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

Log-returns on the S&P 500
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)

\[ N(-20\%, 26\%) \quad N(17\%, 12\%) \]
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**
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

\[ \begin{align*}
\sigma_0 & \quad s_1 = 1 \quad \sigma_1(1) \\
& \quad s_1 = 2 \quad \sigma_1(2) \\
& \quad s_2 = 1 \quad \sigma_2(1,1) \\
& \quad s_2 = 2 \quad \sigma_2(1,2) \\
& \quad s_2 = 1 \quad \sigma_2(2,1) \\
& \quad s_2 = 2 \quad \sigma_2(2,2)
\end{align*} \]
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)

\[
\begin{aligned}
\sigma_0 & \rightarrow \sigma_1(1) \quad \text{s}_1 = 1 \\
\sigma_0 & \rightarrow \sigma_1(2) \quad \text{s}_1 = 2 \\
\sigma_1(1) & \rightarrow h_1 \\
\sigma_1(2) & \rightarrow h_1 \\
\sigma_2(1) & \rightarrow h_2 \\
\sigma_2(2) & \rightarrow h_2 \\
\end{aligned}
\]

\[h_t = \text{Var}(y_t | y_{t-1}, y_{t-2}, \ldots)\]
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 $\theta$ and $y$

Simulate $\theta$ 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 using the EM algorithm.
• **EM Algorithm:** Dempster et al. (1977)

• **Insight:** let $\ell(\theta)$ represent the log-likelihood

$$\ell(\theta) = E[\log f(y, S | \theta) | y, \theta'] - E[\log f(S | y, \theta) | y, \theta']$$

$$= Q(\theta | \theta') - H(\theta | \theta')$$

• We wish to find a better value than $\theta'$, i.e., we need

$$\ell(\theta) - \ell(\theta') = [Q(\theta | \theta') - Q(\theta' | \theta')]$$

$$- [H(\theta | \theta') - H(\theta' | \theta')] > 0$$

$$\leq 0$$
**E-Step**

\[
Q(\theta \mid \theta^{(r-1)}) = \int \log[f(y, S \mid \theta)] f(S \mid y, \theta^{(r-1)}) \, dS
\]

\[
\approx \frac{1}{m_r} \sum_{i=1}^{m_r} \log[f(y, S_i^{(r)} \mid \theta)] = \hat{Q}_{m_r}(\theta \mid \theta^{(r-1)})
\]

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

**M-Step**

\[
\theta^{(r)} = \arg \max_\theta Q(\theta \mid \theta^{(r-1)})
\]
E-Step: Gibbs sampler

• How can we obtain draws from $f(S \mid y, \theta^{(r-1)})$?

  **Gibbs sampler (single-move)**
  
  • Full conditional distribution

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

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

\[
\omega_i = \frac{f(y, S_i^* \mid \theta^{(r-1)})}{f(y, S_i^* \mid \theta^*)}
\]

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

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

\[
w^{(r)}_i = (1 - \gamma_r) w^{(r-1)}_i + \gamma_r \frac{\omega^{(r)}_i}{\sum_{i=1}^{m} \omega^{(r)}_i}
\]
The algorithm

**Strategy**

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

<table>
<thead>
<tr>
<th></th>
<th>(m_1)</th>
<th>(m_2)</th>
<th>(m_3)</th>
<th>(m_4)</th>
<th>(m_5)</th>
<th>(m_6)</th>
<th>(m_7)</th>
<th>(m_8)</th>
<th>(m_9)</th>
<th>(m_{10})</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>500</td>
<td>1,000</td>
<td>2,000</td>
<td>4,000</td>
<td>8,000</td>
<td>12,000</td>
<td>16,000</td>
<td>20,000</td>
<td>28,000</td>
<td>40,000</td>
</tr>
</tbody>
</table>

2) Do **5** steps with importance sampling

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

\[
\frac{1}{n^{1/2}}, \ n = 2, \ldots, 6
\]
Results – Sample size of 500

<table>
<thead>
<tr>
<th></th>
<th>Value</th>
<th>MLE</th>
<th>RMSE</th>
<th>A-StDev</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_1$</td>
<td>0.06</td>
<td>0.062</td>
<td>0.051</td>
<td>0.039</td>
</tr>
<tr>
<td>$\mu_2$</td>
<td>-0.09</td>
<td>-0.090</td>
<td>0.219</td>
<td>0.211</td>
</tr>
<tr>
<td>$\omega_1$</td>
<td>0.30</td>
<td>0.301</td>
<td>0.087</td>
<td>0.096</td>
</tr>
<tr>
<td>$\omega_2$</td>
<td>2.00</td>
<td>2.573</td>
<td>1.668</td>
<td>1.539</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>0.35</td>
<td>0.344</td>
<td>0.139</td>
<td>0.110</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>0.10</td>
<td>0.117</td>
<td>0.154</td>
<td>0.106</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>0.20</td>
<td>0.194</td>
<td>0.158</td>
<td>0.162</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>0.60</td>
<td>0.480</td>
<td>0.332</td>
<td>0.279</td>
</tr>
<tr>
<td>$p_{11}$</td>
<td>0.98</td>
<td>0.977</td>
<td>0.012</td>
<td>0.010</td>
</tr>
<tr>
<td>$p_{22}$</td>
<td>0.96</td>
<td>0.953</td>
<td>0.029</td>
<td>0.023</td>
</tr>
</tbody>
</table>
## Results – Sample size of 1500

<table>
<thead>
<tr>
<th></th>
<th>Value</th>
<th>MLE</th>
<th>RMSE</th>
<th>A-StDev</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\mu_1)</td>
<td>0.06</td>
<td>0.061</td>
<td>0.023</td>
<td>0.025</td>
</tr>
<tr>
<td>(\mu_2)</td>
<td>-0.09</td>
<td>-0.092</td>
<td>0.114</td>
<td>0.114</td>
</tr>
<tr>
<td>(\omega_1)</td>
<td>0.30</td>
<td>0.301</td>
<td>0.057</td>
<td>0.054</td>
</tr>
<tr>
<td>(\omega_2)</td>
<td>2.00</td>
<td>2.310</td>
<td>1.129</td>
<td>1.006</td>
</tr>
<tr>
<td>(\alpha_1)</td>
<td>0.35</td>
<td>0.351</td>
<td>0.064</td>
<td>0.071</td>
</tr>
<tr>
<td>(\alpha_2)</td>
<td>0.10</td>
<td>0.089</td>
<td>0.058</td>
<td>0.060</td>
</tr>
<tr>
<td>(\beta_1)</td>
<td>0.20</td>
<td>0.199</td>
<td>0.098</td>
<td>0.091</td>
</tr>
<tr>
<td>(\beta_2)</td>
<td>0.60</td>
<td>0.561</td>
<td>0.189</td>
<td>0.187</td>
</tr>
<tr>
<td>(p_{11})</td>
<td>0.98</td>
<td>0.979</td>
<td>0.007</td>
<td>0.006</td>
</tr>
<tr>
<td>(p_{22})</td>
<td>0.96</td>
<td>0.959</td>
<td>0.014</td>
<td>0.011</td>
</tr>
</tbody>
</table>
## Results – Sample size of 5000

<table>
<thead>
<tr>
<th>Value</th>
<th>MLE</th>
<th>RMSE</th>
<th>A-StDev</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_1$</td>
<td>0.06</td>
<td>0.061</td>
<td>0.013</td>
</tr>
<tr>
<td>$\mu_2$</td>
<td>-0.09</td>
<td>-0.090</td>
<td>0.060</td>
</tr>
<tr>
<td>$\omega_1$</td>
<td>0.30</td>
<td>0.302</td>
<td>0.026</td>
</tr>
<tr>
<td>$\omega_2$</td>
<td>2.00</td>
<td>2.059</td>
<td>0.525</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>0.35</td>
<td>0.353</td>
<td>0.037</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>0.10</td>
<td>0.094</td>
<td>0.034</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>0.20</td>
<td>0.196</td>
<td>0.048</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>0.60</td>
<td>0.597</td>
<td>0.097</td>
</tr>
<tr>
<td>$p_{11}$</td>
<td>0.98</td>
<td>0.980</td>
<td>0.003</td>
</tr>
<tr>
<td>$p_{22}$</td>
<td>0.96</td>
<td>0.959</td>
<td>0.006</td>
</tr>
</tbody>
</table>
Thank You!

Questions?