate a Markov chain on the parameter space Θ for which, under certain conditions, $\pi(\theta) \in \mathcal{P}_\theta$

is the invariant (also called 'equilibrium' or 'long-run') distribution. The Markov chain of draws of θ can be used to construct simulation-based estimates of the required integrals, and functionals $\mathcal{Q}(\theta)$ of θ that are expressed as integrals. These functionals include objects of interest for inference on θ such as quantiles of $\pi(\theta)$.

In a generic MCMC scheme, the Markov chain sampling mechanism specifies a method for generating a sequence of random variables $\{\theta_k\}_{k=1}^K$, starting from an initial point θ_0, in the form of conditional distributions for the draws $\theta_{k+1}|\theta_k \sim G(\theta_k)$. Under relatively weak regularity conditions (Robert and Casella, 2004), the average of the Markov chain converges to the expectation under the stationary distribution:

$$\lim_{K \to \infty} \frac{1}{K} \sum_{k=1}^K \mathcal{Q}(\theta_k) = E_\pi[\mathcal{Q}(\theta)]$$

A Markov chain with this property is called ergodic. As a means of approximation the analyst relies on large but finite number of draws $K \in \mathbb{N}$ which the analyst selects in applications.

$G(\theta_k)$ can be obtained from a given model and its corresponding likelihood $\pi(\theta)$. Typically, $\pi(\theta)$ has a complicated form which precludes direct sampling in which case the Metropolis-Hastings (M-H) principle is usually used for $\theta_{k+1}|\theta_k$ from $G(\theta_k)$; see Chib and Greenberg (1995) for a detailed overview. The M-H algorithm proceeds according to the following steps. Suppose we have a proposal-generating density $q(\theta_{k+1}^*|\theta_k)$ where θ_{k+1}^* is a proposed state given the current state θ_k of the Markov chain. The M-H principle stipulates that θ_{k+1}^* be accepted as the next state θ_{k+1} with the acceptance probability

$$\alpha(\theta_k, \theta_{k+1}^*) = \min\left[\frac{\pi(\theta_{k+1}^*)q(\theta_k|\theta_{k+1}^*)}{\pi(\theta_k)q(\theta_{k+1}^*|\theta_k)}, 1\right] \tag{14}$$

otherwise $\theta_{k+1} = \theta_k$. Then, the Markov chain satisfies the so-called detailed balance condition

$$\pi(\theta_k)q(\theta_{k+1}^*|\theta_k)\alpha(\theta_k, \theta_{k+1}^*) = \pi(\theta_{k+1}^*)q(\theta_k|\theta_{k+1}^*)\alpha(\theta_{k+1}^*, \theta_k)$$

which is sufficient for ergodicity. $\alpha(\theta_{k+1}^*, \theta_k)$ is the probability of the move $\theta_k|\theta_{k+1}^*$ if the dynamics of the proposal generating mechanism are reversed. The proposal-generating density $q(\theta_{k+1}^*|\theta_k)$ can be chosen to be sampled conveniently even though it may be difficult or expensive to sample from $\pi(\theta)$. The popular Gibbs sampler arises as a special case when

the M-H sampler is factored into conditional densities. The proposal draws $\theta^*_{k+1}|\theta_k$ from $q(\theta^*_{k+1}|\theta_k)$ in (14) are generated in one step. We will refer to this procedure as random walk MCMC (RW-MCMC).

3.1 CHMC Principle

Hamiltonian (or Hybrid) Monte Carlo (HMC) originates in the physics literature where it was introduced as a fast method for molecular dynamics simulations (Duane et al., 1987). It has since become an important tool for Bayesian statistical inference (Neal, 1993, 2011; Ishwaran, 1999; Liu, 2004; Beskos et al., 2013) with applications in a number of diverse fields including statistical physics (Akhmatskaya et al., 2009; Gupta et al., 1988) and computational chemistry (Tuckerman et al., 1993). Detailed description of the CHMC mechanism from the statistics perspective is provided in Neal (2011). Nonetheless, its usage in economics and finance has so far been relatively scarce. HMC and its constrained version, CHMC, applies to a general class of models that is parametrized by a Euclidean vector $\theta \in \Theta$ for which all information in the sample is contained in the model likelihood $\pi(\theta)$ assumed known up to an integrating constant. Formally, this class of models is characterized by a family $\mathcal{P}_\theta$ of probability measures on a measurable space $(\Theta, \mathcal{B})$ where $\mathcal{B}$ is the Borel σ-algebra.

In contrast to a traditional MCMC scheme utilizing one-step proposal draws, Hamiltonian Monte Carlo (HMC) uses a sequence of many proposal steps whereby the last step in the sequence is designated as the proposal draw. The proposal is thus drawn in a relatively distant location from the current draw, which facilitates efficient exploration of the parameter space with the resulting Markov chain. The proposal sequence is constructed using difference equations of the law of motion that result in high acceptance probability even for distant proposals. The parameter space Θ is augmented with a set of independent auxiliary stochastic parameters $\gamma \in \Gamma$ that fulfill a supplementary role in the proposal algorithm, facilitating the directional guidance of the proposal mechanism. The proposal sequence takes the form $\{\theta^\ell_k, \gamma^\ell_k\}^L_{\ell=1}$ starting from the current state $(\theta_k, \gamma_k) = (\theta^0_k, \gamma^0_k)$ and yielding a proposal $(\theta_k, \gamma_k) = (\theta^L_k, \gamma^L_k)$. The detailed balance is then satisfied using the acceptance probability

$$\alpha(\theta_k, \gamma_k; \theta^*_{k+1}, \gamma^*_{k+1}) = \min\left[\frac{\pi(\theta^*_{k+1}, \gamma^*_{k+1})q(\theta_k, \gamma_k|\theta^*_{k+1}, \gamma^*_{k+1})}{\pi(\theta_k, \gamma_k)q(\theta^*_{k+1}, \gamma^*_{k+1}|\theta_k, \gamma_k)}, 1\right] \qquad (15)$$

In constrained HMC (CHMC), constraints are embedded in the HMC proposal mechanism via "hard walls" representing a barrier against which the proposal sequence, simulating a particle movement, bounces off elastically. Thus, violating a constraint along the proposal sequence does not result in proposal rejection but rather an altered course of the proposal path, eliminating any associated redundancies. Heuristically, the constraint is checked at each step of the proposal sequence and if it is violated then the trajectory of the sequence is reflected off the hard wall posed by the constraint. This facilitates efficient exploration of the parameter space even in the presence of highly complex parameter constraints.

Let $\gamma \sim \Phi(\gamma; 0, M)$ where Φ denotes the Gaussian distribution with mean vector 0 and covariance matrix M. Denote the constraint on the parameter space by a density kernel $\exp\left(C_k(\theta)\right)$ where

$$
C_k(\theta) = \begin{cases} 0 \text{ if the constraint is satisfied} \\ C_k > 0 \text{ s.t. } \lim_{k \to \infty} C_k = \infty \text{ if the constraint is violated} \end{cases}
$$

Denote the joint density of (θ, γ) by $\pi(\theta, \gamma)$. The negative of the logarithm of the joint density of (θ, γ) is given by the constrained Hamiltonian equation

$$
H(\theta, \gamma) = -C_k(\theta) - \ln \pi(\theta) + \frac{1}{2} \ln \left((2\pi)^d |M|\right) + \frac{1}{2} \gamma' M^{-1} \gamma \tag{16}
$$

where $d = \dim(\theta) = \dim(\gamma)$. In the physics literature, the Hamiltonian $H(\theta, \gamma)$ represents the total energy of the (θ, γ)-system where θ denotes the position (or state) variable and $-\ln \pi(\theta)$ describes its potential energy. γ is the momentum variable with kinetic energy $\gamma' M^{-1} \gamma / 2$ where M is a constant, symmetric, positive-definite "mass" matrix, typically set to the identity matrix.

A canonical iteration of Constrained Hamiltonian Monte Carlo (CHMC) is performed in the following steps:

1. Draw an initial auxiliary parameter vector $\gamma_k^0 \sim \Phi(\gamma; 0, M)$;

2. Transition from (θ_k, γ_k) to $(\theta_k^L, \gamma_k^L) = (\theta_{k+1}^*, \gamma_{k+1}^*)$ according to the constrained Hamiltonian dynamics;

3. Accept $(\theta^*_{k+1}, \gamma^*_{k+1})$ with probability

$$\alpha(\theta_k, \gamma_k; \theta^*_{k+1}, \gamma^*_{k+1}) = \min \left[\exp \left(-H(\theta^*_{k+1}, \gamma^*_{k+1}) + H(\theta^0_k, \gamma^0_k) \right), 1 \right], \qquad (17)$$

otherwise keep (θ_k, γ_k) as the next MC draw.

The constraints on the parameter space Θ are reflected in Step 2. We will now describe each step in detail.

Step 1 updates the system akin to a RW draw in order to make the resulting Markov chain $\{(\theta_k, \gamma_k)\}^R_{k=1}$ irreducible and aperiodic (Ishwaran, 1999).

Step 2 constructs a sequence $\{\theta^\ell_k, \gamma^\ell_k\}^L_{\ell=1}$ according to the Hamiltonian dynamics starting from the current state (θ^0_k, γ^0_k), with the newly drawn γ^0_k, and setting the last member of the sequence as the CHMC new state proposal $(\theta^*_{k+1}, \gamma^*_{k+1}) = (\theta^L_k, \gamma^L_k)$. The transition from (θ^0_k, γ^0_k) to (θ^L_k, γ^L_k) via the proposal sequence $\{\theta^\ell_k, \gamma^\ell_k\}^L_{\ell=1}$ taken according to the discretized Hamiltonian dynamics described below is fully deterministic, placing a Dirac delta probability mass $\delta(\theta^\ell_k, \gamma^\ell_k) = 1$ on each $(\theta^\ell_k, \gamma^\ell_k)$ conditional on (θ^0_k, γ^0_k). The CHMC acceptance probability in (17) is specified in terms of the difference between the Hamiltonian (16) evaluated at the initial (θ^0_k, γ^0_k) and at the proposal $(\theta^*_{k+1}, \gamma^*_{k+1})$. The role of the Hamiltonian dynamics is to ensure that the acceptance probability (17) for $(\theta^*_{k+1}, \gamma^*_{k+1})$ is kept close to 1. This is achieved by maintaining the difference $-H(\theta^*_{k+1}, \gamma^*_{k+1}) + H(\theta^0_k, \gamma^0_k)$ close to zero throughout the sequence $\{\theta^\ell_k, \gamma^\ell_k\}^L_{\ell=1}$. This property can be achieved by conceptualizing θ and γ as functions of continuous time t and specifying their evolution using the Hamiltonian dynamics equations

$$\frac{d\theta_i}{dt} = \frac{\partial H(\theta, \gamma)}{\partial \gamma_i} = \left[M^{-1} \gamma \right]_i \qquad (18)$$

$$\frac{d\gamma_i}{dt} = -\frac{\partial H(\theta, \gamma)}{\partial \theta_i} = \nabla_{\theta_i} \ln \pi(\theta) \qquad (19)$$

for $i = 1, \ldots, d$. For any discrete time interval of duration s, (18)–(19) define a mapping T_s from the state of the system at time t to the state at time $t+s$. The differential equations (18)–(19) are generally solved by numerical methods, typically the Stormer-Verlet (or leapfrog) numerical integrator (Leimkuhler and Reich, 2004). For each step of $\{\theta^\ell_k, \gamma^\ell_k\}^L_{\ell=1}$, CHMC discretizes the Hamiltonian dynamics (18)–(19) as follows: for some small $\varepsilon \in \mathbb{R}$ first take a

half-ε step in γ

$$\gamma(t + \varepsilon/2) = \gamma(t) + (\varepsilon/2)\nabla_\theta \ln \pi(\theta(t)), \tag{20}$$

then take a full ε step in θ

$$\theta(t + \varepsilon) = \theta(t) + \varepsilon M^{-1}\gamma(t + \varepsilon/2), \tag{21}$$

and check the constraint at $\theta(t + \varepsilon)$ for each dimension i of θ. If the constraint is violated for any i then set $\gamma_i(t + \varepsilon/2) = -\gamma_i(t + \varepsilon/2)$ reversing the proposal dynamics and take further steps in θ_i until the constraint is satisfied, then finish with another half-ε step in γ

$$\gamma(t + \varepsilon) = \gamma(t + \varepsilon/2) + (\varepsilon/2)\nabla_\theta \ln \pi(\theta(t + \varepsilon)). \tag{22}$$

Intuitively, the proposal trajectory bounces off the "walls" given by the constraint. Since a move in γ only occurs conditional on θ for which the constraint is satisfied in which case $C_k(\theta) = 0$, and the move in θ only depends on $\partial H(\theta, \gamma)/\partial \gamma_i$ which does not involve $C_k(\theta)$, the exact functional form of $C_k(\theta)$ is inconsequential as long as it is differentiable in θ to define valid Hamiltonian dynamics.

Even though $\ln \pi(\theta_k^\ell)$ can deviate substantially from $\ln \pi(\theta_k^0)$, the additional Hamiltonian terms parametrized by γ in (16) compensate for this deviation maintaining the overall level of $H(\theta_k^\ell, \gamma_k^\ell)$ close to constant over $\ell = 1, \ldots, L$ when used in accordance with (20)–(22), since $\frac{\partial H(\theta, \gamma)}{\partial \gamma_i}$ and $\frac{\partial H(\theta, \gamma)}{\partial \theta_i}$ enter with the opposite signs in (18)–(19). In contrast, without the additional parametrization with γ, if only $\ln \pi(\theta_k^\ell)$ were to be used in the proposal mechanism as is the case in RW style samplers, the M-H acceptance probability would often drop to zero relatively quickly.

Step 3 applies a Metropolis correction to the proposal $(\theta_{k+1}^*, \gamma_{k+1}^*)$. In continuous time, or for $\varepsilon \to 0$, (18)–(19) would keep $-H(\theta_{k+1}^*, \gamma_{k+1}^*) + H(\theta_k, \gamma_k) = 0$ exactly resulting in $\alpha(\theta_k, \theta_{k+1}^*) = 1$ but for discrete $\varepsilon > 0$, in general, $-H(\theta^*, \gamma^*) + H(\theta, \gamma) \neq 0$ necessitating the Metropolis step.

The system (20)–(22) is time reversible and symmetric in (θ, γ), which implies that the forward and reverse transition probabilities $q(\theta_k^L, \gamma_k^L|\theta_k^0, \gamma_k^0)$ and $q(\theta_k^0, \gamma_k^0|\theta_k^L, \gamma_k^L)$ are equal: this simplifies the Metropolis-Hastings acceptance ratio in (15) to the Metropolis form $\pi(\theta_{k+1}^*, \gamma_{k+1}^*)/\pi(\theta_k^0, \gamma_k^0)$. From the definition of the Hamiltonian $H(\theta, \gamma)$ in (16) as the nega-

tive of the log-joint densities, the joint density of (θ, π) is given by

$$\pi(\theta, \gamma) = \exp\left[-H(\theta, \gamma)\right] = \pi(\theta)\left((2\pi)^d |M|\right)^{-1/2} \exp\left(-\frac{1}{2}\gamma' M^{-1}\gamma\right) \tag{23}$$

Hence, the Metropolis acceptance probability takes the form

$$\alpha(\theta_r, \gamma_r; \theta^*_{r+1}, \gamma^*_{r+1}) = \min\left[\frac{\pi(\theta^*_{r+1}, \gamma^*_{r+1})}{\pi(\theta^0_r, \gamma^0_r)}, 1\right]$$
$$= \min\left[\exp\left(-H(\theta^*_{r+1}, \gamma^*_{r+1}) + H(\theta^0_r, \gamma^0_r)\right), 1\right]$$

The expression for $\alpha(\theta_k, \gamma_k; \theta^*_{k+1}, \gamma^*_{k+1})$ shows, as noted above, that the CHMC acceptance probability is given in terms of the difference of the Hamiltonian equations $H(\theta^0_k, \gamma^0_k) - H(\theta^*_{k+1}, \gamma^*_{k+1})$. As this Hamiltonian difference approaches zero, the acceptance probability approaches one. A key feature of the Hamiltonian dynamics (18)–(19) in Step 2 is that they maintain $H(\theta, \gamma)$ constant over the parameter space in continuous time conditional on $H(\theta^0_k, \gamma^0_k)$ obtained in Step 1, while their discretization (20)–(22) closely approximates this property for discrete time steps $\varepsilon > 0$ with a global error of order ε^2 corrected by the Metropolis update in Step 3.

In the next Section we apply CHMC to the CMGARCH model of Section 2.

4 Application

We apply CHMC and RW-MCMC to the daily closing prices of S&P500 equities for the period 02/Jan/2015 to 29/Jun/2018, with $T = 880$, obtained from the Compustat database. We chose two companies with the highest overall transaction volumes: Citigroup (C) and Bank of America (BAC). The log return series are plotted in Figure 2 and summary statistics are presented in Table 1.

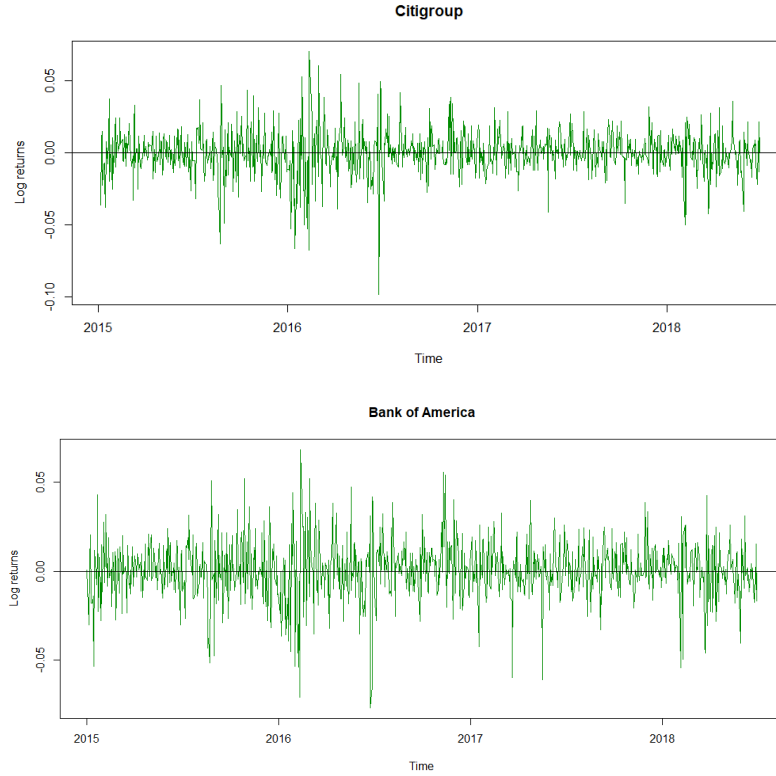|	Mean	St.Dev.	Skewness	Kurtosis
C	0.241×10^{-3}	0.015	-0.39	6.90
BAC	0.516×10^{-3}	0.016	-0.29	5.32

Table 1: Summary Statistics

Figure 2: Daily equity log returns

In theory, the dimensionality of the portfolio is only limited by the sample size. Nonetheless, in practice the portfolio size for feasible implementation will also be restricted by the computing power available. Our implementation was run with a Coarray Fortran 2008 code using Intel 2016 compiler on 3.1 GHz desktop computer[2]. We set $m = 2$ and $\nu_i = 2$ for $i = 1, ..., N$. In RW-MCMC we tuned the step sizes for each parameter to achieve transition acceptance rate of about 30% (Roberts et al., 1997). In CHMC tuned the step sizes to achieve transition acceptance rate of about 60%. Theoretical analysis of optimal step sizes and acceptance rates for HMC is provided in Beskos et al. (2013).

We used $L = 20$ as the length of the proposal sequence in CHMC. Even though each iteration takes longer to complete for CHMC than for RW-MCMC, the resulting parameter draws exhibit superior mixing properties for the former, even when normalized per unit of time. In order to make both methods comparable from the user's perspective, we ran both procedures for 10 minutes of wallclock time and discarded one extra initial minute of runtime

---

[2]Intel Xeon E5-2687w v3, Windows 10 operating system

as a burn-in part of the chain. CHMC completed just over 1,100 iterations while RW-MCMC completed just over 31,000 iterations. In order to minimize autocorrelation of the chain and render the output comparable to CHMC in terms of size per unit of time, we thinned the RW-MCMC chain by retaining every 30-th iteration. Table 2 reports comparison of the resulting chains in terms of effective sample size (ESS), evaluated by the R package coda (Plummer et al., 2006). ESS of a Bayesian Markov chain parameter draw is the number of effectively independent draws from the posterior distribution to which the Markov chain draw is equivalent Kass et al. (1998). The ESS has been commonly used as a standard measure for the quality of the posterior approximation for individual chains.

parameter	CHMC	RW-MCMC
η	91.69	8.74
a	43.94	1.77
b	31.54	2.63
μ_1	52.43	4.23
ω_1	109.74	10.32
α_1	32.88	26.29
β_1	60.20	5.02
μ_2	48.01	3.21
ω_2	85.23	8.90
α_2	16.83	7.75
β_2	51.33	2.79
$r_{1,2}$	66.99	2.26

Table 2: Normalized Effective Sample Size

Markov chain convergence diagnostics, obtained using the R package coda, for both methods are reported in Table 3. Specifically, we evaluated the z-scores of the Geweke (1992) convergence test standardized z-score and p-values of the Heidelberger and Welch (1983) stationary distribution test. As evidenced by the z-scores, while for CHMC all chains have converged within the burn-in section, for RW-MCMC half of the parameter chains have failed to reach convergence. The p-values output confirms that both methods drew chains from a stationary distribution.

We further visually illustrate the CHMC versus RW-MCMC comparison in Figures 3-5 in the Appendix, showing the trace plots of the Markov chains. The CHMC trace (left column) mixes very well, exploring the tails of the posterior, while the RW-MCMC trace (right column) suffers from relatively poorer mixing with scant tail exploration. This is consistent with the outcomes of the convergence diagnostics reported above. Overall, our results favor

the use of the CHMC as an effective method of inference for the complex posterior of the CMGARCH model.

parameter	CHMC		RW-MCMC	
	z-score	HW p-value	z-score	HW p-value
η	0.342	0.848	-1.323	0.850
a	0.345	0.263	-13.412	0.265
b	-0.472	0.587	-16.012	0.169
μ_1	1.580	0.007	-3.936	0.057
ω_1	1.647	0.671	-0.765	0.133
α_1	1.005	0.844	1.304	0.053
β_1	-1.630	0.858	-1.455	0.506
μ_2	1.433	0.008	-3.122	0.447
ω_2	-1.213	0.276	1.613	0.022
α_2	-1.111	0.502	3.021	0.075
β_2	1.038	0.283	-1.543	0.189
$r_{1,2}$	0.144	0.555	-2.199	0.597

Table 3: Geweke z-score and Heidelberger and Welch (HW) diagnostics p-value

We used the CHMC output to evaluate the posterior mean value of the tail dependence index (9) as

$$\overline{\widehat{\lambda}} = \frac{1}{KT} \sum_{k=1}^{K} \sum_{t=1}^{T} 2t_{\eta^{(k)}+1} \left(\frac{-(\eta^{(k)}+1)^{1/2}(1-\rho_t^{(k)})^{1/2}}{(1+\rho_t^{(k)})^{1/2}} \right)$$

where the superscript (k) indicated the k-th draw. We then used the posterior distribution of the tail dependence index to quantify its 95% Bayesian Credible Set (BCS). The posterior mean was thus 0.176, and the 95% BCS (0.157, 0.194).

5   Conclusions

In this paper we build on a recent stream of literature merging the dynamic MGARCH modeling framework with the copula dependence structure, yielding a family of flexible yet parsimonious non-elliptical multivariate distributions for modeling the features of financial portfolio series. The main advantage of this approach is that the individual marginal densities of the series can be defined separately from their dependence structure. However, the resulting CMGARCH model requires a set of parameter constraints maintaining positivity

of variances and covariance stationarity of the modeled stochastic processes. The model likelihood is then highly irregular, skewed, asymmetric, and truncated in regions of relatively high posterior density, hindering the applicability and accuracy of asymptotic inference. In this paper, we show that the Constrained Hamiltonian Monte Carlo exhibits a number of advantages over traditional random-walk based Markov Chain Monte Carlo methods utilized for the CMGARCH in the previous literature. Our analysis is performed in the context of an application to modelling a portfolio of S&P500 equity returns, quantifying the coefficient of tail dependence.

References

Akhmatskaya, E., N. Bou-Rabee, and S. Reich (2009). A comparison of generalized hybrid monte carlo methods with and without momentum flip. *Journal of Computational Physics 228*(6), 2256–2265.

Ausín, C. and H. Lopes (2010, 11). Time-varying joint distribution through copulas. *Computational Statistics & Data Analysis 54*, 2383–2399.

Bauwens, L., S. Laurent, and J. Rombouts (2006). Multivariate garch models: a survey. *Journal of Applied Econometrics 21*(1), 79–109.

Beskos, A., N. Pillai, G. Roberts, J.-M. Sanz-Serna, and A. Stuart (2013, 11). Optimal tuning of the hybrid monte carlo algorithm. *Bernoulli 19*(5A), 1501–1534.

Bollerslev, T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. *The Review of Economics and Statistics 69*(3), 542–47.

Chib, S. and E. Greenberg (1995). Understanding the metropolis-hastings algorithm. *American Statistician 49*(4), 327–335.

Dias, A. and P. Embrechts (2004, 12). Dynamic copula models for multivariate high-frequency data in finance. Discussion Paper.

Duane, S., A. Kennedy, B. Pendleton, and D. Roweth (1987). Hybrid monte carlo. *Physics Letters B 195*(2), 216–222.

Embrechts, P., C. Klüppelberg, and T. Mikosch (1997). *Modelling Extremal Events for Insurance and Finance*. Springer-Verlag.

Embrechts, P., F. Lindskog, and E. J. Mcneil (2001, 10). Modelling dependence with copulas and applications to risk management. *Handbook of Heavy Tailed Distributions in Finance 8*.

Embrechts, P., A. McNeil, and D. Straumann (2001). Correlation and dependency in risk management: properties and pitfalls. In M. Dempster and H. Moffatt (Eds.), *Risk Management: Value at Risk and Beyond*, pp. 176–223. Cambridge University Press.

Engle, R. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica 50*(4), 987–1007.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In J. Bernado, J. O. Berger, A. Dawid, and A. Smith (Eds.), *Bayesian Statistics 4*. Clarendon Press, Oxford, UK.

Gupta, R., G. Kilcup, and S. Sharpe (1988). Tuning the hybrid monte carlo algorithm. *Physical Review D 38*(4), 1278–1287.

Heidelberger, P. and P. Welch (1983). Simulation run length control in the presence of an initial transient. *Operations Research 31*, 1109–44.

Hu, L. (2006). Dependence patterns across financial markets: a mixed copula approach. *Applied Financial Economics 16*(10), 717–729.

Ishwaran, H. (1999). Applications of hybrid monte carlo to generalized linear models: Quasi-complete separation and neural networks. *Journal of Computational and Graphical Statistics 8*, 779–799.

Jondeau, E. and M. Rockinger (2006). The copula-garch model of conditional dependencies: An international stock market application. *Journal of International Money and Finance 25*(5), 827–853.

Kass, R. E., B. P. Carlin, A. Gelman, and R. M. Neal (1998). Markov chain monte carlo in practice: A roundtable discussion. *The American Statistician 52*(2), 93–100.

Lee, T.-H. and X. Long (2009, June). Copula-based multivariate GARCH model with uncorrelated dependent errors. *Journal of Econometrics 150*(2), 207–218.

Leimkuhler, B. and S. Reich (2004). *Simulating Hamiltonian Dynamics*. Cambridge University Press.

Liu, J. S. (2004). *Monte Carlo Strategies in Scientific Computing*. Springer Series in Statistics.

Liu, Y. and R. Luger (2009, 04). Efficient estimation of copula-garch models. *Computational Statistics & Data Analysis 53*, 2284–2297.

Neal, R. M. (1993). Probabilistic inference using markov chain monte carlo methods. Technical report crg-tr-93-1, Dept. of Computer Science, University of Toronto.

Neal, R. M. (2011). Mcmc using hamiltonian dynamics. In S. Brooks, A. Gelman, G. Jones, and X.-L. Meng (Eds.), *Handbook of Markov Chain Monte Carlo*. Chapman & Hall / CRC Press.

Nelsen, R. B. (1999). *An Introduction to Copulas*, Volume 139 of *Lecture Notes in Statistics*. New York, NY: Springer-Verlag.

Nelsen, R. B. (2006). *An Introduction to Copulas* (Second ed.). Springer Series in Statistics. New York, NY: Springer.

Patton, A. (2006). Estimation of multivariate models for time series of possibly different lengths. *Journal of Applied Econometrics 21*(2), 147–173.

Plummer, M., N. Best, K. Cowles, and K. Vines (2006). CODA: Convergence diagnosis and output analysis for MCMC. *R News 6*(1), 7–11.

Robert, C. P. and G. Casella (2004). *Monte Carlo statistical methods* (Second ed.). New York: Springer.

Roberts, O. G., A. Gelman, and R. W. Gilks (1997, 04). Weak convergence and optimal scaling of random walk metropolis algorithms. *The Annals of Applied Probability 7*.

Rodríguez, J. (2007). Measuring financial contagion: A copula approach. *Journal of Empirical Finance 14*(3), 401–423.

Salvadori, G., C. De Michele, N. T. Kottegoda, and R. Rosso (2007). *Extremes in Nature: An approach using Copulas.* Springer.

Tse, Y. and A. Tsui (2002). A multivariate generalized autoregressive conditional heteroscedasticity model with time-varying correlations. *Journal of Business and Economics Statistics 20*, 351–362.

Tuckerman, M., B. Berne, G. Martyna, and M. Klein (1993). Efficient molecular dynamics and hybrid monte carlo algorithms for path integrals. *The Journal of Chemical Physics 99*(4), 2796–2808.
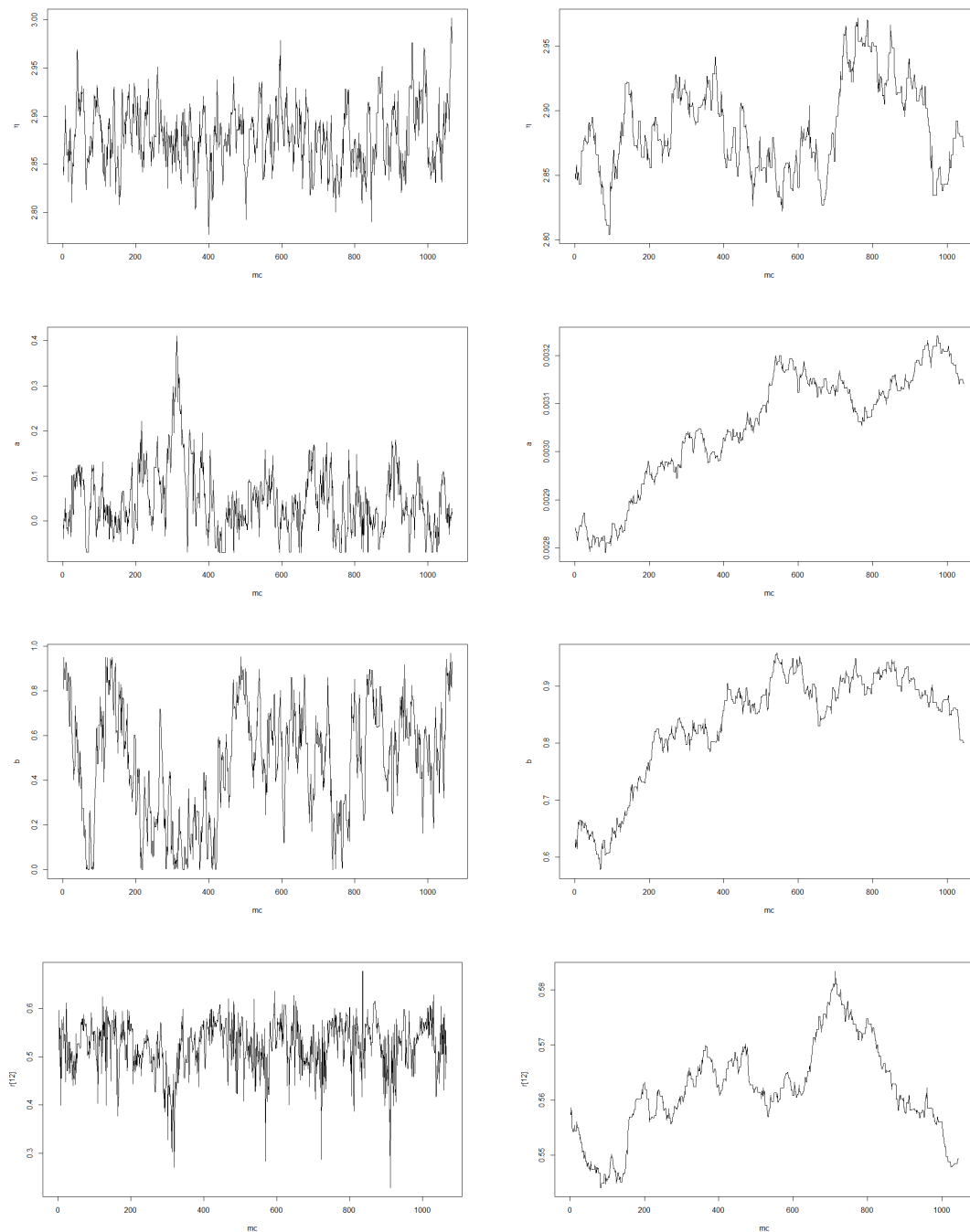
6 Appendix

Figure 3: Chains of parameter draws for η, a, b, and $r_{1,2}$; CHMC (left) and RW-MCMC (right)
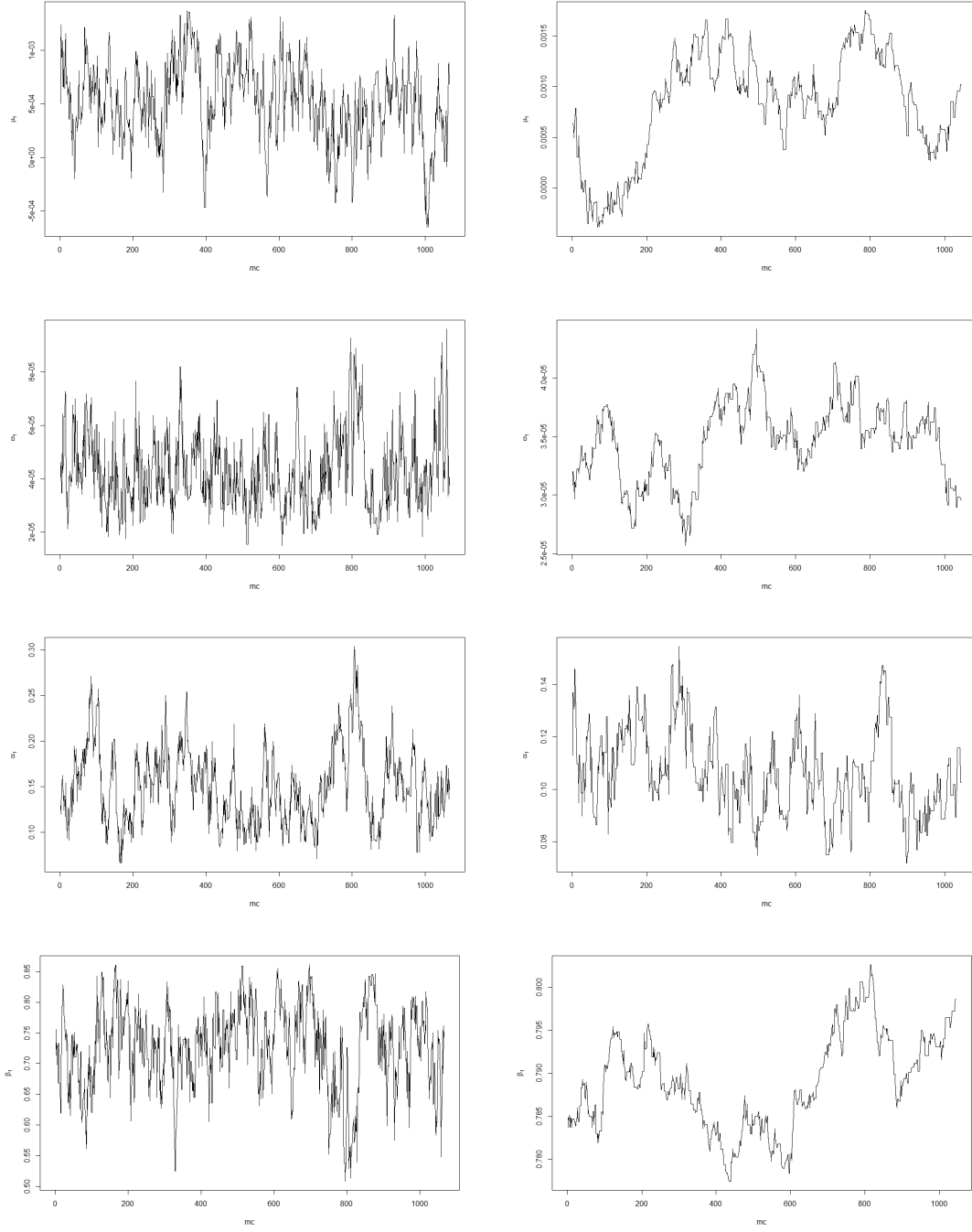
Figure 4: Chains of parameter draws for μ_1, ω_1, α_1, and β_1; CHMC (left) and RW-MCMC (right)

22
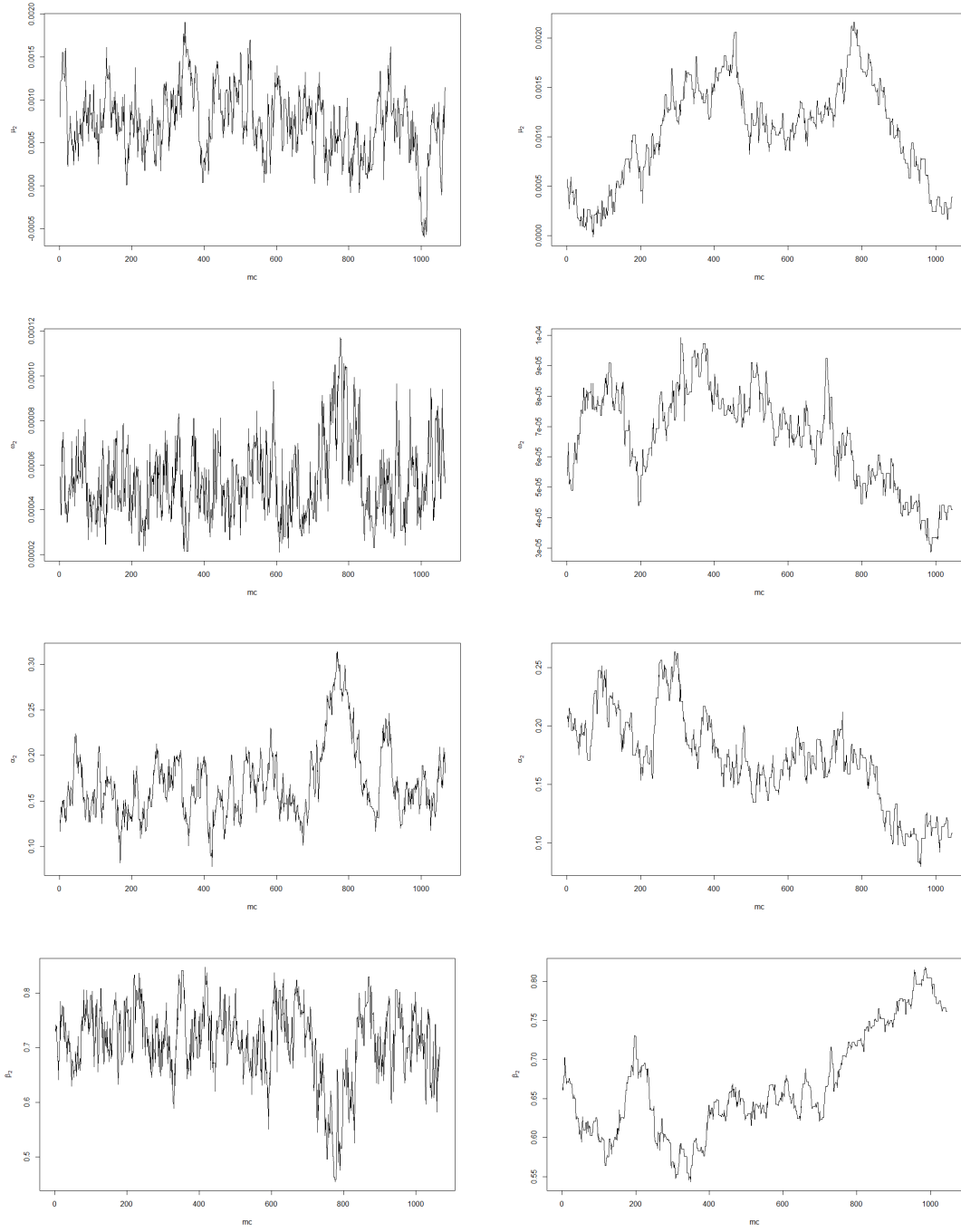
Figure 5: Chains of parameter draws for μ_2, ω_2, α_2, and β_2; CHMC (left) and RW-MCMC (right)