Bayesian analysis of chaos: The joint return-volatility dynamical system

Tsionas, Mike G. and Michaelides, Panayotis G.

Lancaster University Management School, Athens University of Economics Business, Systemic Risk Centre, London School of Economics, National Technical University of Athens

2017

Bayesian analysis of chaos: The joint return-volatility dynamical system

Mike G. Tsionas*and Panayotis G. Michaelides †

Abstract

We use a novel Bayesian inference procedure for the Lyapunov exponent in the dynamical system of returns and their unobserved volatility. In the dynamical system, computation of largest Lyapunov exponent by traditional methods is impossible as the stochastic nature has to be taken explicitly into account due to unobserved volatility. We apply the new techniques to daily stock return data for a group of six world countries, namely USA, UK, Switzerland, Netherlands, Germany and France, from 2003 to 2014 by means of Sequential Monte Carlo for Bayesian inference. The evidence points to the direction that there is indeed noisy chaos both before and after the recent financial crisis. However, when a much simpler model is examined where the interaction between returns and volatility is not taken into consideration jointly, the hypothesis of chaotic dynamics does not receive much support by the data ("neglected chaos").

Key Words: Noisy Chaos; Lyapunov exponent; Neural networks; Bayesian analysis; Sequential Monte Carlo, World Economy.

*Lancaster University Management School, LA1 4YX U.K. & Athens University of Economics and Business, Greece, m.tsionas@lancaster.ac.uk

†National Technical University of Athens, School of Applied Mathematics and Physics, Greece, pmichael@central.ntua.gr

1 Introduction

The field of chaos was developed several decades ago in Physics to explain some strange-looking behaviours that lacked order. According to Lahmiri (2017): "a chaotic system is a random-looking nonlinear deterministic process with irregular periodicity and sensitivity to initial conditions". Up until recently, the various tests for chaotic behavior have been implemented primarily in the fields of Science, especially in meteorological and climate change data where noise is usually absent. However, over the last years, chaos has been the subject of research in various fields, including those of economics and finance, where a typical approach in these fields has been that the dominant dynamics are of a stochastic nature, usually described by a given probability function. Relatively recently, some researchers in the field of finance and financial economics, have found some evidence in favour of chaotic dynamics. For instance, Mishra et al. (2011) found some evidence of chaotic dynamics in the Indian stock market, Yousefpoor et al. (2008) in the Tehran stock market, Kyrtsou and Terraza (2002) in the French CAC 40 index, Çoban and Büyüklü (2009), in the Turkish Lira-USD exchange rate, and Das and Das (2007) in exchange rates.

In this context, tests for chaos using artificial neural networks (ANN) have gained considerable popularity, recently (BenSaïda, 2014, BenSaïda and Latimi, 2013). In financial time series, however, it is not sufficient to account for a flexible functional form to represent state dynamics. Time-varying conditional variance is a key characteristic of these series and, most often, this is ignored or modeled with simple models such as the EGARCH (Lahmiri, 2017). Although the EGARCH model (Nelson, 1991) is popular, the consensus is that stochastic volatility models (Jacquier et al., 2014) are better suited for financial data but considerably more computationally intensive. As BenSaïda (2014) put it: "[F]ew studies have considered studying the dynamics of financial and economic time series in times of political or economic instability to better understand their behaviour from an econophysics perspective"; for examples and applications see Cajueiro and Tabak (2009), Sensoy (2013), Lahmiri (2015), Morales et al. (2012).

In this paper, we consider a bivariate dynamical system consisting of returns and their volatility. Both functional forms are modeled via ANNs and feedback is allowed between returns and volatility. We propose a novel way of computing the largest Lyapunov exponent as the dynamical system is inherenty stochastic due to the presence of stochastic volatility. More precisely, in this work, we test for the presence of noisy chaotic dynamics before and after the 2008 international financial crisis. The data employed consist of stock indices for six major countries, namely: USA, UK, Switzerland, Netherlands, Germany and France, and their corresponding implied volatility indices. The sample covers the period from 15th May 2003 to 25th November 2014 in order to capture and study the financial crisis.

2 Model

2.1 General

Suppose a time series $\{x_t; t = 1, ..., T\}$ has the representation

$$y_t = f(y_{t-L}, y_{t-2L}, ..., y_{t-mL}) + u_t,\ u_t|\sigma_t \sim N(0, \sigma_t^2), t = 1, ..., T, \tag{1}$$

where σ_t^2 is the conditional variance, m is the embedding dimension (or the length of past dependence) and L is the time delay. The state space representation is:

$$F : \begin{bmatrix} y_{t-L} \\ y_{t-2L} \\ \vdots \\ y_{t-mL} \end{bmatrix} \rightarrow \begin{bmatrix} y_t = f(y_{t-L}, y_{t-2L}, ..., y_{t-mL}) + \varepsilon_t \\ y_{t-L} \\ \vdots \\ y_{t-(m-1)L} \end{bmatrix}. \tag{2}$$

Given initial conditions y_0 and a perturbation $\triangle y_0$ the time series after t periods changes by $\triangle y(y_0, t)$. The Lyapunov exponent is defined as:

$$\lambda = \lim_{\tau \to \infty} \tau^{-1} \ln \frac{|\triangle y(y_0, \tau)|}{|\Delta y_0|}, \tag{3}$$

and measures the average exponential divergence (positive exponent) or convergence (negative exponent) rate between nearby trajectories within time horizons that differ in terms of initial conditions by an infinitesimal amount. The Jacobian matrix J at a point $\boldsymbol{\chi}$is

$$J^t(\boldsymbol{x}) = \frac{df^t(\boldsymbol{\chi})}{d\boldsymbol{\chi}} \tag{4}$$

The Jacobian of the map in (2) is:

$$J_t = \begin{bmatrix} \frac{\partial f}{\partial y_{t-L}} & \frac{\partial f}{\partial y_{t-2L}} & \cdots & \frac{\partial f}{\partial y_{t-(m-1)L}} & \frac{\partial f}{\partial y_{t-mL}} \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}. \tag{5}$$

We write (1) compactly as:

$$y_t = f(\mathbf{x}_t) + u_t, \tag{6}$$

where $\mathbf{x}_t = [x_{t-L}, ..., x_{t-mL}] \in \Re^d$.

The Lyapunov exponent (following Eckmann and Ruelle , 1985) is:

$$\lambda = \lim_{M \to \infty} \frac{1}{2M} \log \nu^*, \tag{7}$$

where ν^* is the largest eigenvalue of matrix $T_M' T_M$, with

$$T_M = \prod_{t=1}^{M-1} J_{M-1}, \tag{8}$$

and $M \leq T$ is the block-length of equally spaced evaluation points, and J is the Jacobian matrix of the chaotic map f. One can take $M = T^{2/3}$(BenSaida, 2014, see also BenSaïda, 2014, BenSaïda and Latimi, 2013).

In this work, we use Artificial Neural Networks (ANNs) because they can approximate any smooth, nonlinearity, as the number of hidden units gets larger (Hornik, et al. 1989). Actually, it has been shown in Hornik (1991) that ANNs act as global approximations to various functions. For instance, see Michaelides et al. (2015) for an application to banking data. In this work, we approximate the map using a neural network:

$$y_t = \sum_{g=1}^{G} \alpha_g \varphi(\mathbf{x}_t' \boldsymbol{\beta}_g) + u_t, \tag{9}$$

where $\varphi(z) = \frac{1}{1+\exp(-z)},\ z \in \Re$ is the sigmoid activation function, and $\boldsymbol{\beta}_g \in \Re^L,\ \forall g = 1, ..., G$. We impose the identifiability constraints $\alpha_1 \leq ... \leq \alpha_G$. Asympotically normal tests for chaos have been proposed by considering the variance of the Lyapunov exponent which depend on the eigengvalues of matrix $T_M' T_M$. One can assume that the conditional variance follows a stochastic volatility model:

$$\log \sigma_t^2 = \mu + \rho \log \sigma_{t-1}^2 + \varepsilon_t,\ \varepsilon_t \sim iidN\left(0, \omega^2\right). \tag{10}$$

It is well known that even with linear models estimation of (10) is quite difficult. Insread of (9) and (10) we use a more flexible model:

$$\begin{array}{l} y_t = \sum_{g=1}^{G_1} \alpha_g \varphi(\mathbf{x}_t' \boldsymbol{\beta}_g + \sum_{l=1}^{l_2} \gamma_{gl} \sigma_{t-lL_2}^2) + u_t,\ u_t | \sigma_t \sim N(0, \sigma_t^2), t = 1, ..., T, \\ \log \sigma_t^2 = \sum_{g=1}^{G} \delta_g \varphi(\mathbf{x}_t' \boldsymbol{\zeta}_g + \sum_{l=1}^{l_2} \psi_{gl} \sigma_{t-lL_2}^2) + \varepsilon_t,\ \varepsilon_t \sim N(0, \omega^2), t = 1, ..., T, \end{array} \tag{11}$$

where L_2 is the time delay for the conditional variance. In this model: i) Returns depend on past values as well as lagged volatilities. ii) Volatility is given by a flexible neural network specification and depends on both past returns and lagged volatility values. We impose the identifiability constraints $\alpha_1 \leq ... \leq \alpha_G$ and $\delta_1 \leq ... \leq \delta_G$.

2.2 Computation of Lyapunov exponent

The Lyapunov exponent for the dynamical system in (11) is difficult to compute as we have a bivariate *stochastic* system whose variance is part of the state vector. We can write

$$\begin{array}{l} y_t = \sum_{g=1}^{G_1} \alpha_g \varphi(\mathbf{x}_t' \boldsymbol{\beta}_g + \sum_{l=1}^{l_2} \gamma_{gl} \sigma_{t-lL_2}^2) + \sigma_t \xi_{t1},\ \xi_{t1} \sim N(0,1), t = 1, ..., T, \\ \log \sigma_t^2 = \sum_{g=1}^{G} \delta_g \varphi(\mathbf{x}_t' \boldsymbol{\zeta}_g + \sum_{l=1}^{l_2} \psi_{gl} \sigma_{t-lL_2}^2) + \omega \xi_{t2},\ \xi_{t2} \sim N(0,1), t = 1, ..., T. \end{array} \tag{12}$$

The advantage is that σ_t appears explicitly in the first equation. In section 3 we show how likelihood-based inference can be performed for this model. Clearly, however, we cannot ignore ξ_t as in existing approaches to computing the Lyapunov exponent(s).

Our proposal is the following.

1. For given parameter values, simulate two values $\boldsymbol{\xi}^* = [\xi_1^*, \xi_2^*]' \sim iidN(0,1)$. Set $\xi_{t1} = \xi_1^*$ and $\xi_{t2} = \xi_2^*$. In this way, when $\overset{*}{2} = 0$ (12) is converted to a deterministic multivariate nonlinear state space model.

2. Compute the *conditional Lyapunov exponent*:

$$\lambda(\boldsymbol{\xi}^*) = \lim_{n \to \infty} n^{-1} \sum_{i=0}^{n-1} \log ||\partial \boldsymbol{\Phi}(\boldsymbol{\chi}_i) / \partial \boldsymbol{\chi}_i||, \tag{13}$$

where $\boldsymbol{\chi} = \left[y_{t-1}, ...y_{t-m_1 L_1}, \sigma_{t-1}^2, ..., \sigma_{t-m_2 L_2}^2\right]$, $\boldsymbol{\Phi}(\boldsymbol{\chi})$ represents the right-hand-side of (11), and $||.||$ represents the absolute value of the determinant. The Lyapunov exponent is, again, computed as:

$$\lambda(\boldsymbol{\xi}^*) = \lim_{M \to \infty} \frac{1}{2M} \log \nu^*, \tag{14}$$

where ν^* is the largest eigenvalue of matrix $T_M' T_M$, with:

$$T_M = \prod_{t=1}^{M-1} J_{M-1}, \tag{15}$$

and $M \leq T$ with $M = T^{2/3}$, and $J_i = \partial \boldsymbol{\Phi}(\boldsymbol{\chi}_i) / \partial \boldsymbol{\chi}_i$.

3. Repeat this a large number of times and let Ξ^* be the set of simulated values for $\boldsymbol{\xi}^*$.

4. Finally, as a conservative measure take:

$$\lambda = \min_{\boldsymbol{\xi}^* \in \Xi^*} \lambda(\boldsymbol{\xi}^*). \tag{16}$$

If $\lambda > 0$ then we have noisy chaos but negative λ does not necessarily imply stability and we have to examine the issue in more detail. Since $\boldsymbol{\xi}^*$follows a bivariate standard normal distribution, simulation is not really necessary as we can take, say, S =20 different values for ξ_1^* and ξ_2^* in the interval (-3.5, 3.5) which contains almost all probability mass. We would end up with S^2=400 points at which to evaluate the conditional Lyapunov exponents but the computation is not excessive although we have to repeat it for each SMC draw that is, for every drawn value of the parameters. Of course, nothing precludes to plot $\lambda(\boldsymbol{\xi}^*)$ as a function of ξ_1^* or ξ_2^* or both. This allows us to examine the issue in more detail as there may be values of noise (ξ_1^* and ξ_2^*) that give different results for noisy chaotic versus noisy stable dynamics.

3 Bayesian inference

On the one hand, the complication arisingwhen choosing the parameter values is that low parameters may prevent the neural network from reasonably approximating the specifications. On the other hand, large parameters increase the computational complexity because of the number of coefficients to be estimated (BenSaïda, 2014).This is the reason why we chose to select the parameters based on the (normalized) marginal likelihoods adopted as a strategy. For a discussion of other strategies, see Nychka, et al. (1992), BenSaïda and Litimi (2013).

The likelihood function of the model is:

$$\begin{aligned}\mathcal{L}(\boldsymbol{\theta};\mathcal{Y}) = (2\pi\omega^2)^{-T/2} \int_{\Re_+^T} \prod_{t=1}^{T} (2\pi)^{-1/2} (\sigma_t^2)^{-1} \exp\left\{ -\frac{[y_t - \sum_{g=1}^{G_1} \alpha_g \varphi(\mathbf{x}_t' \boldsymbol{\beta}_g + \sum_{l=1}^{l_2} \gamma_{gl} \sigma_{t-lm_2}^2)]^2}{2\sigma_t^2} \right\} \cdot, \\ \exp\left\{ -\frac{(\log \sigma_t^2 - \sum_{g=1}^{G} \delta_g \varphi(\mathbf{x}_t' \boldsymbol{\zeta}_g + \sum_{l=1}^{l_2} \psi_{gl} \sigma_{t-lm_2}^2))^2}{2\omega^2} \right\} d\boldsymbol{\sigma}^2,\end{aligned} \tag{17}$$

where the parameter vector $\boldsymbol{\theta} = \left[\boldsymbol{\alpha}', \boldsymbol{\beta}', \boldsymbol{\gamma}', \boldsymbol{\delta}', \boldsymbol{\psi}', \omega\right]' \subseteq \Re^p$, $\boldsymbol{\sigma}^2 = [\sigma_1^2, ..., \sigma_T^2]$, and $\mathcal{Y} = [y_1, ..., y_T]$. If $\lambda \geq 0$ then we have (noisy) chaotic dynamics, under a rich structure for the conditional variance, which has been often ignored in practice. We use a flat prior on all parameters:

$$p(\boldsymbol{\theta}) \propto \omega^{-1}. \tag{18}$$

In addition, we have to determine L, L_2, l_1, l_2, m_1, m_2 and G_1, G_2. We make the simplifying assumptions $L_1 = L_2 = L$, $l_1 = l_2 = l$, $m_1 = m_2 = m$ and $G_1 = G_2 = G$. The posterior distribution is:

$$p(\boldsymbol{\theta}|\mathcal{Y}) \propto \mathcal{L}(\boldsymbol{\theta};\mathcal{Y}) \cdot p(\boldsymbol{\theta}). \tag{19}$$

To perform Bayesian analysis we use a Sequential Monte Carlo / Particle Filtering (SMC/PF) algorithm. See Appendix for technical details. We determine L and G using the maximal value of the marginal likelihood or evidence:

$$\mathfrak{M}(\mathcal{Y}) = \int_{\Re^p} \mathcal{L}(\boldsymbol{\theta};\mathcal{Y}) \cdot p(\boldsymbol{\theta}) d\boldsymbol{\theta}. \tag{20}$$

The marginal likelihood is a byproduct of SMC (see Technical Appendix).

4 Empirical results

The data employed consist of stock indices for six countries: USA, UK, Switzerland, Netherlands, Germany and France (S&P 500, FTSE 100, SMI, AEX, DAX and CAC 40), and their corresponding implied volatility indices: VIX, VFTSE, VSMI, VAEX, VDAX-NEW, and VCAC. The volatility indices have thirty days to maturity and reflect the volatility of the respective stock markets. All data are extracted from Bloomberg. The sample covers the period from 15th May 2003 to 25th November 2014 in order to study the financial crisis.

Our empirical results are summarized in Figure 1 where we present the marginal posterior distributions of λ in the system (11).

Normalized marginal likelihoods for selection of G, L and n are presented in Figure 2. We normalize to 1.0 the value of $\mathfrak{M}(\mathcal{Y})$ when $G = 1$, $L = 1$ or n is equal to its minimal value (10). These figures are drawn (for example) for G assuming L, m and n are at their optimal values.

In Figure 3 we present marginal posterior distributions of λ in (1) or (9) with $\sigma_t = \sigma$. The marginal posterior distributions of λ for the squared residuals from (1) -volatility are presented in Figure 4.

In Figures 5 and 6 we provide the conditional Lyapunov exponent, defined in (13), for the U.S and the U.K, respectively.

5 Discussion

As can be seen in Figure 1, the estimated Lyapunov exponent, based on the marginal posterior distributions, is positive for the estimated system in (11) for all the economies examined in the sample, before and after crisis, which indicates presence of chaos in these series. In fact, in some economies, such as Switzerland or Germany, the estimated Lyapunov exponent is positive and larger in the after-crisis period, a finding which deserves careful screening by policy circles. Also, the multimodality often observed in the posterior distributions of various countries such as France before the crisis, could be seen as an expression of the locally unstable character of the Lyapunov exponent mirroring the economic and financial situation in those countries.

In the meantime, when adopting a simpistic approach where returns and volatility are not considered jointly, the model does not support the presence of chaotic dynamics and especially the presence of noisy volatility. We call this phenomenon "neglected chaos" and we believe that it could have important implications for the development of an early warning device. Also, of note is that the multimodality comment made earlier in the joint estimation, is in force here for the volatility case.

Furthermore, as we have seen, when the number of hidden layers G of the Neural Networks employed gets larger, the network can approximate any smooth non-linearity. We can see that the number of hidden layers before the crisis is larger than the one after the crisis, implying that more layers were needed to capture the chaotic behavior of the economy before the crisis. This finding could be seen as an expression of the unstable character of the economic and financial systems, before the global crisis. Finally, when the character of each equation in the joint system is taken into separate consideration we map the conditional Lyapunov exponent for all the vectors (1*, 2*). We can visually infer that for 2*=0, and for all the values of 1, the behaviour is chaotic before and after the crisis, both for the returns and for the volatilities confirming our previous finding.

In conclusion, our findings show strong evidence that for all the economies investigated the joint return-volatility dynamical system is chaotic both before and after the crisis. Put differently, the dynamics in these series, have not changed significantly after the crisis, and still remain chaotic

6 Conclusions

In this article, we put forward a novel Bayesian preocedure to compute the Lyapunov exponent jointly in the dynamical system of returns and volatility, since its direct computation, based on the already existing methods, was impossible as the stochastic nature had to be taken into consideration. To this end, we applied the new approach to daily stock return data for a selected group of six world countries, namely USA, UK, Switzerland, Netherlands, Germany and France, from 2003 to 2014. Sequential Monte Carlo techniques have been employed for Bayesian inference. Our findings show clearly that there is indeed noisy chaos both before and after the recent financial crisis. However, interestingly enough, when a much simpler model is examined where the interaction between returns and volatility is not taken into consideration jointly following common wisedom, the hypothesis of chaotic dynamics does not receive much support by the data, a situation that we call "neglected chaos" and could have important policy implications for the development of an early warning mechanism.

TECHNICAL APPENDIX

Particle filtering

The particle filter methodology can be applied to state space models of the general form:

$$y_T \sim p(y_t|x_t),\ s_t \sim p(s_t|s_{t-1}), \tag{A.1}$$

where s_t is a state variable. For general introductions see Gordon (1997), Gordon et al. (1993), Doucet et al (2001) and Ristic et al. (2004).

Given the data Y_t the posterior distribution $p(s_t|Y_t)$ can be approximated by a set of (auxiliary) particles $\left\{s_t^{(i)}, i = 1, ..., .N\right\}$ with probability weights $\left\{w_t^{(i)}, i = 1, ..., N\right\}$ where $\sum_{i=1}^N w_t^{(i)} = 1$. The predictive density can be approximated by:

$$p(s_{t+1}|Y_t) = \int p(s_{t+1}|s_t)p(s_t|Y_t)ds_t \simeq \sum_{i=1}^N p(s_{t+1}|s_t^{(i)})w_t^{(i)}, \tag{A.2}$$

and the final approximation for the filtering density is:

$$p(s_{t+1}|Y_t) \propto p(y_{t+1}|s_{t+1})p(s_{t+1}|Y_t) \simeq p(y_{t+1}|s_{t+1}) \sum_{i=1}^N p(s_{t+1}|s_t^{(i)})w_t^{(i)}. \tag{A.3}$$

The basic mechanism of particle filtering rests on propagating $\left\{s_t^{(i)}, w_t^{(i)}, i = 1, \ldots, N\right\}$ to the next step, viz. $\left\{s_{t+1}^{(i)}, w_{t+1}^{(i)}, i = 1, \ldots, N\right\}$ but this often suffers from the weight degeneracy problem. If parameters $\theta \in \Theta \in \Re^k$ are available, as is often the case, we follow Liu and West (2001) parameter learning takes place via a mixture of multivariate normals:

$$p(\theta|Y_t) \simeq \sum_{i=1}^N w_t^{(i)} N(\theta|a\theta_t^{(i)} + (1-a)\bar{\theta}_t, b^2 V_t), \tag{A.4}$$

where $\bar{\theta}_t = \sum_{i=1}^N w_t^{(i)}\theta_t^{(i)}$, and $V_t = \sum_{i=1}^N w_t^{(i)}(\theta_t^{(i)} - \bar{\theta}_t)(\theta_t^{(i)} - \bar{\theta}_t)'$. The constants a and b are related to shrinkage and are determined via a discount factor $\delta \in (0, 1)$ as $a = (1 - b^2)^{1/2}$ and $b^2 = 1 - [(3\delta - 1)/2\delta]^2$. See also Casarin and Marin (2007).

Andrieu and Roberts (2009), Flury and Shephard (2011) and Pitt et al. (2012) provide the Particle Metropolis-Hastings (PMCMC) technique which uses an unbiased estimator of the likelihood function $\hat{p}_N(Y|\theta)$ as $p(Y|\theta)$ is often not available in closed form.

Given the current state of the parameter $\theta^{(j)}$ and the current estimate of the likelihood, say $L^j = \hat{p}_N(Y|\theta^{(j)})$, a candidate θ^c is drawn from $q(\theta^c|\theta^{(j)})$ yielding $L^c = \hat{p}_N(Y|\theta^c)$. Then, we set $\theta^{(j+1)} = \theta^c$ with the Metropolis - Hastings probability:

$$A = \min\left\{1,\ \frac{p(\theta^c)L^c}{p(\theta^{(j)}L^j}\frac{q(\theta^{(j)}|\theta^c}{q(\theta^c|\theta^{(j)})}\right\}, \tag{A.5}$$

otherwise we repeat the current draws: $\left\{\theta^{(j+1)}, L^{j+1}\right\} = \left\{\theta^{(j)}, L^j\right\}$.

Hall, Pitt and Kohn (2014) propose an auxiliary particle filter which rests upon the idea that adaptive particle filtering (Pitt et al., 2012) used within PMCMC requires far fewer particles that the standard particle filtering algorithm to approximate $p(Y|\theta)$. From Pitt and Shephard (1999) we know that auxiliary particle filtering can be implemented easily once we can evaluate the state transition density $p(s_t|s_{t-1})$. When this is not possible, Hall, Pitt and Kohn (2014) present a new approach when, for instance, $s_t = g(s_{t-1}, u_t)$ for a certain disturbance. In this case we have:

$$p(y_t|s_{t-1}) = \int p(y_t|s_t)p(s_t|s_{t-1})ds_t, \tag{A.6}$$

$$p(u_t|s_{t-1};y_t) = p(y_t|s_{t-1},u_t)p(u_t|s_{t-1})/p(y_t|s_{t-1}). \tag{A.7}$$

If one can evaluate $p(y_t|s_{t-1})$ and simulate from $p(u_t|s_{t-1};y_t)$ the filter would be fully adaptable (Pitt and Shephard, 1999). One can use a Gaussian approximation for the first-stage proposal $g(y_t|s_{t-1})$ by matching the first two moments of $p(y_t|s_{t-1})$. So in some way we find that the approximating density $p(y_t|s_{t-1}) = N\left(\mathbb{E}(y_t|s_{t-1}), \mathbb{V}(y_t|s_{t-1})\right)$. In the second stage, we know that $p(u_t|y_t,s_{t-1}) \propto p(y_t|s_{t-1},u_t)p(u_t)$. For $p(u_t|y_t,s_{t-1})$ we know it is multimodal so suppose it has M modes are $\hat{u}_t^m$, for $m = 1,\ldots,M$. For each mode we can use a Laplace approximation. Let $l(u_t) = log\left[p(y_t|s_{t-1},u_t)p(u_t)\right]$. From the Laplace approximation we obtain:

$$l(u_t) \simeq l(\hat{u}_t^m) + \tfrac{1}{2}(u_t - \hat{u}_t^m)'\nabla^2 l(\hat{u}_t^m)(u_t - \hat{u}_t^m). \tag{A.8}$$

Then we can construct a mixture approximation:

$$g(u_t|x_t,s_{t-1}) = \sum_{m=1}^{M} \lambda_m (2\pi)^{-d/2}|\Sigma_m|^{-1/2} \exp\left\{\tfrac{1}{2}(u_t - \hat{u}_t^m)'\Sigma_m^{-1}(u_t - \hat{u}_t^m\right\}, \tag{A.9}$$

where $\Sigma_m = -\left[\nabla^2 l(\hat{u}_t^m)\right]^{-1}$ and $\lambda_m \propto \exp\left\{l(u_t^m)\right\}$ with $\sum_{m=1}^{M} = 1$. This is done for each particle s_t^i. This is known as the Auxiliary Disturbance Particle Filter (ADPF).

An alternative is the independent particle filter (IPF) of Lin et al. (2005). The IPF forms a proposal for s_t directly from the measurement density $p(y_t|s_t)$ although Hall, Pitt and Kohn (2014) are quite right in pointing out that the state equation can be very informative.

In the standard particle filter of Gordon et al. (1993) particles are simulated through the state density $p(s_t^i|s_{t-1}^i)$ and they are re-sampled with weights determined by the measurement density evaluated at the resulting particle, viz. $p(y_t|s_t^i)$.

The ADPF is simple to construct and rests upon the following steps:

For $t = 0,\ldots,T-1$ given samples $s_t^k \sim p(s_t|Y_{1:t})$ with mass π_t^k for $k = 1,...,N$.

1) For $k = 1,\ldots,N$ compute $\omega_{t|t+1}^k = g(y_{t+1}|s_t^k)\pi_t^k$, $\pi_{t|t+1}^k = \omega_{t|t+1}^k / \sum_{i=1}^{N} \omega_{t|t+1}^i$.

2) For $k = 1,\ldots,N$ draw $\tilde{s}_t^k \sim \sum_{i=1}^{N} \pi_{t|t+1}^i \delta_{s_t}^i(ds_t)$.

3) For $k = 1,\ldots,N$ draw $u_{t+1}^k \sim g(u_{t+1}|\tilde{s}_t^k, y_{t+1})$ and set $s_{t+1}^k = h(s_t^k; u_{t+1}^k)$.

4) For $k = 1,\ldots,N$ compute

$$\omega_{t+1}^k = \frac{p(y_{t+1}|s_{t+1}^k)p(u_{t+1}^k)}{g(y_{t+1}|s_t^k)g(u_{t+1}^k|\tilde{s}_t^k, y_{t+1})}, \pi_{t+1}^k = \frac{\omega_{t+1}^k}{\sum_{i=1}^{N}\omega_{t+1}^i}. \tag{A.10}$$

It should be mentioned that the estimate of likelihood from ADPF is:

$$p(Y_{1:T}) = \prod_{t=1}^{T}\left(\sum_{i=1}^{N}\omega_{t-1|t}^i\right)\left(N^{-1}\sum_{i=1}^{N}\omega_t^i\right). \tag{A.11}$$

Particle Metropolis adjusted Langevin filters

Nemeth, Sherlock and Fearnhead (2014) provide a particle version of a Metropolis adjusted Langevin algorithm (MALA). In Sequential Monte Carlo we are interested in approximating $p(s_t|Y_{1:t},\theta)$. Given that:

$$p(s_t|Y_{1:t},\theta) \propto g(y_t|x_t,\theta)\int f(s_t|s_{t-1},\theta)p(s_{t-1}|y_{1:t-1},\theta)ds_{t-1}, \tag{A.12}$$

where $p(s_{t-1}|y_{1:t-1},\theta)$ is the posterior as of time $t-1$. If at time $t-1$ we have a set set of particles $\left\{s_{t-1}^i, i = 1,\ldots,N\right\}$ and weights $\left\{w_{t-1}^i, i = 1,\ldots.N\right\}$ which form a discrete approximation for $p(s_{t-1}|y_{1:t-1},\theta)$ then we have the ap-

proximation:

$$\hat{p}(s_{t-1}|y_{1:t-1},\theta) \propto \sum_{i=1}^{N} w_{t-1}^{i} f(s_t|s_{t-1}^{i},\theta). \tag{A.13}$$

See Andrieu et al. (2010) and Cappe at al. (2005) for reviews. From (A.13) Fernhead (2007) makes the important observation that the joint probability of sampling particle s_{t-1}^{i} and state s_t is:

$$\omega_t = \frac{w_{t-1}^{i} g(y_t|s_t,\theta) f(s|s_{t-1}^{i},\theta)}{\xi_t^{i} q(s_t|s_{t-1}^{i},y_t,\theta)}, \tag{A.14}$$

where $q(s_t|s_{t-1}^{i},y_t,\theta)$ is a density function amenable to simulation and

$$\xi_t^{i} q(s_t|s_{t-1}^{i},y_t,\theta) \simeq c g(y_t|s_t,\theta) f(s_t|s_{t-1}^{i},\theta), \tag{A.15}$$

and c is the normalizing constant in (A.12).

In the MALA algorithm of Roberts and Rosenthal (1998)[1] we form a proposal:

$$\theta^{c} = \theta^{(s)} + \lambda z + \tfrac{\lambda^2}{2} \nabla \log p(\theta^{(s)}|Y_{1:T}), \tag{A.16}$$

where $z \sim N(0,I)$ which should result in larger jumps and better mixing properties, plus lower autocorrelations for a certain scale parameter λ. Acceptance probabilities are:

$$a(\theta^{c}|\theta^{(s)}) = \min\left\{1, \frac{p(Y_{1:T}|\theta^{c}) q(\theta^{(s)}|\theta^{c})}{p(Y_{1:T}|\theta^{(s)}) q(\theta^{c}|\theta^{(s)})}\right\}. \tag{A.17}$$

Using particle filtering it is possible to create an approximation of the score vector using Fisher's identity:

$$\nabla \log p(Y_{1:T}|\theta) = E\left[\nabla \log p(s_{1:T},Y_{1:T}|\theta)|Y_{1:T},\theta\right], \tag{A.18}$$

which corresponds to the expectation of:

$$\nabla \log p(s_{1:T},Y_{1:T}|\theta) = \nabla \log p(|s_{1:T-1},Y_{1:T-1}|\theta) + \nabla \log g(y_T|s_T,\theta) + \nabla \log f(s_T|s|_{T-1},\theta),$$

over the path $s_{1:T}$. The particle approximation to the score vector results from replacing $p(s_{1:T}|Y_{1:T},\theta)$ with a particle approximation $\hat{p}(s_{1:T}|Y_{1:T},\theta)$. With particle i at time t-1 we can associate a value $\alpha_{t-1}^{i} = \nabla \log p(s_{1:t-1}^{i},Y_{1:t-1}|\theta)$ which can be updated recursively. As we sample κ_i in the APF (the index of particle at time $t-1$ that is propagated to produce the ith particle at time t) we have the update:

$$\alpha_t^{i} = a_{t-1}^{\kappa_i} + \nabla \log g(y_t|s_t^{i},\theta) + \nabla \log f(s_t^{i}|s_{t-1}^{i},\theta). \tag{A.19}$$

To avoid problems with increasing variance of the score estimate $\nabla \log p(Y_{1:t}|\theta)$ we can use the approximation:

$$\alpha_{t-1}^{i} \sim N(m_{t-1}^{i}, V_{t-1}). \tag{A.20}$$

[1]The benefit of MALA over Random-Walk-Metropolis arises when the number of parameters n is large. This happens because the scaling parameter λ is $O(n^{-1/2})$for Random-Walk-Metropolis but it is $O(n^{-1/6})$ for MALA, see Roberts et al. (1997) and Roberts and Rosenthal (1998)

The mean is obtained by shrinking α_{t-1}^i towards the mean of α_{t-1} as follows:

$$m_{t-1}^i = \delta\alpha_{t-1}^i + (1-\delta)\sum_{i=1}^{N} w_{t-1}^i \alpha_{t-1}^i, \tag{A.21}$$

where $\delta \in (0,1)$ is a shrinkage parameter. Using Rao-Blackwellization one can avoid sampling α_t^i and instead use the following recursion for the means:

$$m_t^i = \delta m_{t-1}^{\kappa_i} + (1-\delta)\sum_{i=1}^{N} w_{t-1}^i m_{t-1}^i + \nabla \log g(y_t|s_t^i,\theta) + \nabla \log f(s_t^i|s_{t-1}^{\kappa_i},\theta), \tag{A.22}$$

which yields the final score estimate:

$$\nabla \log \hat{p}(Y_{1:t}|\theta) = \sum_{i=1}^{N} w_t^i m_t^i. \tag{A.23}$$

As a rule of thumb Nemeth, Sherlock and Fearnhead (2014) suggest taking $\delta = 0.95$. Furthermore, they show the important result that the algorithm should be tuned to the asymptotically optimal acceptance rate of 15.47% and the number of particles must be selected so that the variance of the estimated log-posterior is about 3. Additionally, if measures are not taken to control the error in the variance of the score vector, there is no gain over a simple random walk proposal.

Of course, the marginal likelihood is:

$$p(Y_{1:T}|\theta) = p(y_1|\theta)\prod_{t=2}^{T} p(y_t|Y_{1:t-1},\theta), \tag{A.24}$$

where

$$p(y_t|Y_{1:t-1},\theta) = \int g(y_t|s_t)\int f(s_t|s_{t-1},\theta)p(s_{t-1}|Y_{1:T-1},\theta)ds_{t-1}ds_t, \tag{A.25}$$

provides, in explicit form, the predictive likelihood.

References

Andrieu, C., and G.O. Roberts. (2009) The pseudo-marginal approach for efficient computation. Ann. Statist., 37, 697–725.

Andrieu, C., A. Doucet, and R. Holenstein (2010), "Particle Markov chain Monte Carlo methods," Journal of the Royal Statistical Society: Series B, 72, 269–342.

BenSaïda, A. (2014). Noisy chaos in intraday financial data: Evidence from the American index. Applied Mathematics and Computation 226, 258-265.

BenSaïda, A., H. Litimi (2013). High level chaos in the exchange and index markets, Chaos Solitons Fractals 54 (2013) 90–95.

D.O. Cajueiro, B.M. Tabak, Testing for long-range dependence in the Brazilian term structure of interest rates, Chaos Solitons Fractals 40 (2009) 1559–1573.

Cappé, O., E. Moulines, E., and T. Rydén (2005). Inference in Hidden Markov Models. Springer, Berlin.

Casarin, R., J.-M. Marin (2007). Online data processing: Comparison of Bayesian regularized particle filters. University of Brescia, Department of Economics. Working Paper n. 0704.

Catherine Kyrtsou, Michel Terraza, Stochastic chaos or ARCH effects in stock series?: A comparative study, International Review of Financial Analysis, Volume 11, Issue 4, 2002, Pages 407-431.

Gürsan Çoban, Ali H. Büyüklü, Deterministic flow in phase space of exchange rates: Evidence of chaos in filtered series of Turkish Lira–Dollar daily growth rates, Chaos, Solitons & Fractals, Volume 42, Issue 2, 30 October 2009, Pages 1062-1067.

Das, A., P. Das (2007). Chaotic analysis of the foreign exchange rates, Appl. Math. Comput. 185, 388–396.

Doucet, A., N. de Freitas, and N. Gordon (2001). Sequential Monte Carlo Methods in Practice. New York: Springer.

Eckmann, J. D. Ruelle (1985) Ergodic theory of chaos and strange attractors, Rev. Mod. Phys. 57 (3) (1985) 617–656.

Fearnhead, P. (2007). Computational methods for complex stochastic systems: a review of some alternatives to MCMC. Statistics and Computing, 18(2):151-171.

Flury, T., and N. Shephard, (2011) Bayesian inference based only on simulated likelihood: particle filter analysis of dynamic economic models. Econometric Theory 27, 933-956.

Gordon, N.J. (1997). A hybrid bootstrap filter for target tracking in clutter. IEEE Transactions on Aerospace and Electronic Systems 33: 353– 358.

Gordon, N.J., D.J. Salmond, and A.F.M. Smith (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. IEEE-Proceedings-F 140: 107–113.

Hall, J., M.K. Pitt, and R. Kohn (2014). Bayesian inference for nonlinear structural time series models. Journal of Econometrics 179 (2), 99–111.

Kurt Hornik, Maxwell Stinchcombe, Halbert White, Multilayer feedforward networks are universal approximators, Neural Networks, Volume 2, Issue 5, 1989, Pages 359-366

Kurt Hornik, Approximation capabilities of multilayer feedforward networks, Neural Networks, Volume 4, Issue 2, 1991, Pages 251-257,

Jacquier, Eric, Nicholas G. Polson, and Peter E. Rossi, 1994, Bayesian analysis of stochastic volatility models, Journal of Business and Economic Statistics 12, 371{417

Lahmiri, S (2015). Long memory in international financial markets trends and short movements during 2008 financial crisis based on variational mode decomposition and detrended fluctuation analysis, Physica A 437 130–138.

Lahmiri, S. (2017). A study on chaos in crude oil markets before and after 2008 international financial crisis. Psysica A 466, 389-395.

Liu, J., M. West (2001). Combined parameter and state estimation in simulation-based filtering. In: Doucet, A., de Freitas, N., Gordon, N. (Eds.), Sequential Monte Carlo Methods in Practice. Springer-Verlag.

Panayotis G. Michaelides, Efthymios G. Tsionas, Angelos T. Vouldis, Konstantinos N. Konstantakis (2015), Global approximation to arbitrary cost functions: A Bayesian approach with application to US banking, European Journal of Operational Research, Volume 241, Issue 1, Pages 148-160.

Ritesh Kumar Mishra, Sanjay Sehgal, N.R. Bhanumurthy, A search for long-range dependence and chaotic structure in Indian stock market, Review of Financial Economics, Volume 20, Issue 2, May 2011, Pages 96-104.

R. Morales, T. Di Matteo, R. Gramatica, T. Aste, Dynamical generalized Hurst exponent as a tool to monitor unstable periods in financial time series, Physica A 391 (2012) 3180–3189.

D.B. Nelson, Conditional heteroskedasticity in asset returns: A new approach, Econometrica 59 (1991) 347–370.

Nemeth, C., P. Fearnhead (2014). Particle Metropolis adjusted Langevin algorithms for state-space models. Pre-print arXiv:1402.0694v1.

D. Nychka, S. Ellner, R.A. Gallant, D. McCaffrey, Finding chaos in noisy system, J. R. Stat. Soc. Ser. B 54 (2) (1992) 399–426.

Pitt, M.K., R.S. Silva, P. Giordani, and R. Kohn (2012). On some properties of Markov chain Monte Carlo simulation methods based on the particle filter. J. Econom. 171(2), 134–151.

B. Ristic, S. Arulampalam, and N. Gordon (2004). Beyond Kalman Filters: Particle Filters for Applications. Norwood, MA: Artech House.

A. Sensoy, Effects of monetary policy on the long memory in interest rates: Evidence from an emerging market, Chaos Solitons Fractals 57 (2013) 85–88.

P. Yousefpoor, M.S. Esfahani, H. Nojumi, Looking for systematic approach to select chaos tests, Applied Mathematics and Computation, Volume 198, Issue 1, 15 April 2008, Pages 73-91

Figure 1: Marginal posterior distributions of λ

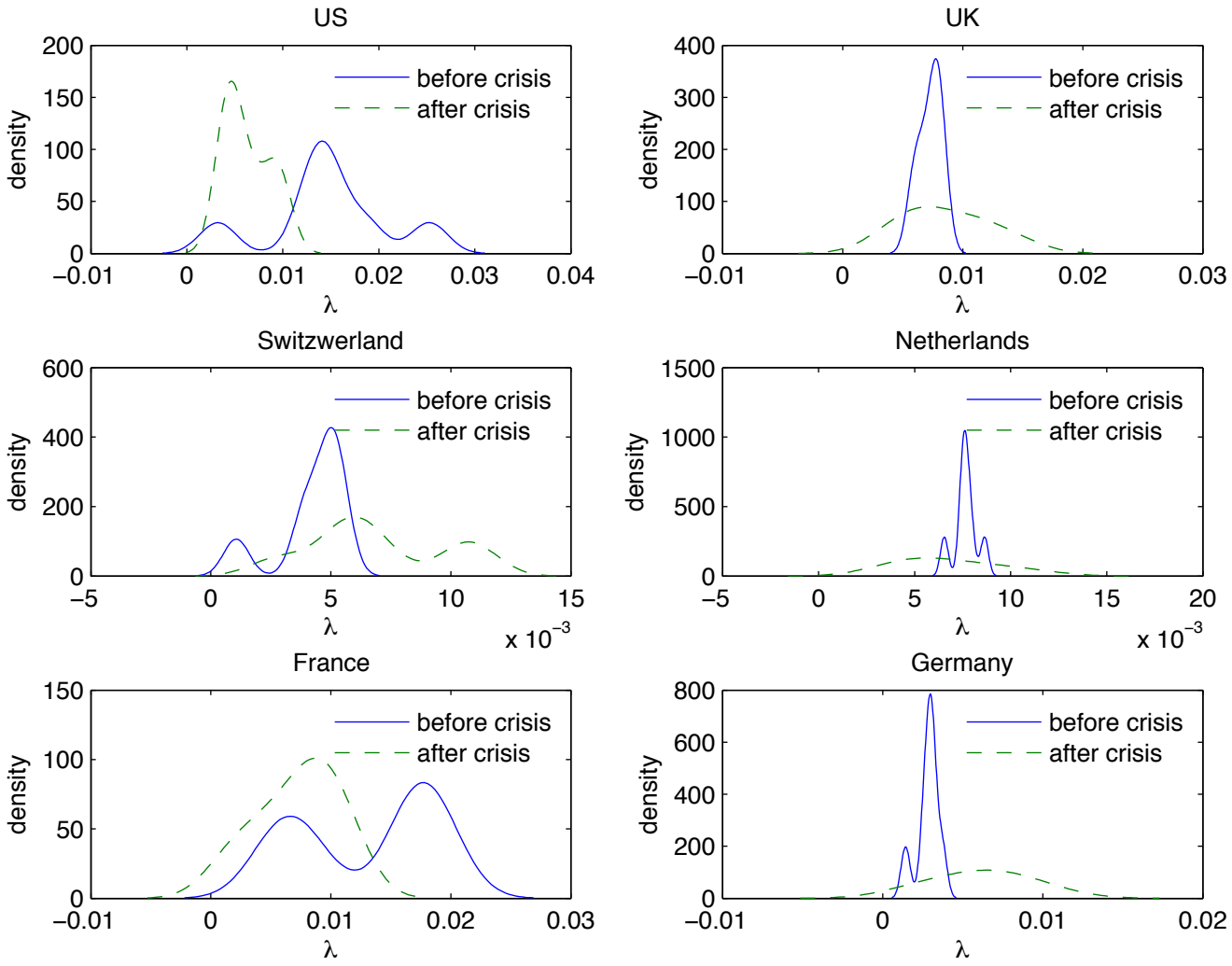

Figure 2: Lyapunov exponents $\lambda(\boldsymbol{\xi}^*)$ for U.S

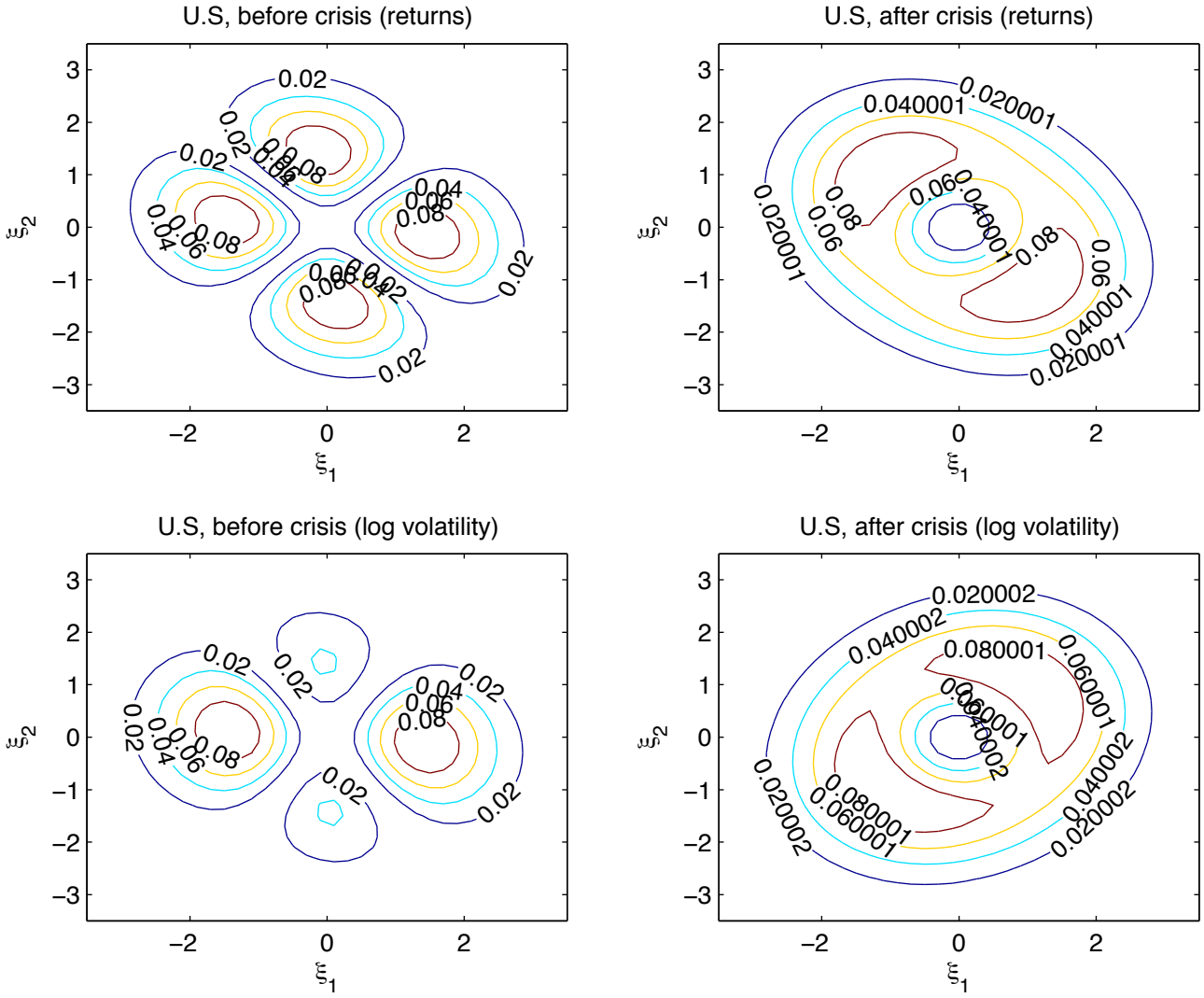

Figure 3: Lyapunov exponents $\lambda(\boldsymbol{\xi}^*)$ for U.K

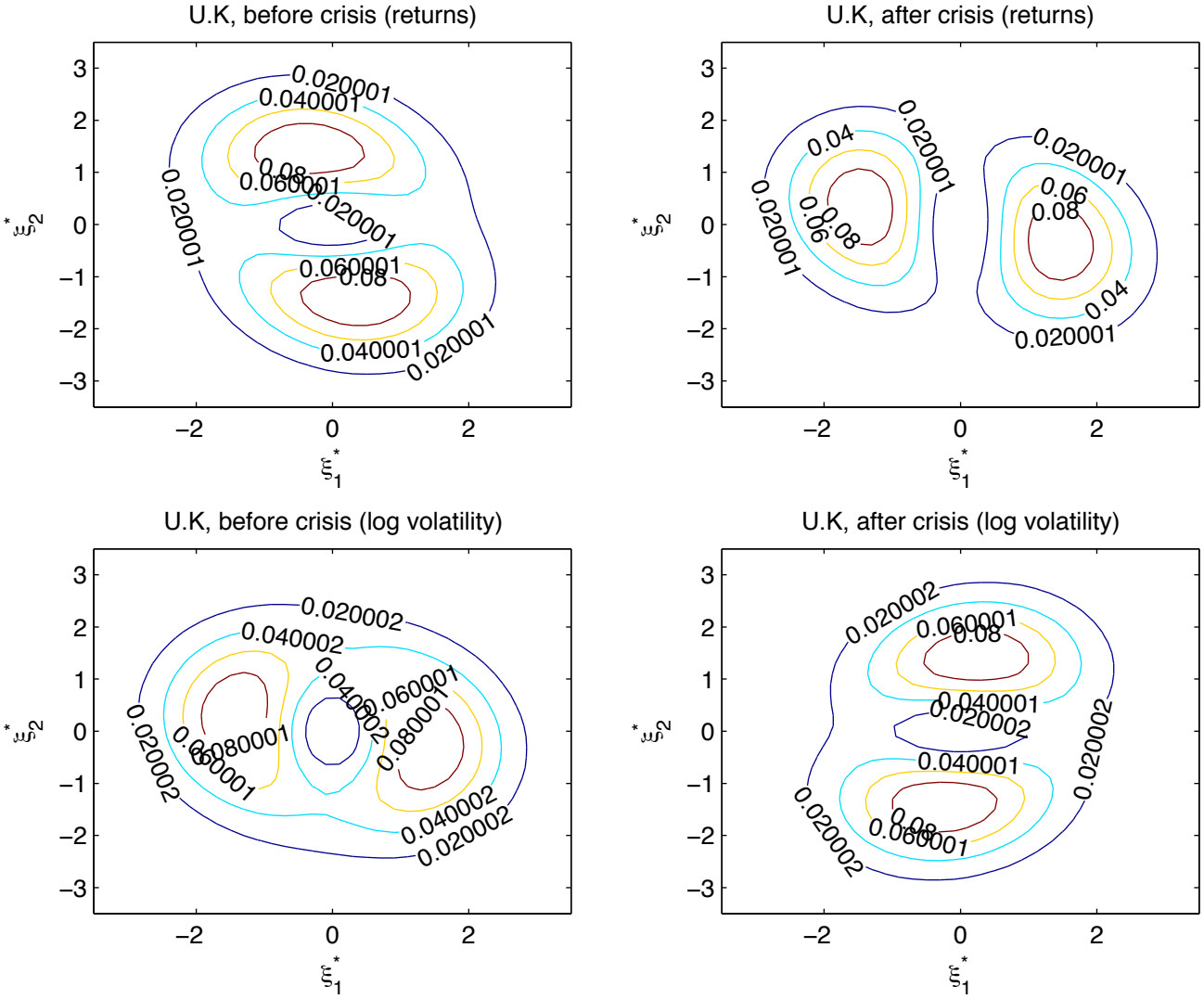

Figure 4: Normalized marginal likelihoods for selection of G,L, n and embedding dimension m

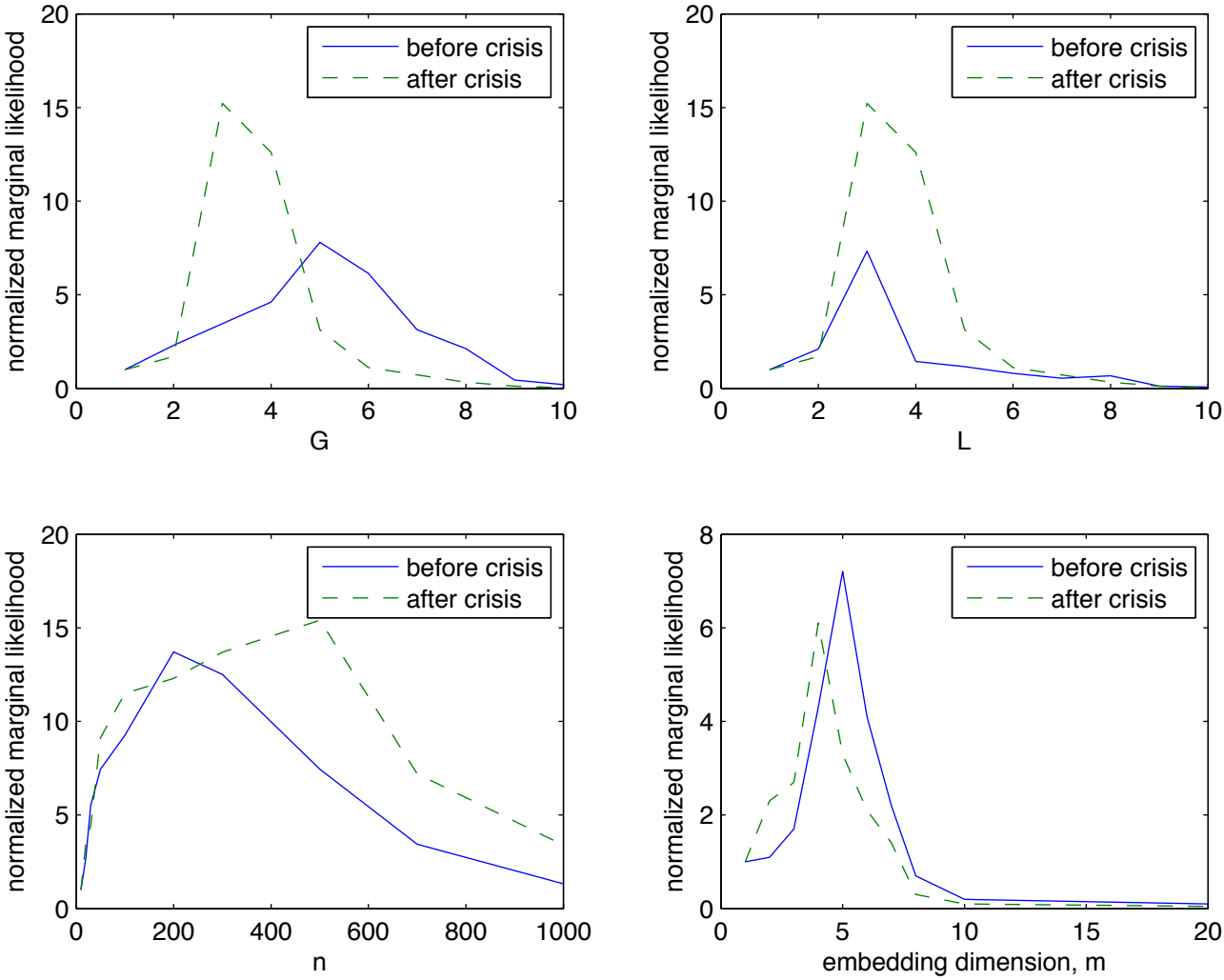

Figure 5: Marginal posteriors of λ in (1)

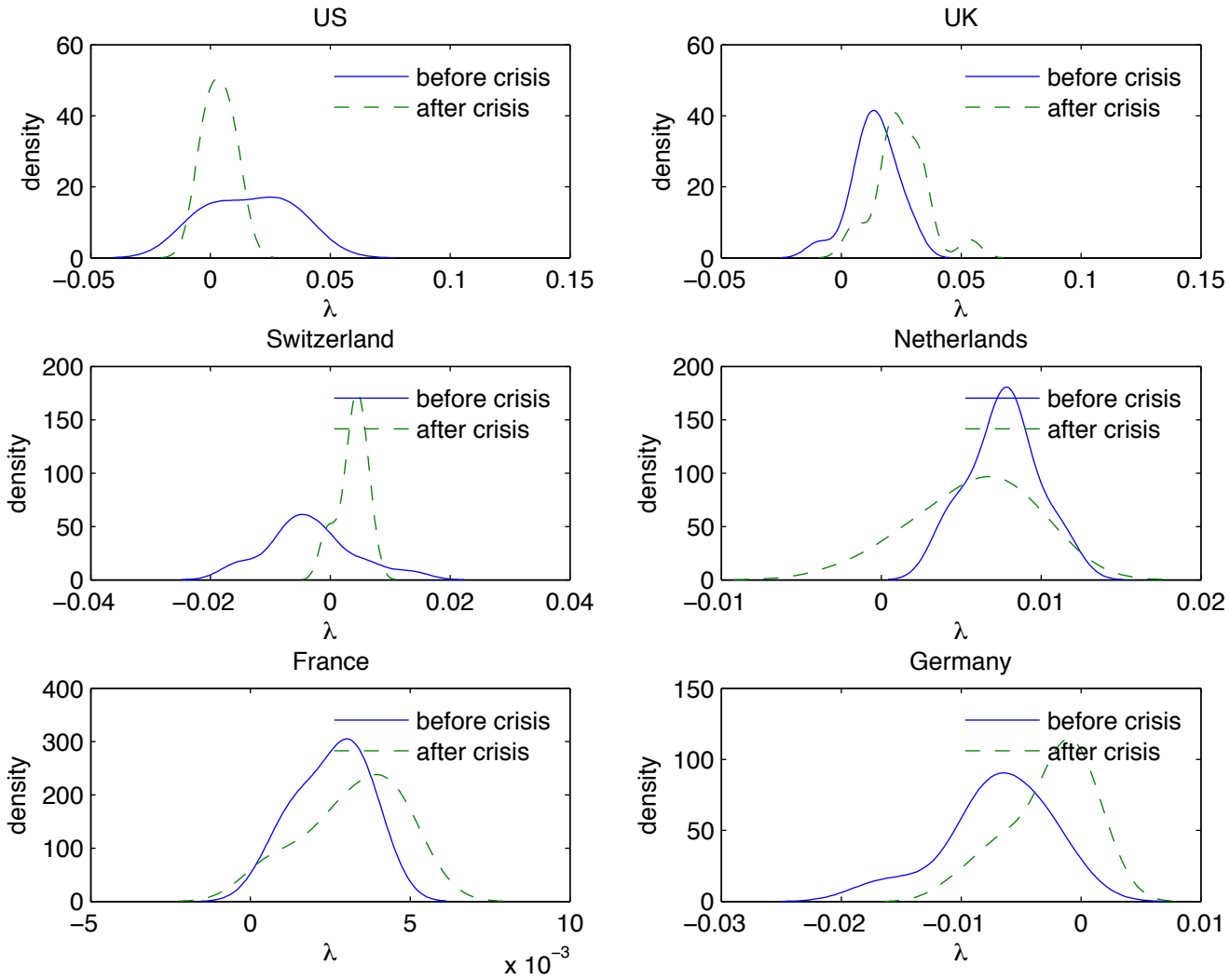

Figure 6: Marginal posteriors of λ for the volatility in (1)