

Constrained Hamiltonian Monte Carlo in BEKK GARCH with Targeting*

Martin Burda[†]

May 8, 2013

Abstract

The GARCH class of models for dynamic conditional covariances trades off flexibility with parameter parsimony. The unrestricted BEKK GARCH dominates its restricted scalar and diagonal versions in terms of model fit but its parameter dimensionality increases quickly with the number of variables. Covariance targeting has been proposed as a way of reducing parameter dimensionality but for the BEKK with targeting the imposition of positive definiteness on the conditional covariance matrices presents a significant challenge. In this paper we suggest an approach based on Constrained Hamiltonian Monte Carlo that can deal effectively both with the nonlinear constraints resulting from BEKK targeting and the complicated nature of the BEKK likelihood in relatively high dimensions. We perform a model comparison of the full BEKK and the BEKK with targeting, indicating that the latter dominates the former in terms of marginal likelihood. Thus, we show that the BEKK with targeting presents an effective way of reducing parameter dimensionality without compromising the model fit, unlike the scalar or diagonal BEKK. The model comparison is conducted in the context of an application concerning a multivariate dynamic volatility analysis of a foreign exchange rate returns portfolio.

JEL: C11, C15, C32, C63

Keywords: *Dynamic conditional covariances, Markov chain Monte Carlo*

*I would like to thank John Maheu for valuable comments. This work was made possible by the facilities of the Shared Hierarchical Academic Research Computing Network (SHARCNET: www.sharcnet.ca) and it was supported by grants from the Social Sciences and Humanities Research Council of Canada (SSHRC: www.sshrc-crsh.gc.ca).

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

1. Introduction

Interest in modeling volatility dynamics of time-series data has been growing in many areas of empirical economics and finance. The management of financial asset portfolios involves estimation and forecasting of financial asset returns dynamics. Analysis of the dynamic evolution of variances and covariances of asset portfolios is used for a number of purposes, such as forecasting value-at-risk (VaR) thresholds to determine compliance with the Basel Accord.

In a recent paper by Caporin and McAleer (2012), henceforth CM, the BEKK model (Engle and Kroner, 1995) and the dynamic conditional correlation (DCC) model (Engle, 2002) are singled out as the "two most widely used models of conditional covariances and correlations" in the class of multivariate GARCH models. CM provide a deep and insightful theoretical analysis of the similarities and differences between the BEKK and DCC. CM note that DCC appears to be preferred to BEKK empirically because of the perceived curse of dimensionality associated with the latter. In an important insight, CM argue that this is a misleading interpretation of the suitability of the two models to be used in practice, as the full unconstrained comparable model versions are both affected by the curse of dimensionality alike. CM argue that either model is able to do "virtually everything the other can do" and hence their usage should depend on the object of interest: BEKK is best suited for conditional covariances while DCC for conditional correlations.

For either model type, parameter parsimony can be achieved by imposition of parametric restrictions on the full model version (Ding and Engle, 2001; Engle, Shephard, and Sheppard, 2009; Bilio and Caporin, 2009). However, doing so is traded off with model specification issues. For example, Burda and Maheu (2012) show that restricted versions of the BEKK, the scalar and diagonal models, are clearly dominated by the full BEKK version in terms of model fit. Indeed, such parameter restrictions generally operate on the parameters driving the model dynamics.

An alternative way of reducing dimensionality is via so-called *targeting* which imposes a structure on the model intercept based on sample information. Targeting yields model constraints in a structured way so that its implied long-run solution of the covariance (or

correlation) ensures its independence from any of the parameters driving the model dynamics. Hence, targeting differs from merely setting a subset of parameters to zero as in the scalar and diagonal models. Variance targeting estimation was originally proposed by Engle and Mezrich (1996) as a two-step estimation procedure, whereby the unconditional covariance matrix of the observed process is estimated in the first step and the remaining parameters are estimated in the second step. However, Aielli (2011) shows that the advantage of targeting as a tool for controlling the curse of dimensionality does not hold for DCC models, as the sample correlation is an inconsistent estimator of the long-run correlation matrix. CM also argue that targeting can be rigorously applied only to BEKK and its use in DCC models inappropriate.

At the same time, enforcing positive definiteness of the conditional covariance matrices in the BEKK with targeting implies a set of model constraints that are nonlinear in parameters and, according to CM, "extremely complicated, except for the scalar case" (p. 742). Furthermore, the model fit of the targeted BEKK relative to the full BEKK remains an open empirical question: does parameter dimensionality reduction via targeting trade off with a model fit that is substantially worse, as in the case of the scalar and diagonal BEKK? Perhaps due to these difficulties and unknowns the BEKK with targeting has not been used extensively in practice, despite its potential benefits.

In this paper, we suggest an approach to estimation and inference of the BEKK model with targeting based on Constrained Hamiltonian Monte Carlo (CHMC) (Neal, 2011), a recent statistical method that deals very effectively with complicated nonlinear model constraints in the context of relatively high-dimensional problems and computationally costly log-likelihoods. It is particularly useful in cases where it is difficult to accurately approximate the log-likelihood surface around the current parameter draw in real time necessary for obtaining sufficiently high acceptance probabilities in importance sampling (IS) methods, such as for recursive models in finance. Complicated constraints pose further challenge for IS-based methods. In such situations one would typically resort to random walk (RW) style sampling that is fast to run and does not require the knowledge of the properties of the underlying log-likelihood. Jin and Maheu (2012) used the RW sampler for covariance targeting in a class of dynamic component models. However, RW mechanisms can lead to very slow exploration of the parameter space with high autocorrelations among draws which might

require a prohibitively large size of the Markov chain to be obtained in implementation to achieve satisfactory mixing and convergence. CHMC combines the advantages of sampling that is relatively cheap with RW-like intensity but superior parameter space exploration.

CHMC falls within a general class of Markov chain Monte Carlo (MCMC) methods that can be used under both the Bayesian and the classical paradigm, applied to posterior densities or directly to model likelihoods without prior information (Chernozhukov and Hong, 2003). To the best of our knowledge, the constrained version of HMC has not yet been used in the economics literature. We elaborate in detail on the CHMC procedure, its application to the BEKK with targeting, and its usefulness in applicability to a wider class of problems of similar nature. We also contrast CHMC performance with RW sampling.

Using CHMC, we perform a model comparison of the BEKK with targeting and the full BEKK, providing evidence that favors the BEKK with targeting in terms of marginal likelihood. Thus, we show that the BEKK with targeting presents an effective way of reducing parameter dimensionality without compromising the model fit, unlike the scalar or diagonal BEKK. The model comparison is conducted in the context of an application concerning a multivariate dynamic volatility analysis of a foreign exchange rate returns portfolio. We believe that the suggested estimation approach and comparison evidence, along with the provided computer code for implementation, will encourage wider use of the BEKK model with targeting as a valuable tool for analysts and practitioners.

The remainder of the paper is organized as follows. Section 2 describes the BEKK model and the targeting constraints. Section 3 elaborates on Constrained Hamiltonian Monte Carlo. Section 4 introduces the application and reports on model comparison results. Section 5 concludes.

2. BEKK with Targeting

The BEKK class of multivariate GARCH models was introduced by Engle and Kroner (1995). Their specification was sufficiently general to allow the inclusion of special factor structures (Bauwens, Laurent, and Rombouts, 2006). Here we focus on the BEKK with all orders set to one, which is the predominantly used version in applications (Silvennoinen and Teräsvirta, 2009). Let r_t be a $N \times 1$ vector of asset returns with $t = 1, \dots, T$ and

denote the information set as $\mathcal{F}_{t-1} = \{r_1, \dots, r_{t-1}\}$. Under the BEKK model the returns follow

$$(2.1) \quad r_t | \mathcal{F}_{t-1} \sim NID(0, \Sigma_t)$$

$$(2.2) \quad \Sigma_t = CC' + Ar_{t-1}r'_{t-1}A' + B\Sigma_{t-1}B'$$

Σ_t is a positive definite $N \times N$ conditional covariance matrix of r_t given information at time $t - 1$, C is a lower triangular matrix and A and B are $N \times N$ matrices. We maintain a Gaussian assumption and a zero intercept for simplicity. The total number of parameters in this model is $N(N+1)/2 + 2N^2$, collected into the vector $\theta = (\text{vech}(C)', \text{vech}(A)', \text{vech}(B)')$.

The BEKK model with targeting is defined as

$$(2.3) \quad \Sigma_t = \bar{\Sigma} + A(r_{t-1}r'_{t-1} - \bar{\Sigma})A' + B(\Sigma_{t-1} - \bar{\Sigma})B'$$

where

$$\begin{aligned} E[\Sigma_t] &= E[\Sigma_{t-1}] = \bar{\Sigma} \\ E[r_{t-1}r'_{t-1}] &= \bar{\Sigma} \end{aligned}$$

There are two types of constraints on the model that need to be taken into account. The first type, covariance stationarity constraints, are generally simple to impose, as shown by Engle and Kroner (1995). The second type, positive definiteness of the conditional covariance matrices, is satisfied by construction in the full BEKK version, but becomes difficult to deal with in the targeted case. CM note that one way in which positive definiteness of Σ_t can be guaranteed is by imposing positive definiteness on $\bar{\Sigma} - A\bar{\Sigma}A' - B\bar{\Sigma}B'$ but note that such constraint is complicated. Indeed, Figure 1 below shows the effect of this constraint on the parameter space on an example of a log-likelihood plot over the vector field of covariance parameters $a_{12} \times b_{12}$, where $A = [a_{ij}]$, $B = [b_{ij}]$, for $N = 3$ (left) and $N = 4$ (right). In the white areas not covered by a contour plot the constraint is violated. The QMLE argmax is on the parameter space boundary which complicates inference. Implementation of this constraint is relatively cheap, since only one constraint check is required per likelihood evaluation, but its consequences are unfavorable for further QMLE analysis. An alternative way of imposing positive definiteness of Σ_t that we adopt here is to check that all its eigenvalues are positive for each t and take corrective action in the sampling procedure if the constraint is violated. The effect of this constraint on the parameter space is shown in

Figure 2 which plots the log-likelihood around the mode over the same vector space as in Figure 1. In our application, we did not detect a parameter without a well-defined mode due to this constraint.

FIGURE 1. Targeting Constraints on the Intercepts

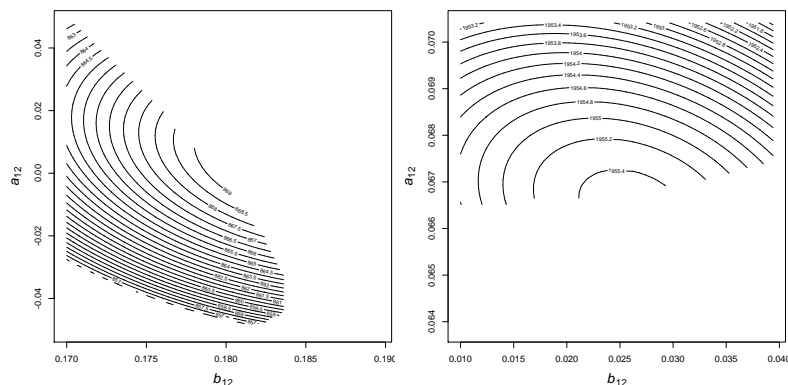
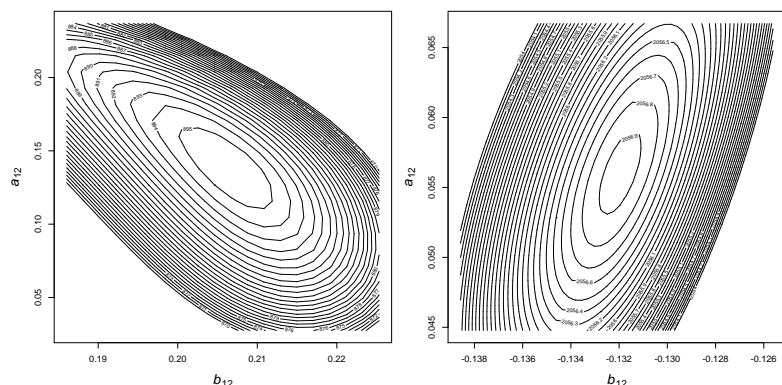


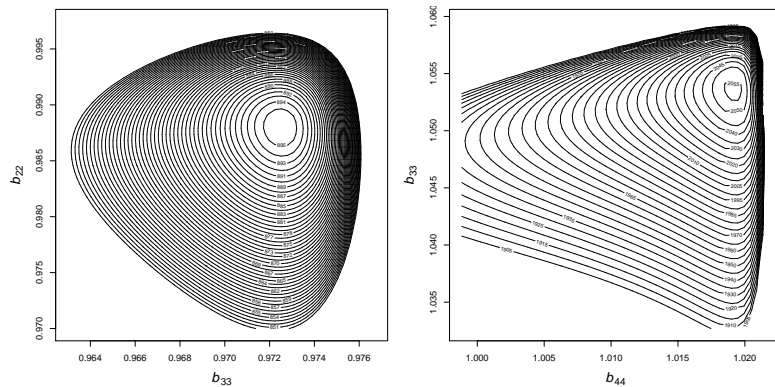
FIGURE 2. Targeting Constraints on Σ_t



Nonetheless, an occurrence of an argmax on or very close the parameter space boundary cannot be ruled out a-priori in general. Furthermore, the log-likelihood surface can be highly asymmetric, as illustrated in Figure 3 on the case of a contour plot over the space of parameter vector $b_{22} \times b_{33}$ for $N = 3$, and $b_{33} \times b_{44}$ for $N = 4$. The asymmetry appears more severe in higher dimensions. Due to both of these features, the accuracy of QMLE asymptotic inference, based on a Fisher information matrix that is symmetric by construction, is questionable. MCMC-based procedures, including CHMC presented in the

next Section, provide an alternative estimation and inference approach for either constraint case, unfettered by the aforementioned likelihood irregularities.

FIGURE 3. Likelihood Kernel Skewness



3. Constrained Hamiltonian Monte Carlo

The original Hamiltonian (or Hybrid) Monte Carlo (HMC) has its roots in the physics literature where it was introduced as a fast method for simulating molecular dynamics (Duane et al., 1987). It has since become popular in a number of application areas including statistical physics (Akhmatskaya et al., 2009; Gupta et al., 1988), computational chemistry (Tuckerman et al., 1993), or a generic tool for Bayesian statistical inference (Neal, 1993, 2011; Ishwaran, 1999; Liu, 2004; Beskos et al., 2010). HMC and its constrained version, CHMC, applies to a general class of models, nesting the BEKK, 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 can be characterized by a family \mathcal{P}_θ of probability measures on a measurable space (Θ, \mathcal{B}) where \mathcal{B} is the Borel σ -algebra. The purpose of 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 $h(\theta)$ of θ that are expressed as integrals. These functionals include objects of interest for inference on θ such as quantiles of $\pi(\theta)$.

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

$$\lim_{R \rightarrow \infty} \frac{1}{R} \sum_{r=1}^R h(\theta_r) = E_{\pi}[h(\theta)]$$

A Markov chain with this property is called ergodic. As a means of approximation we rely on large but finite number of draws $R \in \mathbb{N}$ which the analyst has the discretion to select in applications.

$G(\theta_r)$ can be obtained from a given (economic) 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 employed for $\theta_{r+1}|\theta_r$ from $G(\theta_r)$; see Chib and Greenberg (1995) for a detailed overview. Suppose we have a proposal-generating density $q(\theta_{r+1}^*|\theta_r)$ where θ_{r+1}^* is a proposed state given the current state θ_r of the Markov chain. The M-H principle stipulates that θ_{r+1}^* be accepted as the next state θ_{r+1} with the acceptance probability

$$(3.1) \quad \alpha(\theta_r, \theta_{r+1}^*) = \min \left[\frac{\pi(\theta_{r+1}^*)q(\theta_r|\theta_{r+1}^*)}{\pi(\theta_r)q(\theta_{r+1}^*|\theta_r)}, 1 \right]$$

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

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

which is sufficient for ergodicity. $\alpha(\theta_{r+1}^*, \theta_r)$ is the probability of the move $\theta_r|\theta_{r+1}^*$ if the dynamics of the proposal generating mechanism were to be reversed. The proposal-generating density $q(\theta_{r+1}^*|\theta_r)$ can be chosen to be sampled easily even though $\pi(\theta)$ may be difficult or expensive to sample from. The popular Gibbs sampler arises as a special case when the M-H sampler is factored into conditional densities. The proposal draws $\theta_{r+1}^*|\theta_r$ from $q(\theta_{r+1}^*|\theta_r)$ in (3.1) are generated in one step.

In contrast, Hamiltonian Monte Carlo (HMC) uses a whole sequence of proposal steps whereby the last step in the sequence becomes the proposal draw. This 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 yielding 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 $\{\theta_r^k, \gamma_r^k\}_{k=1}^L$ starting from the current state $(\theta_r, \gamma_r) = (\theta_r^0, \gamma_r^0)$ and yielding a proposal $(\theta_r, \gamma_r) = (\theta_r^L, \gamma_r^L)$. The detailed balance is then satisfied using the acceptance probability

$$(3.2) \quad \alpha(\theta_r, \gamma_r; \theta_{r+1}^*, \gamma_{r+1}^*) = \min \left[\frac{\pi(\theta_{r+1}^*, \gamma_{r+1}^*)q(\theta_r, \gamma_r | \theta_{r+1}^*, \gamma_{r+1}^*)}{\pi(\theta_r, \gamma_r)q(\theta_{r+1}^*, \gamma_{r+1}^* | \theta_r, \gamma_r)}, 1 \right]$$

In CHMC, constraints are incorporated into the HMC proposal mechanism via "hard walls" representing a barrier against which the proposal sequence, simulating a particle movement, bounces off elastically. Constraints thus do not provide grounds for proposal rejection, 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. We further synthesize the technical principles of CHMC in a generally accessible form in the Appendix A. In the next Section we apply CHMC to the task of model comparison of the full BEKK and the BEKK with targeting.

4. Application and Model Comparison

The model comparison is performed in the context of an empirical application with data on percent log-differences of foreign exchange spot rates for AUD/USD, GBP/USD, CAD/USD, and EUR/USD, from 2000/01/04 to 2011/12/30, total of $T = 3,009$ observations. We consider cases with two, three and four variables ($N = 2, 3, 4$) with data used in the order presented. The associated parameter dimensionality for the full BEKK is 11, 24, and 42, and for the BEKK with targeting 8, 18, and 32, respectively. A time series plot of the four series is shown in Figure B.1 and summary statistics are provided in Table B.1 in Appendix B. The sample mean for all series is close to 0 and skewness is small. The sample correlations indicate that all series tend to move together.

For identification, the diagonal elements of C and the first element of both A and B are restricted to be positive (Engle and Kroner, 1995). All priors, other than the model identification conditions, are set to be diffuse. In the implementation, we utilize the analytical expressions for the gradient of the BEKK log-likelihood from Hafner and Herwartz (2008), which in the case of the BEKK with targeting is subject to the targeting constraints (2.3). Although numerical estimates of the gradients could also be used in forming the proposals, evaluation of analytical expressions increases the implementation speed. The initial Σ_1 in the GARCH recursion is set to the sample covariance. Starting from the modal parameter values we collect a total of 50,000 posterior draws for inference, with 5,000 burnin section. The length of the proposal sequence was set to $L = 50, 30, 20$ for the cases $N = 2, 3, 4$ respectively, and the stepsize ϵ tuned to achieve acceptance rates close to 0.8.

Table B.2 reports model comparison results of the full BEKK and BEKK with targeting in terms of marginal log-likelihood (Gelfand and Dey, 1994; Geweke, 2005). Since all parameters are integrated out, marginal likelihood does not explicitly depend on the dimensionality of the parameter space and hence is suitable for comparison of models with different dimensions. In all cases, $N = 2, 3, 4$, the evidence strongly favors the BEKK model with targeting over the full BEKK model, and this effect increases with dimensionality of variables N . The evolution of the mean of the conditional covariances over time is presented in Figure B.2 for the full BEKK, and in Figure B.3 for the BEKK with targeting, in four variables. Both plots are virtually identical when examined closely, indicating a minimal loss of model prediction capability as a result of targeting.

Markov chain convergence diagnostics, obtained using the R package coda, for both models and variable dimensions are reported in Table B.3 confirming the validity of model inference and comparison. All chains have converged within the burnin section, as evidenced by the reported z-scores (mean, minimum, maximum, and standard deviation) of the Geweke (1992) convergence test standardized z-score and p-values of the Heidelberger and Welch (1983) stationary distribution test. The Raftery burnin control diagnostic (Raftery and Lewis, 1995) confirms the sufficiency of the length of the burnin section. The CPU run time, also reported in Table B.3, took in the order of hours and was virtually equivalent to the wall-clock run time. All chains were obtained using fortran code with Intel compiler on a 2.8 GHz Unix machine.

The advantages of HMC-based procedures over random walk (RW) sampling have been well documented (Neal 2011; Pakman and Paninski 2012). We illustrate the CHMC versus RW comparison in Figure B.4, showing the trace plot of the Markov chain for the conditional variance parameter a_{44} , in the BEKK with targeting, with 4 variables and 32 parameters. The CHMC trace (left) mixes very well, exploring the tails of the likelihood kernel, while the RW trace (right) stays within a relatively narrow band with minimal tail exploration and without any signs of convergence. Overall, our results support the use of the BEKK with targeting, and CHMC provides a suitable method to sample effectively from its likelihood kernel with nonlinear constraints.

5. Conclusions

In this paper we suggest an effective approach for estimation and inference on the BEKK GARCH model with targeting. The approach is based on Constrained Hamiltonian Monte Carlo (CHMC), a recent statistical technique devised to handle nonlinear constraints in the context of relatively costly and irregular likelihoods. Based on a model comparison with the unrestricted version of the BEKK we present evidence favoring the BEKK with targeting in terms of marginal likelihood. Due to its potential and applicability to similar types of problems, we detail CHMC in a generally accessible form, and provide computer code for its implementation. We also provide a comparison of the CHMC and random walk sampling. We believe that the elaborated estimation approach and comparison evidence will encourage wider use of both the BEKK model with targeting and CHMC as valuable tools for analysts and practitioners.

Appendix A. Statistical Properties of Constrained Hamiltonian Monte Carlo

In this Section we provide the stochastic background for CHMC. This synthesis is based on previously published material (Neal 2011; Burda and Maheu, 2012; and references therein). However, the bulk of literature presenting HMC methods does so in terms of the physical laws of motion based on preservation of total energy in the phase-space. Here we take a fully stochastic perspective familiar to the applied econometrician. The CHMC principle is thus presented in terms of the joint density over the augmented parameter space leading to a Metropolis acceptance probability update.

A.1. CHMC Principle

Consider a vector of parameters of interest $\theta \in \mathbb{R}^d$ distributed according to the density kernel $\pi(\theta)$. Let $\gamma \in \mathbb{R}^d$ denote a vector of auxiliary parameters with $\gamma \sim \Phi(\gamma; 0, M)$ where Φ denotes the Gaussian distribution with mean vector 0 and covariance matrix M , independent of θ . M can be either set to the identity matrix or to a covariance matrix estimated by the inverse of the Fisher information matrix around the mode, as we do in our implementation. Denote the joint density of (θ, γ) by $\pi(\theta, \gamma)$. Denote the constraint on the parameter space by a density kernel $\exp(C_r(\theta))$ where

$$C_r(\theta) = \begin{cases} 0 & \text{if the constraint is satisfied} \\ C_r > 0 \text{ s.t. } \lim_{r \rightarrow \infty} C_r = \infty & \text{if the constraint is violated} \end{cases}$$

Then the negative of the logarithm of the joint density of (θ, γ) is given by the constrained Hamiltonian equation¹

$$(A.1) \quad H(\theta, \gamma) = -C_r(\theta) - \ln \pi(\theta) + \frac{1}{2} \ln \left((2\pi)^d |M| \right) + \frac{1}{2} \gamma' M^{-1} \gamma$$

Constrained Hamiltonian Monte Carlo (CHMC) is formulated in the following steps:

- (1) Draw an initial auxiliary parameter vector $\gamma_r^0 \sim \Phi(\gamma; 0, M)$;
- (2) Transition from (θ_r, γ_r) to $(\theta_r^L, \gamma_r^L) = (\theta_{r+1}^*, \gamma_{r+1}^*)$ according to the constrained Hamiltonian dynamics;
- (3) Accept $(\theta_{r+1}^*, \gamma_{r+1}^*)$ with probability

$$(A.2) \quad \alpha(\theta_r, \gamma_r; \theta_{r+1}^*, \gamma_{r+1}^*) = \min \left[\exp \left(-H(\theta_{r+1}^*, \gamma_{r+1}^*) + H(\theta_r, \gamma_r) \right), 1 \right],$$

otherwise keep (θ_r, γ_r) 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 provides a stochastic initialization of the system akin to a RW draw in order to make the resulting Markov chain $\{(\theta_r, \gamma_r)\}_{r=1}^R$ irreducible and aperiodic (Ishwaran, 1999).

In contrast to RW, this so-called refreshment move is performed on the auxiliary variable γ

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

as opposed to the original parameter of interest θ , setting $\theta_r^0 = \theta_r$. The initial refreshment draw of γ_r^0 is equivalent to a Gibbs step on the parameter space of (θ, γ) accepted with probability 1. Since it only applies to γ , it will leave the target joint distribution of (θ, γ) invariant and subsequent steps can be performed conditional on γ_r^0 (Neal, 2011).

Step 2 constructs a sequence $\{\theta_r^k, \gamma_r^k\}_{k=1}^L$ according to the Hamiltonian dynamics starting from the current state (θ_r^0, γ_r^0) , with the newly refreshed γ_r^0 , and setting the last member of the sequence as the CHMC new state proposal $(\theta_{r+1}^*, \gamma_{r+1}^*) = (\theta_r^L, \gamma_r^L)$. The transition from (θ_r^0, γ_r^0) to (θ_r^L, γ_r^L) via the proposal sequence $\{\theta_r^k, \gamma_r^k\}_{k=1}^L$ taken according to the discretized Hamiltonian dynamics (A.5)–(A.7) is fully deterministic, placing a Dirac delta probability mass $\delta(\theta_r^k, \gamma_r^k) = 1$ on each (θ_r^k, γ_r^k) conditional on (θ_r^0, γ_r^0) . The CHMC acceptance probability in (A.2) is specified in terms of the difference between the Hamiltonian (A.1) evaluated at the initial (θ_r^0, γ_r^0) and at the proposal $(\theta_{r+1}^*, \gamma_{r+1}^*)$. The role of the Hamiltonian dynamics is to ensure that the acceptance probability (A.2) for $(\theta_{r+1}^*, \gamma_{r+1}^*)$ is kept close to 1. This corresponds to maintaining the difference $-H(\theta_{r+1}^*, \gamma_{r+1}^*) + H(\theta_r^0, \gamma_r^0)$ close to zero throughout the sequence $\{\theta_r^k, \gamma_r^k\}_{k=1}^L$. This property of the transition from (θ_r, γ_r) to $(\theta_{r+1}^*, \gamma_{r+1}^*)$ can be achieved by conceptualizing θ and γ as functions of continuous time t and specifying their evolution using the Hamiltonian dynamics equations²

$$(A.3) \quad \frac{d\theta_i}{dt} = \frac{\partial H(\theta, \gamma)}{\partial \gamma_i} = [M^{-1}\gamma]_i$$

$$(A.4) \quad \frac{d\gamma_i}{dt} = -\frac{\partial H(\theta, \gamma)}{\partial \theta_i} = \nabla_{\theta_i} \ln \pi(\theta)$$

for $i = 1, \dots, d$. For any discrete time interval of duration s , (A.3)–(A.4) define a mapping T_s from the state of the system at time t to the state at time $t + s$. The differential equations (A.3)–(A.4) are generally solved by numerical methods, typically the Stormer-Verlet (or leapfrog) numerical integrator (Leimkuhler and Reich, 2004). For each step in constructing the proposal sequence $\{\theta_r^k, \gamma_r^k\}_{k=1}^L$, CHMC discretizes the Hamiltonian dynamics (A.3)–(A.4) as follows: for some small $\varepsilon \in \mathbb{R}$ first take a half- ε step in γ

$$(A.5) \quad \gamma(t + \varepsilon/2) = \gamma(t) + (\varepsilon/2)\nabla_{\theta} \ln \pi(\theta(t)),$$

²In the physics literature, the Hamiltonian dynamics describe the evolution of (θ, γ) that keeps the total energy $H(\theta, \gamma)$ constant.

then take a full ε step in θ

$$(A.6) \quad \theta(t + \varepsilon) = \theta(t) + \varepsilon M^{-1} \gamma(t + \varepsilon/2),$$

check the constraint at $\theta(t + \varepsilon)$ for each dimension i of θ , if for any i the constraint is violated 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

$$(A.7) \quad \gamma(t + \varepsilon) = \gamma(t + \varepsilon/2) + (\varepsilon/2) \nabla_{\theta} \ln \pi(\theta(t + \varepsilon)).$$

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 $C_r(\theta) = 0$, and the move in θ only depends on $\partial H(\theta, \gamma) / \partial \gamma_i$ which does not involve $C_r(\theta)$, the exact functional form of $C_r(\theta)$ is inconsequential as long as it is differentiable in θ to define valid Hamiltonian dynamics.

From the statistical perspective, γ plays the role of an auxiliary variable that parametrizes (a functional of) $\pi(\theta)$ providing it with an additional degree of flexibility to maintain the acceptance probability close to one for every k . Even though $\ln \pi(\theta_r^k)$ can deviate substantially from $\ln \pi(\theta_r^0)$, resulting in favorable mixing for θ , the additional terms in γ in (A.1) compensate for this deviation maintaining the overall level of $H(\theta_r^k, \gamma_r^k)$ close to constant over $k = 1, \dots, L$ when used in accordance with (A.5)–(A.7), since $\frac{\partial H(\theta, \gamma)}{\partial \gamma_i}$ and $\frac{\partial H(\theta, \gamma)}{\partial \theta_i}$ enter with the opposite signs in (A.3)–(A.4). In contrast, without the additional parametrization with γ , if only $\ln \pi(\theta_r^k)$ 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 $(\theta_{r+1}^*, \gamma_{r+1}^*)$. In continuous time, or for $\varepsilon \rightarrow 0$, (A.3)–(A.4) would keep $-H(\theta_{r+1}^*, \gamma_{r+1}^*) + H(\theta_r, \gamma_r) = 0$ exactly resulting in $\alpha(\theta_r, \theta_{r+1}^*) = 1$ but for discrete $\varepsilon > 0$, in general, $-H(\theta^*, \gamma^*) + H(\theta, \gamma) \neq 0$ necessitating the Metropolis step.

The system (A.5)–(A.7) is time reversible and symmetric in (θ, γ) , which implies that the forward and reverse transition probabilities $q(\theta_r^L, \gamma_r^L | \theta_r^0, \gamma_r^0)$ and $q(\theta_r^0, \gamma_r^0 | \theta_r^L, \gamma_r^L)$ are equal: this simplifies the Metropolis-Hastings acceptance ratio in (3.2) to the Metropolis form $\pi(\theta_{r+1}^*, \gamma_{r+1}^*) / \pi(\theta_r^0, \gamma_r^0)$. From the definition of the Hamiltonian $H(\theta, \gamma)$ in (A.1) as the

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

$$(A.8) \quad \pi(\theta, \gamma) = \exp[-H(\theta, \gamma)] = \pi(\theta) \left((2\pi)^d |M| \right)^{-1/2} \exp\left(-\frac{1}{2} \gamma' M^{-1} \gamma\right)$$

Hence, the Metropolis acceptance probability takes the form

$$\begin{aligned} \alpha(\theta_r, \gamma_r; \theta_{r+1}^*, \gamma_{r+1}^*) &= \min \left[\frac{\pi(\theta_{r+1}^*, \gamma_{r+1}^*)}{\pi(\theta_r^0, \gamma_r^0)}, 1 \right] \\ &= \min \left[\exp(-H(\theta_{r+1}^*, \gamma_{r+1}^*) + H(\theta_r^0, \gamma_r^0)), 1 \right] \end{aligned}$$

The expression for $\alpha(\theta_r, \gamma_r; \theta_{r+1}^*, \gamma_{r+1}^*)$ shows, as noted above, that the CHMC acceptance probability is given in terms of the difference of the Hamiltonian equations $H(\theta_r^0, \gamma_r^0) - H(\theta_{r+1}^*, \gamma_{r+1}^*)$. The closer can we keep this difference to zero, the closer the acceptance probability is to one. A key feature of the Hamiltonian dynamics (A.3)–(A.4) in Step 2 is that they maintain $H(\theta, \gamma)$ constant over the parameter space in continuous time conditional on $H(\theta_r^0, \gamma_r^0)$ obtained in Step 1, while their discretization (A.5)–(A.7) 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.

Appendix B. Application and Model Comparison Results

FIGURE B.1. Time-series of Log-differences in Foreign Exchange Rates

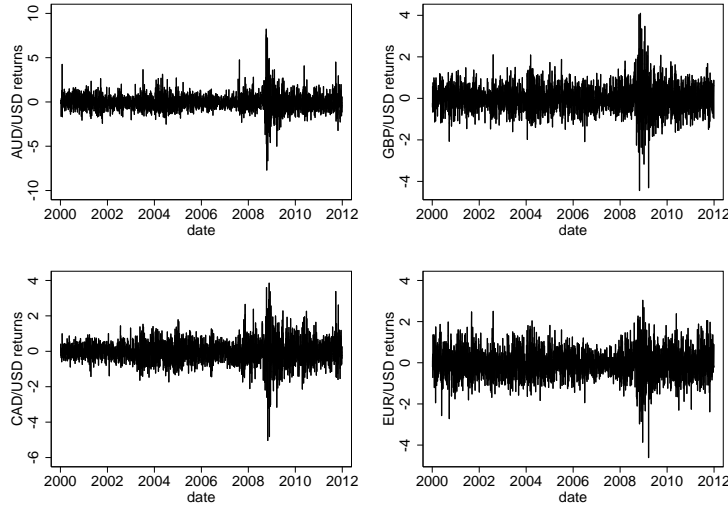


TABLE B.1. Summary Statistics

	Mean	s.d.	Skewness	Sample Correlation			
AUD/USD	-0.0148	0.9062	0.6607	1	0.5621	0.6361	0.5906
GBP/USD	0.0017	0.6233	0.1466		1	0.4613	0.6794
CAD/USD	-0.0118	0.6188	-0.0500			1	0.4873
EUR/USD	-0.0076	0.6682	-0.1068				1

TABLE B.2. Marginal log-Likelihood

Variables	BEKK		BEKK with Targeting	
	Parameters	Marginal lnL	Parameters	Marginal lnL
2	11	-30,447.06	8	-22,146.67
3	24	-65,402.69	18	-48,806.99
4	42	-114,063.13	32	-86,417.73

TABLE B.3. Convergence Diagnostics

	BEKK			BEKK with Targeting		
	2	3	4	2	3	4
Variable dimension	2	3	4	2	3	4
Parameter dimension	11	24	42	8	18	32
CPU Time (hrs:min)	1:35	5:50	9:53	1:14	4:53	6:45
Geweke z-score mean	0.220	-0.199	-0.127	0.040	0.022	-0.285
Geweke z-score min	-0.803	-1.704	-1.896	-0.761	-1.522	-1.669
Geweke z-score max	1.109	1.705	1.738	1.213	1.430	1.477
Geweke z-score s.d.	0.778	1.171	0.868	0.824	0.878	0.877
Heidel p-value mean	0.544	0.414	0.307	0.415	0.537	0.484
Heidel p-value min	0.137	0.086	0.063	0.244	0.089	0.053
Heidel p-value max	0.956	0.892	0.991	0.596	0.971	0.986
Heidel p-value s.d.	0.321	0.223	0.262	0.116	0.268	0.266
Raftery burnin mean	9	27	90	9	87	132
Raftery burnin min	6	9	5	4	30	40
Raftery burnin max	20	75	351	16	216	280
Raftery burnin s.d.	4	16	65	4	49	56
Proposal steps L	50	30	20	50	30	20

FIGURE B.2. Conditional Covariances: Full BEKK

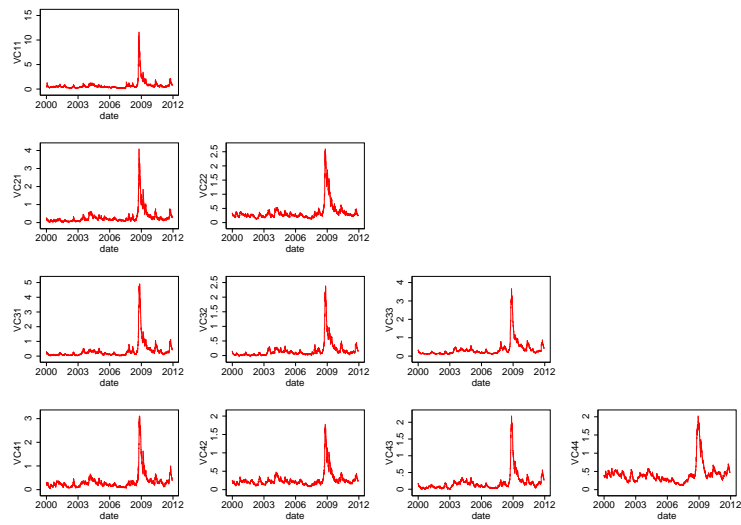


FIGURE B.3. Conditional Covariances: BEKK with Targeting

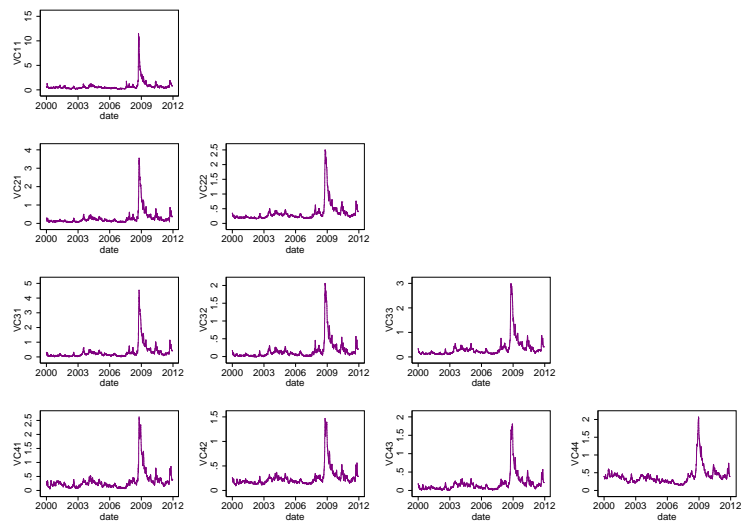
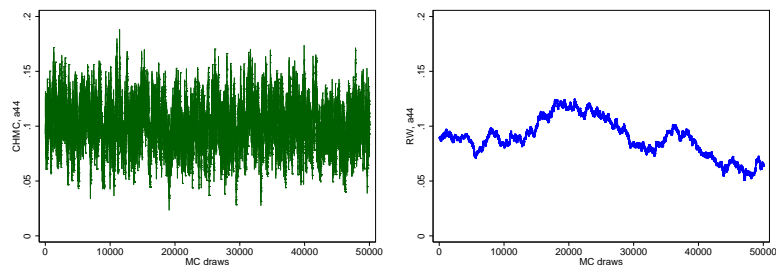


FIGURE B.4. CHMC (left) vs Random Walk Sampling (right)



References

- Aielli, G. P. (2011). Dynamic conditional correlation: On properties and estimation. Available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1507743, SSRN Working Paper,.
- 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.
- Bauwens, L., S. Laurent, and J. V. K. Rombouts (2006). Multivariate garch models: a survey. *Journal of Applied Econometrics* 21, 79–109.
- Beskos, A., N. S. Pillai, G. O. Roberts, J. M. Sanz-Serna, and A. M. Stuart (2010). Optimal tuning of the hybrid monte-carlo algorithm. Working paper, arxiv:1001.4460v1 [math.pr].
- Billio, M. and M. Caporin (2009). A generalized dynamic conditional correlation model for portfolio risk evaluation. *Mathematics and Computers in Simulation* 79(8), 2566–2578.
- Burda, M. and J. M. Maheu (2012). Bayesian adaptively updated hamiltonian monte carlo with an application to high-dimensional BEKK GARCH models. *Studies in Nonlinear Dynamics & Econometrics*, forthcoming.
- Caporin, M. and M. McAleer (2012). Do we really need both BEKK and DCC? a tale of two multivariate garch models. *Journal of Economic Surveys* 26(4), 736–751.
- Chernozhukov, V. and H. Hong (2003). An MCMC approach to classical estimation. *Journal of Econometrics* 115(3), 293–346.
- Chib, S. and E. Greenberg (1995). Understanding the metropolis-hastings algorithm. *American Statistician* 49(4), 327–335.
- Ding, Z. and R. Engle (2001). Large scale conditional covariance matrix modeling, estimation and testing. *Academia Economic Papers* 29, 157–184.
- Duane, S., A. Kennedy, B. Pendleton, and D. Roweth (1987). Hybrid monte carlo. *Physics Letters B* 195(2), 216–222.
- Engle, R. F. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business and Economic Statistics* 20, 339–350.
- Engle, R. F. and K. F. Kroner (1995). Multivariate simultaneous generalized arch. *Econometric Theory* 11(1), 122–150.
- Engle, R. F., N. Shephard, and K. Sheppard (2009). Fitting vast dimensional time-varying covariance models. Available at SSRN: <http://ssrn.com/abstract=1354497>.
- Gelfand, A. and D. Dey (1994). Bayesian model choice: Asymptotics and exact calculations. *Journal of the Royal Statistical Society, Ser. B* 56, 501–514.
- Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In J. Bernardo, J. O. Berger, A. Dawid, and A. Smith (Eds.), *Bayesian Statistics 4*. Clarendon Press, Oxford, UK.
- Geweke, J. (2005). *Contemporary Bayesian Econometrics and Statistics*. Hoboken, New Jersey: Wiley.

- Gupta, R., G. Kilcup, and S. Sharpe (1988). Tuning the hybrid monte carlo algorithm. *Physical Review D* 38(4), 1278–1287.
- Hafner, C. M. and H. Herwartz (2008). Analytical quasi maximum likelihood inference in multivariate volatility models. *Metrika* 67, 219–239.
- Heidelberger, P. and P. Welch (1983). Simulation run length control in the presence of an initial transient. *Operations Research* 31, 1109–44.
- Ishwaran, H. (1999). Applications of hybrid monte carlo to generalized linear models: Quasicomplete separation and neural networks. *Journal of Computational and Graphical Statistics* 8, 779–799.
- Jin, X. and J. Maheu (2012). Modeling realized covariances and returns. forthcoming, *Journal of Financial Econometrics*.
- 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.
- 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.
- Pakman, A. and L. Paninski (2012). Exact hamiltonian monte carlo for truncated multivariate gaussians. Available at <http://arxiv.org/abs/1208.4118>, Columbia University.
- Plummer, M., N. Best, K. Cowles, K. Vines, D. Sarkar, and R. Almond (2012). The R package coda. version 0.16-1.
- Raftery, A. and S. Lewis (1995). The number of iterations, convergence diagnostics and generic metropolis algorithms. In W. Gilks, D. Spiegelhalter, and S. Richardson (Eds.), *Practical Markov Chain Monte Carlo*. London, U.K.: Chapman and Hall.
- Robert, C. P. and G. Casella (2004). *Monte Carlo statistical methods* (Second ed.). New York: Springer.
- Silvennoinen, A. and T. Teräsvirta (2009). Modeling multivariate autoregressive conditional heteroskedasticity with the double smooth transition conditional correlation garch model. *Journal of Financial Econometrics* 7(4), 373–411.
- 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.