Score Permutation Based Finite Sample Inference for GARCH Models

Balázs Csanád Csáji

Institute for Computer Science and Control (SZTAKI) of the Hungarian Academy of Sciences (MTA)

Overview

- Confidence regions (and hypothesis testing) for parameters of GARCH processes
- ScoPe: works by permuting the residuals in the score (gradient of the log-likelihood)
- Centered around the Quasi-Maximum Likelihood Estimate (QMLE)
- Distribution-free (w.r.t. the driving noise; even if it is heavy-tailed and skewed)
- Non-asymptotic (finite sample) guarantees
- Exact (user-chosen) coverage probabilities
- Applicable to nonstationary models, as well
- Confirmation on major stock market indices

Introduction

In many applications it is typical that larger disturbances are more likely followed by larger disturbances, while smaller fluctuations tend to be followed by smaller fluctuations. This phenomenon can be modeled by GARCH processes. Here, we extend the SPS method [1] to GARCH models.

GARCH Models

Formally, a GARCH(p,q) process, $\{X_t\}$, is defined by the following two equations [2]

$$X_t \triangleq \sigma_t \varepsilon_t,$$
$$\sigma_t^2 \triangleq \omega^* + \sum_{i=1}^{p} \alpha_i^* X_{t-i}^2 + \sum_{j=1}^{q} \beta_j^* \sigma_{t-j}^2,$$

where $\{\varepsilon_t\}$ is a strong white noise, i.e., an i.i.d. sequence of real random variables with zero mean and unit variance; variable σ_t^2 defines the conditional variance of X_t, given its own past up to $t-1$; and $\omega^* > 0$ as well as $\alpha_i^*, \beta_j^* \geq 0$ are constants.

Quasi-Maximum Likelihood

GARCH models are typically estimated by Quasi-Maximum Likelihood (QML) methods. They use a Gaussian "working hypothesis", but are guaranteed to work under mild statistical assumptions. The *conditional Gaussian quasi-likelihood* is

$$\mathcal{L}_n(\theta; x) \triangleq \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi\hat{\sigma}_t^2(\theta)}} \exp\left(-\frac{X_t^2}{2\hat{\sigma}_t^2(\theta)}\right),$$

where $x = (X_1, \dots, X_n)$ is the sample and

$$\hat{\sigma}_t^2(\theta) \triangleq \omega + \sum_{i=1}^{p} \alpha_i X_{t-i}^2 + \sum_{j=1}^{q} \beta_j \hat{\sigma}_{t-j}^2(\theta),$$

where $\theta \in \mathbb{R}^{p+q+1}$ is a generic vector encoding the parameters, $\theta \triangleq (\omega, \alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_q)$, while the "true" parameter vector is denoted by θ^*.

The QMLE is any measurable solution of

$$\hat{\theta}_n \triangleq \arg\max_{\theta\in\Theta} \mathcal{L}_n(\theta; x).$$

Which is equivalent to minimizing (namely, take a natural logarithm and drop constants)

$$\ell_n(\theta) \triangleq \frac{1}{n}\sum_{t=1}^{n}\left[\log \hat{\sigma}_t^2(\theta) + \frac{X_t^2}{\hat{\sigma}_t^2(\theta)}\right],$$

where $1/n$ is included for numerical stability.

Asymptotics of QMLE

Under mild regularity conditions (nondeg. noise & identifiability), the QMLE is *strongly consistent*

$$\hat{\theta}_n \xrightarrow{as} \theta^* \quad \text{as} \quad n \to \infty.$$

It can also be proved, assuming $\mathbb{E}[\varepsilon_0^4] < \infty$, that the QMLE is *asymptotically normal*

$$\sqrt{n}(\hat{\theta}_n - \theta^*) \xrightarrow{d} \mathcal{N}(0, \Gamma) \quad \text{as} \quad n \to \infty,$$

for a covariance matrix Γ depending on $\nabla_\theta\, \hat{\sigma}_0^2(\theta^*)$. This can be used to define (asymptotic) *confidence ellipsoids*. Assume Γ_n is an estimate of Γ, then

$$\widetilde{\Theta}_n(s) \triangleq \left\{ \theta \in \mathbb{R}^d : (\theta - \hat{\theta}_n)^{\mathrm{T}}\, \Gamma_n^{-1} (\theta - \hat{\theta}_n) \leq \frac{s}{n} \right\},$$

where $d \triangleq p+q+1$ and the probability that $\theta^* \in \widetilde{\Theta}_n$ is approximately $F_{\chi^2(d)}(s)$, which is the CDF of the χ^2 distribution with d degrees of freedom.

Gaussian Score

The QMLE satisfies the *likelihood equation*

$$\nabla_\theta\, \ell_n(\hat{\theta}_n) = 0,$$

and the gradient of the (conditional) log-likelihood function, the *score* function, can be written as

$$\nabla_\theta\, \ell_n(\theta) = \frac{1}{n}\sum_{t=1}^{n}\left(1 - \hat{\varepsilon}_t^2(\theta)\right)\frac{1}{\hat{\sigma}_t^2(\theta)}\nabla_\theta\, \hat{\sigma}_t^2(\theta),$$

where $\hat{\varepsilon}_t(\theta) \triangleq X_t / \hat{\sigma}_t(\theta)$ is a reconstructed residual for time t assuming parameter θ, and $\hat{\sigma}_t(\theta)$ is an estimate of σ_t, which can be calculated recursively.

Score Permutation

Note that $\hat{\varepsilon}_t(\theta^*) = \varepsilon_t$, for all t, assuming

(P1) The "true" system is in the model class.

(P2) The initial conditions are known.

Since $\{\varepsilon_t\}$ is i.i.d., their joint distribution is maintained under arbitrary index permutation $\pi(\cdot)$,

$$\{\varepsilon_t\} \overset{d}{=} \{\varepsilon_{\pi(t)}\}$$

Given a θ, the main idea is first to "invert" the system to get $\{\hat{\varepsilon}_t(\theta)\}$ and then generate alternative trajectories by randomly permutated residuals,

$$\hat{\varepsilon}_{\pi_i(1)}(\theta), \dots, \hat{\varepsilon}_{\pi_i(n)}(\theta),$$

for all $i \in \{1, \dots, m-1\}$, where m is user-chosen. Let π_0 be the identity permutation, i.e., $\pi_0(t) = t$. The original and the *perturbed score functions* are

$$B(\theta, \pi_i) \triangleq \frac{1}{n}\sum_{t=1}^{n}\frac{(1 - \hat{\varepsilon}_{\pi_i(t)}^2(\theta))}{\bar{\sigma}_t^2(\theta, \pi_i)}\nabla_\theta\, \bar{\sigma}_t^2(\theta, \pi_i),$$

where the perturbed variances $\bar{\sigma}_t^2(\theta, \pi_i)$ are

$$\bar{\sigma}_t^2(\theta, \pi_i) \triangleq \omega + \sum_{k=1}^{p} \alpha_k \bar{X}_{t-k}^2(\theta, \pi_i) + \sum_{j=1}^{q} \beta_j \bar{\sigma}_{t-j}^2(\theta, \pi_i)$$

which gives rise to an *alternative trajectory*

$$\bar{X}_t(\theta, \pi_i) \triangleq \bar{\sigma}_t(\theta, \pi_i)\, \hat{\varepsilon}_{\pi_i(t)}(\theta).$$

The *rank* of $\|B(\theta, \pi_0)\|^2$ within $\{\|B(\theta, \pi_i)\|^2\}$ is

$$\mathcal{R}_m(\theta) \triangleq 1 + \sum_{i=1}^{m-1} \mathbb{I}\left(\|B(\theta, \pi_0)\|^2 \succ \|B(\theta, \pi_i)\|^2\right),$$

where $\mathbb{I}(\cdot)$ is an indicator function and $\succ$ is $>$ with random tie-breaking. The ScoPe confidence set is

$$\widehat{\Theta}_n(m, r) \triangleq \{\, \theta \in \Theta : \mathcal{R}_m(\theta) \leq m - r \,\},$$

where $m > r > 0$ are user-chosen integers.

Main Theorem

Assuming (P1) and (P2), we have that

$$\mathbb{P}\left(\theta^* \in \widehat{\Theta}_n(m, r)\right) = 1 - \frac{r}{m}.$$

Experimental Results

The experiments focused on GARCH(1,1) models

$$X_t \triangleq \sigma_t \varepsilon_t,$$
$$\sigma_t^2 \triangleq \omega^* + \alpha^* X_{t-1}^2 + \beta^* \sigma_{t-1}^2,$$

using both simulated and real-world datasets. ScoPe was compared with asymptotic ellipsoids, residual- and likelihood ratio bootstrap methods. The daily closing prices of Nasdaq 100, S&P 500 and FTSE 100 were used from the entire period of 2014. Models were fitted to the *compound returns*, i.e., for each price sequence $\{P_t\}$, the data were transformed by $R_t = \log(P_t / P_{t-1})$.

Table 1: Relative Areas on Stock Market Indices (2014)

Method	Nasdaq 100	S&P 500	FTSE 100
Asym.Ell.	0.3426	0.1679	0.1535
Res.Boots.	0.3791	0.2549	0.2850
LR.Boots.	0.8150	0.7919	0.8326
ScoPe	0.3801	0.2862	0.2412

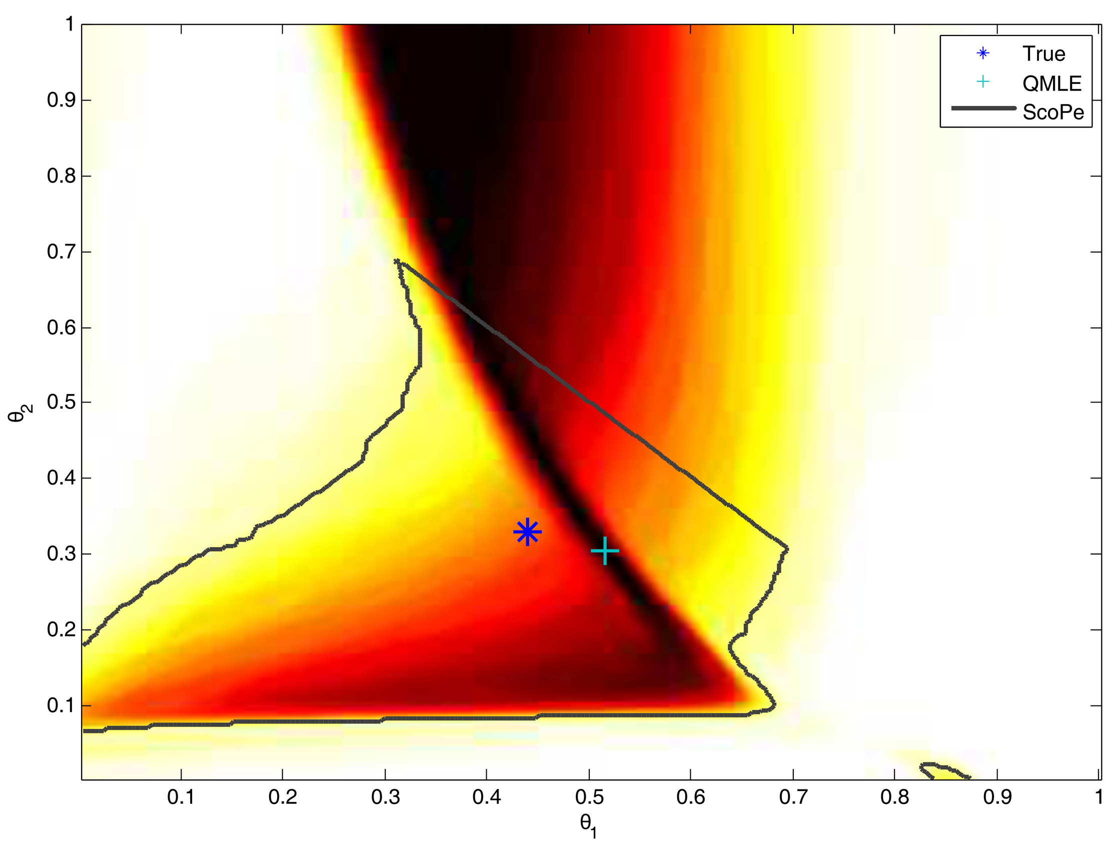

Figure 1: Logistic noise, $n = 100$, $m = 100$, $r = 10$; Exact 90% ScoPe confidence set for a stationary GARCH$(1,1)$.

References

[1] B. Cs. Csáji, M. C. Campi, and E. Weyer. Sign-Perturbed Sums: A new system identification approach for constructing exact non-asymptotic confidence regions in linear regression models. *IEEE Transactions on Signal Processing*, 63(1):169–181, 2015.

[2] C. Francq and J. M. Zakoian. *GARCH Models: Structure, Statistical Inference and Financial Applications.* John Wiley & Sons, 2011.

Contact Information

- Email: balazs.csaji@sztaki.mta.hu
- Website: http://www.sztaki.mta.hu/~csaji/

Poster for the 19th AISTATS Conference, Cadiz, Spain, 2016