

Estimation of the Markov-switching GARCH model by a Monte Carlo EM algorithm

Maciej Augustyniak

Fields Institute
February 3, 2012

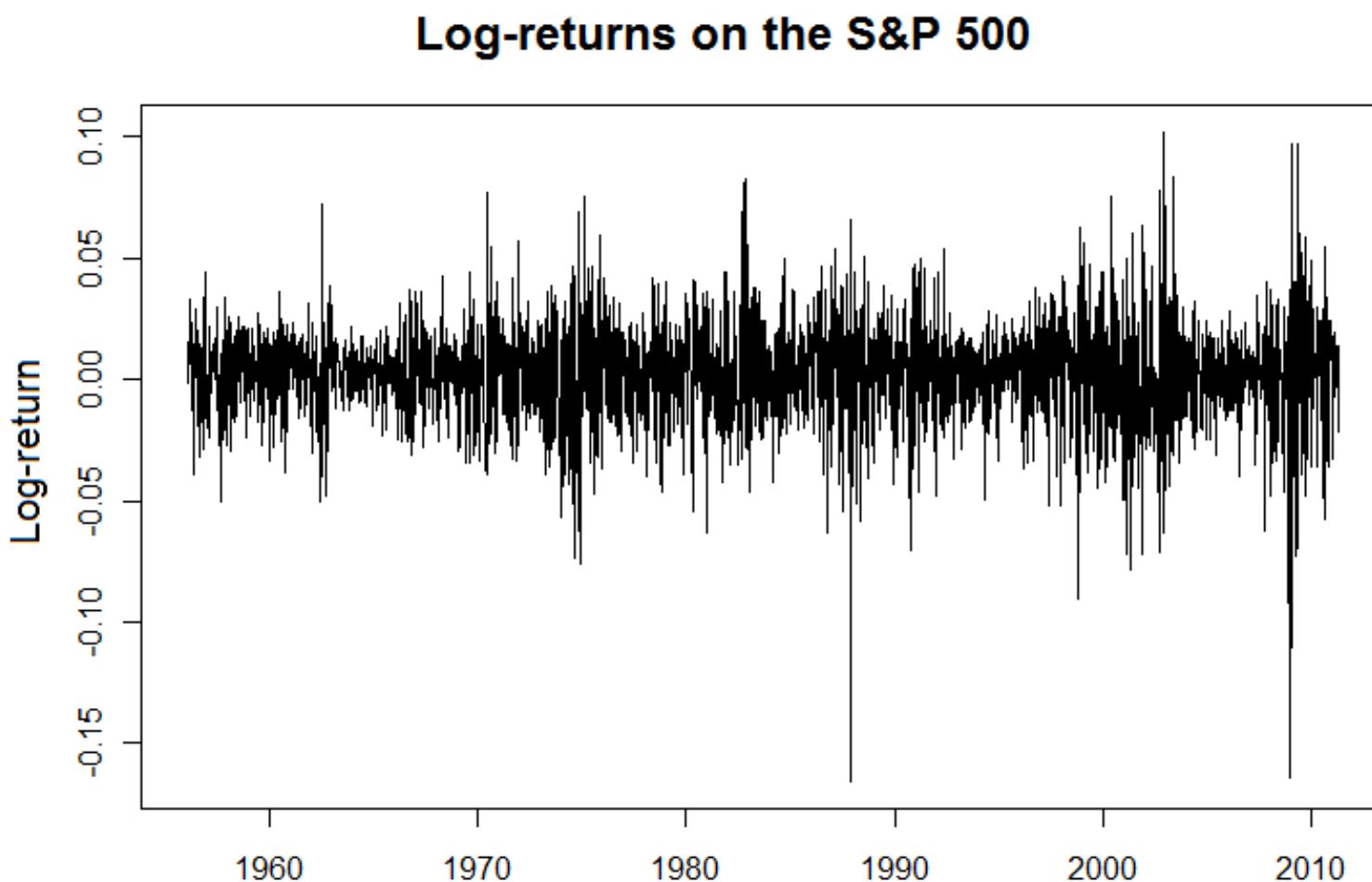
Agenda

- Stylized facts of financial data
- GARCH
- Regime-switching
- MS-GARCH
- Available estimation methods for MS-GARCH models
- EM algorithm and its stochastic variants
- Estimation algorithm for the MS-GARCH based on the Monte Carlo EM algorithm
- Simulation study

Stylized facts of financial data

- There is no (or very weak) correlation in returns
- However, the **square of the returns are highly correlated**; this implies a certain form of dependence between returns
- **Volatility clustering**: periods of high and low volatility
- **Heavy tails and negative skewness**
- **Leverage effect**: a large negative return has a bigger impact on future volatility than a large positive return
- **Jumps** in volatility and returns

Stylized facts of financial data



GARCH

- **GARCH(1,1)**

$$y_t = \mu + \sigma_t \eta_t$$

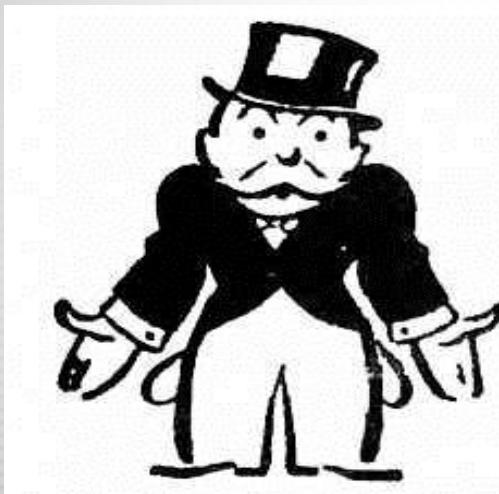
$$\sigma_t^2 = \omega + \alpha \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2$$

$$\epsilon_t = y_t - \mu$$

- Properties
 - Heavy tails
 - Volatility clustering
 - No correlation in returns but correlation in the squares

Regime-Switching

- In regime-switching (RS) models, the distribution generating returns depends on the (**unobservable**) state of the economy (also known as **regime**)



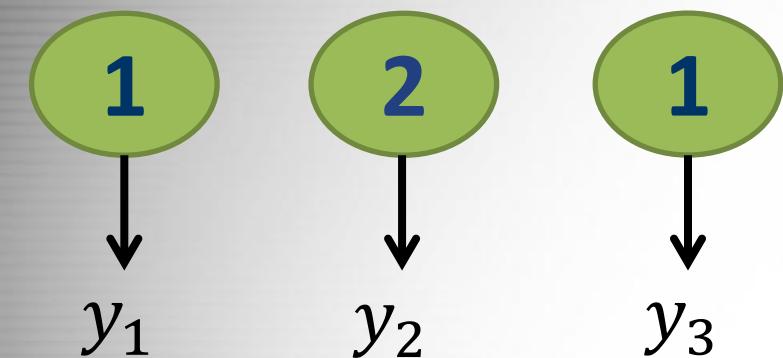
$N(-20\%, 26\%)$



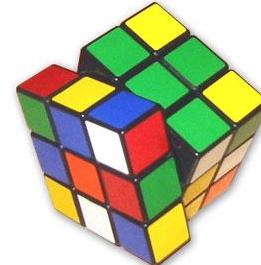
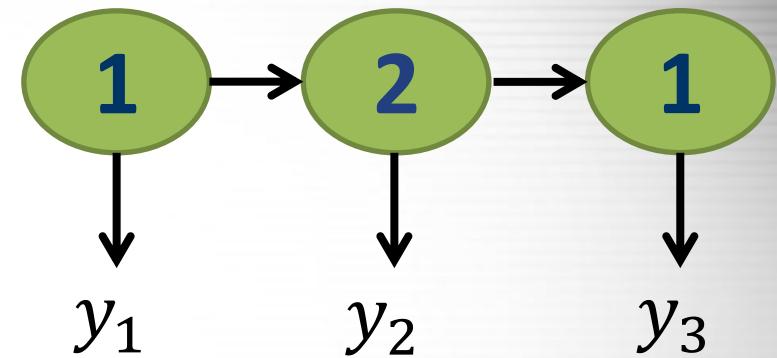
$N(17\%, 12\%)$ Université de Montréal

Regime-Switching RS

Mixture



$$\begin{pmatrix} 0.75 & 0.25 \\ 0.75 & 0.25 \end{pmatrix}$$



$$\begin{pmatrix} 0.95 & 0.05 \\ 0.20 & 0.80 \end{pmatrix}$$

Regime-Switching

- Estimation of RS models
 - Direct maximization of the log-likelihood:
Hamilton filter – Hamilton (1989)
 - EM algorithm: in the context of RS models, it also known as the Baum-Welch or forward-backward algorithm – Hamilton (1990) provides a slight generalization of that algorithm
 - Bayesian methods
- Alternative terms used for a RS model include hidden Markov model (HMM), hidden Markov process, Markov-dependent mixture and **Markov-switching (MS) model**

MS-GARCH

- A natural combination of a RS (or MS) model with a GARCH model is the following:

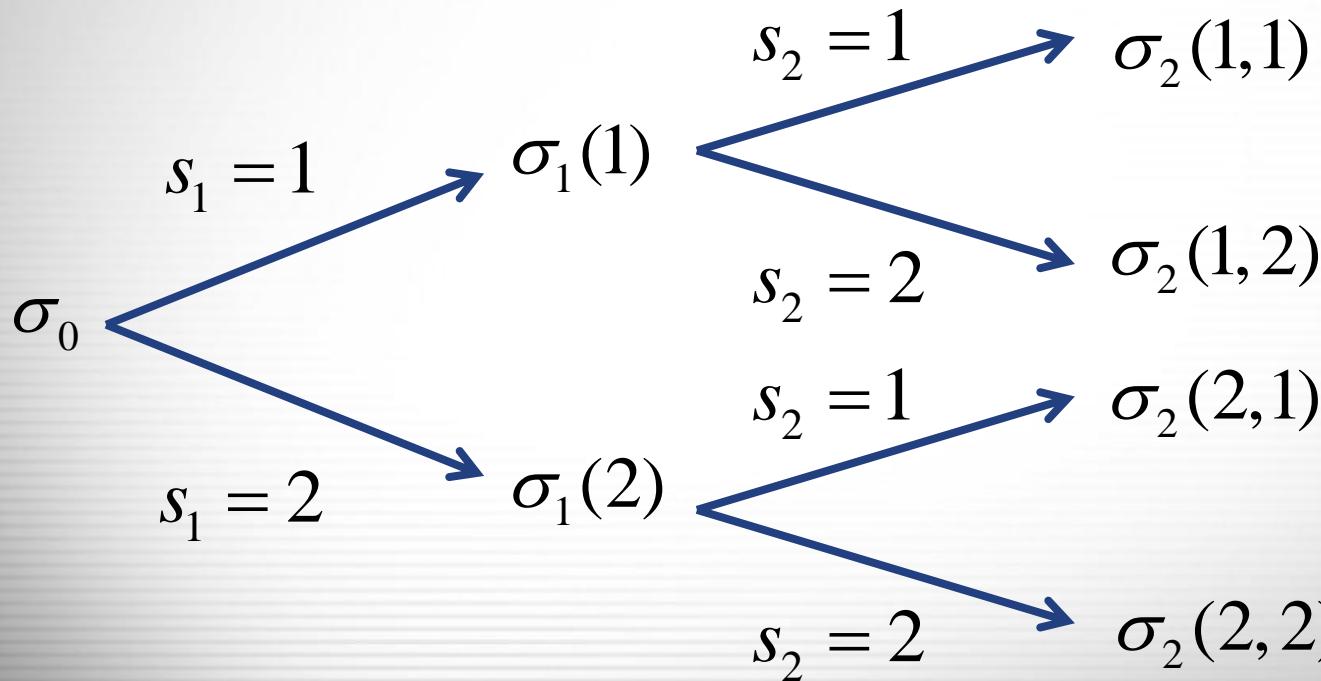
$$y_t = \mu_{s_t} + \sigma_t(s_{1:t})\eta_t$$

$$\sigma_t^2(s_{1:t}) = \omega_{s_t} + \alpha_{s_t} \epsilon_{t-1}^2(s_{t-1}) + \beta_{s_t} \sigma_{t-1}^2(s_{1:t-1})$$

$$\epsilon_{t-1}(s_{t-1}) = y_{t-1} - \mu_{s_{t-1}}$$

MS-GARCH

- The conditional distribution of each observation depends on the whole regime path
- **Path dependence problem**

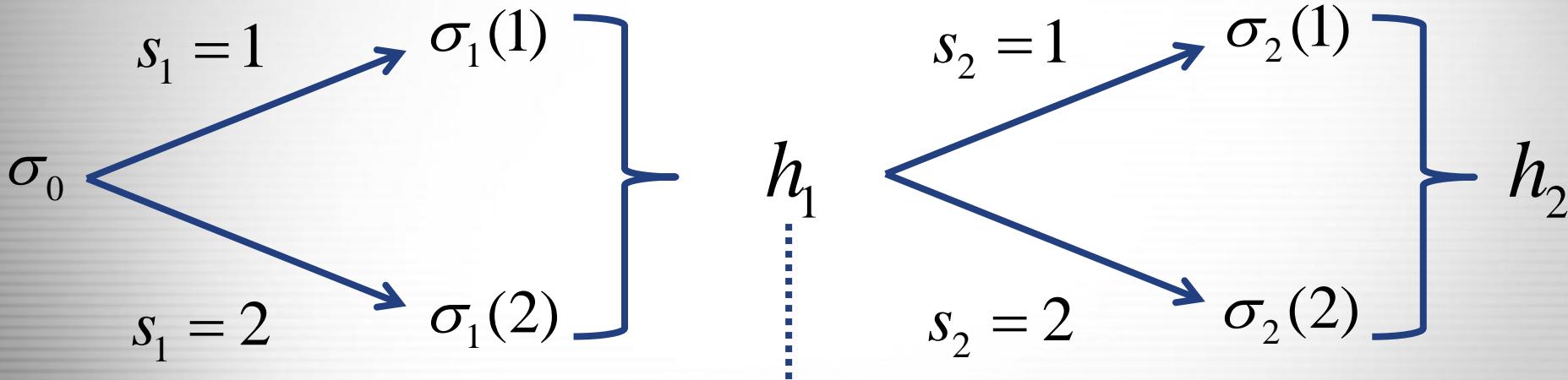


MS-GARCH

- MS-GARCH models are becoming increasingly popular to model financial data
- Due to this popularity, it is essential to develop efficient estimation techniques
- However, estimating these models is a difficult task because of the path dependence problem; **one has yet to propose a method to obtain the maximum likelihood estimator (MLE) of the model!**
- I will now present some methods that were developed in the past to estimate MS-GARCH models

Methods based on collapsing

- Hamilton and Susmel (1994) and Cai (1994) introduce MS-ARCH models that avoid path dependence
- First “MS-GARCH” model: **Gray (1996)**



$$h_t = \text{Var}(y_t \mid y_{t-1}, y_{t-2}, \dots)$$

GMM

- Francq and Zakoïan (2008) are the first to propose a method to estimate the MS-GARCH model **without resorting to a modification of the model**
- They estimate the model using the generalized method of moments (GMM); the path dependence problem is not encountered since the method does not rely on the likelihood
- Their technique relies on the availability of analytic expressions (derived by Francq and Zakoïan, 2005) for $E(y_t^2)$ and $E(y_t^{2m} y_{t-k}^{2m})$, $m \geq 1$ and $k \geq 0$
- Problems: identifiability, robustness and bias

Bayesian MCMC

- Bauwens, Preminger and Rombouts (2010) are among the first to estimate the MS-GARCH model using Bayesian MCMC techniques
- Data augmentation (Tanner et Wong, 1987)

Simulate $s_{1:T}$ conditional on θ and y

Simulate θ conditional on $y^{aug} = (s_{1:T}, y)$

Obtaining the MLE

- The GMM and the Bayesian MCMC offer ways to estimate the MS-GARCH model but one has yet to propose a method to find the MLE
- The estimation approaches that were introduced so far were generally justified by their respective authors with a statement that it is not possible to obtain the MLE because the path dependence problem renders computation of the likelihood infeasible in practice
- While it is true that the likelihood cannot be calculated exactly, this does not imply that the MLE cannot be obtained  **EM algorithm**

- EM Algorithm: Dempster et al. (1977)
- **Insight:** let $\ell(\theta)$ represent the log-likelihood

$$\begin{aligned}\ell(\theta) &= E[\log f(y, \textcolor{blue}{S} | \theta) | y, \textcolor{red}{\theta'}] - E[\log f(\textcolor{blue}{S} | y, \theta) | y, \textcolor{red}{\theta'}] \\ &= Q(\theta | \textcolor{red}{\theta'}) - H(\theta | \textcolor{red}{\theta'})\end{aligned}$$

- We wish to find a better value than θ' , i.e., we need

$$\begin{aligned}\ell(\theta) - \ell(\textcolor{red}{\theta'}) &= [Q(\theta | \textcolor{red}{\theta'}) - Q(\textcolor{red}{\theta'} | \textcolor{red}{\theta'})] \\ &\quad - \underbrace{[H(\theta | \textcolor{red}{\theta'}) - H(\textcolor{red}{\theta'} | \textcolor{red}{\theta'})]}_{\leq 0} > 0\end{aligned}$$

- E-Step

$$Q(\theta | \theta^{(r-1)}) = \int \log[f(y, S | \theta)] f(S | y, \theta^{(r-1)}) dS$$
$$\approx \underbrace{\frac{1}{m_r} \sum_{i=1}^{m_r} \log[f(y, S_i^{(r)} | \theta)]}_{\text{Monte Carlo E-Step}} = \hat{Q}_{m_r}(\theta | \theta^{(r-1)})$$

Monte Carlo E-Step (Wei and Tanner, 1990)

- M-Step

$$\theta^{(r)} = \arg \max_{\theta} Q(\theta | \theta^{(r-1)})$$

E-Step: Gibbs sampler

- How can we obtain draws from $f(\mathbf{S} \mid \mathbf{y}, \boldsymbol{\theta}^{(r-1)})$?
→ **Gibbs sampler (single-move)**
- Full conditional distribution

$$p(s_t \mid s_{1:t-1}^{(i)}, s_{t+1:T}^{(i-1)}, \mathbf{y}, \boldsymbol{\theta}^{(r-1)}) \propto p_{s_{t-1}^{(i)}, s_t} p_{s_t, s_{t+1}^{(i-1)}} \prod_{j=t}^T \sigma_j^{-1} g\left(\frac{y_j - \mu_{s_j}}{\sigma_j}\right)$$

	$s_1^{(i)}$ $s_2^{(i)}$ $s_3^{(i)}$... $s_T^{(i)}$ $s_1^{(i+1)}$ $s_2^{(i)}$ $s_3^{(i)}$... $s_T^{(i)}$ $s_1^{(i+1)}$ $s_2^{(i+1)}$ $s_3^{(i)}$... $s_T^{(i)}$ $s_1^{(i+1)}$ $s_2^{(i+1)}$ $s_3^{(i+1)}$... $s_T^{(i)}$	 $(\mathbf{y}, \boldsymbol{\theta}^{(r-1)})$
--	--	---

M-Step

- The M-Step is straightforward and requires less computational time than the E-Step
- It can be split into two independent maximizations
 - 1) Transition probabilities: **closed-form optimization**
 - 2) GARCH parameters: the optimization must be performed numerically
- The gradient of the function to be maximized can be calculated recursively

Importance sampling

- **Importance sampling (reweighting samples)**

$$\hat{Q}_{m_r}(\theta | \theta^{(r-1)}) = \sum_{i=1}^{m_r} \omega_i \log[f(y, S_i^* | \theta)] \Bigg/ \sum_{i=1}^{m_r} \omega_i, \text{ where}$$

$$\omega_i = \frac{f(y, S_i^* | \theta^{(r-1)})}{f(y, S_i^* | \theta^*)}$$

- **Problem (minor):** At each iteration of the Monte Carlo EM (MCEM) algorithm the parameters are updated and the sample size is (should be) increased; the importance proposal density may become inappropriate

SAEM

- Eventually, we would like to keep the sample size fixed and stop generating states; however, **the MCEM does not converge with a fixed sample size**
- **Solution:** Stochastic Approximation EM (SAEM)
(Delyon, Lavielle and Moulines, 1999)

$$\hat{Q}_r(\theta | \theta^{(r-1)}) = (1 - \gamma_r) \hat{Q}_{r-1}(\theta | \theta^{(r-2)}) + \gamma_r \left(\frac{1}{m_r} \sum_{i=1}^{m_r} \log[f(y, S_i^{(r)} | \theta)] \right)$$

step size

can be held fixed

SAEM

- **Problem:** we must keep track of all the samples
- **Solution:** combine SAEM with importance sampling

$$\hat{Q}_r(\theta | \theta^{(r-1)}) = \sum_{i=1}^m w_i^{(r)} \log[f(y, S_i^* | \theta)], \text{ where}$$

$$w_i^{(r)} = (1 - \gamma_r) w_i^{(r-1)} + \gamma_r \frac{\omega_i^{(r)}}{\sum_{i=1}^m \omega_i^{(r)}}$$

Importance sampling weights

**SAEM &
Importance sampling**



“Importance sampling”

The algorithm

Strategy

- 1) Start with **10** steps of the MCEM algorithm, increasing the sample size at each step

m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9	m_{10}
500	1,000	2,000	4,000	8,000	12,000	16,000	20,000	28,000	40,000

- 2) Do **5** steps with importance sampling
- 3) End with **5** steps of SAEM with importance sampling using the following step sizes

$$1/n^{1/2}, \quad n = 2, \dots, 6$$

Results – Sample size of 500

	Value	MLE	RMSE	A-StDev
μ_1	0.06	0.062	0.051	0.039
μ_2	-0.09	-0.090	0.219	0.211
ω_1	0.30	0.301	0.087	0.096
ω_2	2.00	2.573	1.668	1.539
α_1	0.35	0.344	0.139	0.110
α_2	0.10	0.117	0.154	0.106
β_1	0.20	0.194	0.158	0.162
β_2	0.60	0.480	0.332	0.279
p_{11}	0.98	0.977	0.012	0.010
p_{22}	0.96	0.953	0.029	0.023

Results – Sample size of 1500

	Value	MLE	RMSE	A-StDev
μ_1	0.06	0.061	0.023	0.025
μ_2	-0.09	-0.092	0.114	0.114
ω_1	0.30	0.301	0.057	0.054
ω_2	2.00	2.310	1.129	1.006
α_1	0.35	0.351	0.064	0.071
α_2	0.10	0.089	0.058	0.060
β_1	0.20	0.199	0.098	0.091
β_2	0.60	0.561	0.189	0.187
p_{11}	0.98	0.979	0.007	0.006
p_{22}	0.96	0.959	0.014	0.011

Results – Sample size of 5000

	Value	MLE	RMSE	A-StDev
μ_1	0.06	0.062	0.014	0.012
μ_2	-0.09	-0.091	0.057	0.068
ω_1	0.30	0.301	0.027	0.030
ω_2	2.00	2.064	0.583	0.564
α_1	0.35	0.356	0.038	0.037
α_2	0.10	0.094	0.035	0.031
β_1	0.20	0.195	0.050	0.048
β_2	0.60	0.597	0.110	0.098
p_{11}	0.98	0.980	0.003	0.003
p_{22}	0.96	0.959	0.006	0.006

Thank You!

Questions?