Model estimation through matrix equations in financial econometrics

Federico Poloni[1]
Joint work with Giacomo Sbrana[2]

[1]Technische Universität Berlin (A. Von Humboldt postdoctoral fellow)
[2]Rouen Business School

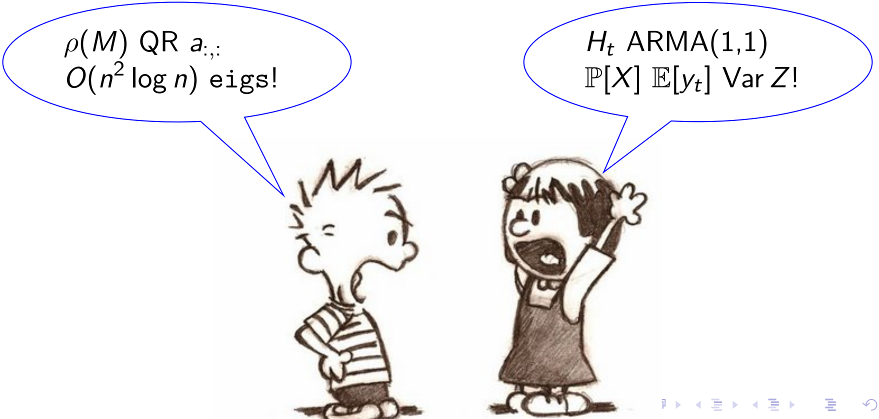
When Probability Meets Computation
Varese, June 2012

Last November, I received an e-mail message from a researcher in Econometrics looking for help with some matrix equations.

We spent some time trying to understand each other's language. . .

Last November, I received an e-mail message from a researcher in Econometrics looking for help with some matrix equations.

We spent some time trying to understand each other's language...



Scalar GARCH

A GARCH(1,1) model is a stochastic time series y_t such that

- $y_t = h_t^{1/2}\epsilon_t$, where ϵ_t is IID "noise" with mean 0 and variance 1
- h_t depends linearly on y_{t-1}^2 and h_{t-1}:

$$h_t = c + a y_{t-1}^2 + b h_{t-1}$$

- $a + b \leq 1$

Popular choice to model stock market volatility [Engle, '82], [Bollerslev, '86]

Estimation problem

Given a large number of observations, estimate the parameters c, a, b [Francq, Zakoïan, book '10]

Example

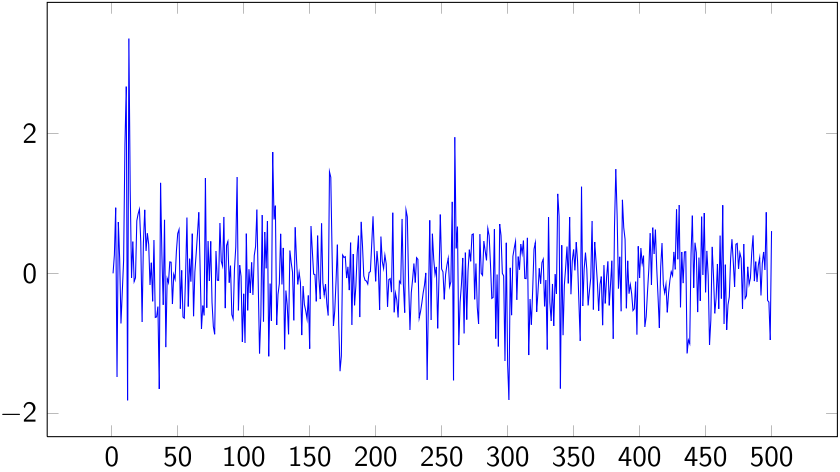


Figure: Example GARCH(1,1) data

Multivariate GARCH

Different possible generalizations.

Simplest one: every entry of the variance H_t may depend linearly on every entry of H_{t-1} and every $(y_{t-1})_i(y_{t-1})_j$

Vectorization: a tool to express this

$$y_t y_t^T \mapsto x_t = \begin{bmatrix} y_1 y_1 \\ y_1 y_2 \\ y_1 y_3 \\ y_2 y_2 \\ y_2 y_3 \\ y_3 y_3 \end{bmatrix}, \quad H_t = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 5 \\ 3 & 5 & 6 \end{bmatrix} \mapsto h_t = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{bmatrix}$$

Multivariate GARCH

$$h_t = c + Ax_{t-1} + Bh_{t-1}$$

Multivariate GARCH

A multivariate GARCH(1,1) model is a time series y_t of $d \times 1$ vectors such that

- $y_t = H_t^{1/2} \epsilon_t$, where ϵ_t is IID "noise" with mean 0 and variance I_d
- H_t is an affine linear function of H_{t-1} and $y_{t-1} y_{t-1}^T$. In formulas

$$h_t = c + A x_{t-1} + B h_{t-1}$$

- all the eigenvalues of $A + B$ are in the unit circle

Estimation problem

Given a large number of observations, estimate the parameters c, A, B [Francq, Zakoïan, book '10]

Example

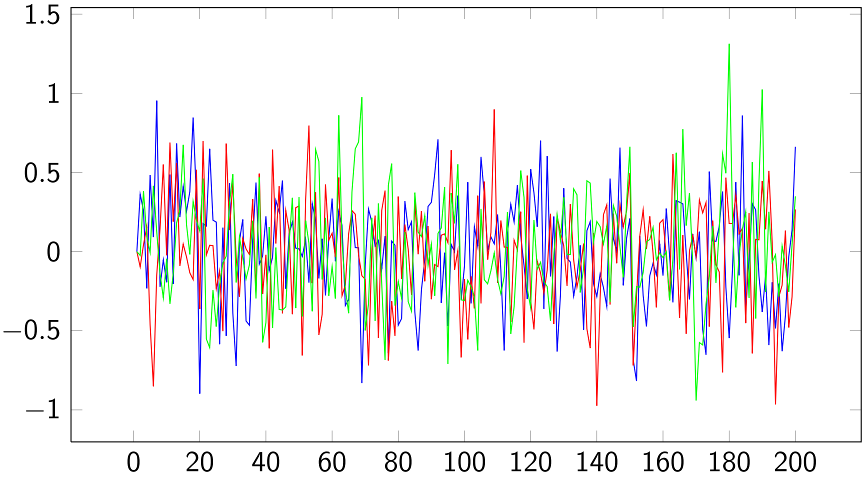


Figure: Example GARCH(1,1) data

Quasi-Maximum Likelihood

Most popular choice for estimating c, A, B: Quasi-Maximum Likelihood

Given a guess $\hat{c}, \hat{A}, \hat{B}$, we can compute $\hat{H}_t$ at each time and then the likelihood ℓ_t that a Gaussian $N(0, H_t)$ generates the observed y_t

Remark the noise is not assumed Gaussian in general, so this is an approximation

A good guess for the parameters is

$$\max_{\hat{c}, \hat{A}, \hat{B}} \prod_{t=1}^{n} \ell_t$$

(but it is not clear how to compute it)

Estimating a GARCH through ML

Most popular choice for estimating c, A, B: Quasi-Maximum Likelihood

Take likelihood function $L(c, A, B)$ and feed it to a “black-box” optimizer

- optimizer proceeds blindly, no knowledge of function expression
- difficult optimization problem: nonconvex, high-dimensional
- the code is delicate: step-size control, numerical derivatives...
- slow, big and scary

Our hope Finding a more manageable estimator

Estimating a GARCH through ML

Most popular choice for estimating c, A, B: Quasi-Maximum Likelihood

Take likelihood function $L(c, A, B)$ and feed it to a "black-box" optimizer

- optimizer proceeds blindly, no knowledge of function expression
- difficult optimization problem: nonconvex, high-dimensional
- the code is delicate: step-size control, numerical derivatives. . .
- slow, big and scary

Our hope Finding a more manageable estimator

Estimating c and $A + B$

Key property $y_t y_t^T$ is "almost" H_t, since $H_t = \text{var}[y_t]$

More precisely:
$\xi_t := x_t - h_t$ is an MDS, i.e., $\mathbb{E}[\xi_t] = 0$ independently of everything that happens up to $t - 1$

$$x_t = \xi_t + c + (A + B)x_{t-1} - B\xi_{t-1}$$

Yule-Walker–type formulas

If we compute autocorrelations $M_k = \mathbb{E}\left[(x_t - \mathbb{E}[x])(x_{t-k} - \mathbb{E}[x])^T\right]$,

$$M_{k+1} = (A + B)M_k \qquad k \geq 1$$

Autocorrelations are easily estimated using sample autocorrelations $\hat{M}_k = \frac{1}{n}\sum_{t=1}^{n} x_{t+k}x_t^T$. From them we can get $A + B$ and c.

Moving average

Now we use the moving average vector

$$j_t := x_t - (A+B)x_{t-1} - c = \xi_t - B\xi_{t-1}$$

(the second form is obtained using the formulas for h_t)

We get, with $\Phi := A + B$ and $\Sigma = \mathrm{Var}(\xi_t)$

$$\Gamma_0 = \mathbb{E}\left[j_t j_t^T\right] = M_0 - M_1\Phi^T - \Phi M_1^T + \Phi M_0 \Phi^T = \Sigma + B\Sigma B^T$$

$$\Gamma_1 = \mathbb{E}\left[j_{t+1} j_t^T\right] = M_1 - \Phi M_0 = -B\Sigma$$

Matrix equations

$$\Gamma_0 = \mathbb{E}\left[j_t j_t^T\right] = M_0 - M_1\Phi^T - \Phi M_1^T + \Phi M_0 \Phi^T = \Sigma + B\Sigma B^T$$
$$\Gamma_1 = \mathbb{E}\left[j_{t+1} j_t^T\right] = M_1 - \Phi M_0 = -B\Sigma$$

We can eliminate either Σ or B, getting

$$\Gamma_1^T - \Gamma_0 B^T + \Gamma_1 (B^T)^2 = 0 \tag{P}$$
$$\Gamma_0 = \Sigma + \Gamma_1 \Sigma^{-1} \Gamma_1^T. \tag{R}$$

(P) is a palindromic matrix equation [Mackey *et al* '05, Gohberg *et al* book '09]
(R) is a "baby-Riccati" nonlinear matrix equation (NME) [Engwerda *et al* '93, Meini '02, + others]

A closed-form estimator

$$\Gamma_1^T - \Gamma_0 B^T + \Gamma_1 (B^T)^2 = 0 \qquad \text{(P)}$$

We can solve (P) via linearization [. . .] or cyclic reduction/doubling [. . .]

The complete procedure

- Compute a few of the first sample autocorrelations $\hat{M}_k$
- Estimate $\widehat{A+B}$ and $\hat{c}$ using Yule-Walker results (how?)
- Estimate $\hat{\Gamma}_0$ and $\hat{\Gamma}_1$, autocorrelations of $j_t = x_t - (\widehat{A+B})x_{t-1} - \hat{c}$
- Solve the matrix equation (P) to get $\hat{B}$

Suggested by [Linton, Kristensen '06] for the univariate GARCH, where (P) is simply a quadratic equation

We can now generalize it to the more interesting multivariate case

Solving matrix equations (when it is possible)

$$\Gamma_1^T - \Gamma_0 B^T + \Gamma_1 (B^T)^2 = 0 \tag{P}$$

$$\Gamma_0 = \Sigma + \Gamma_1 \Sigma^{-1} \Gamma_1^T. \tag{R}$$

Solvability theory, putting together mostly known results for (R):

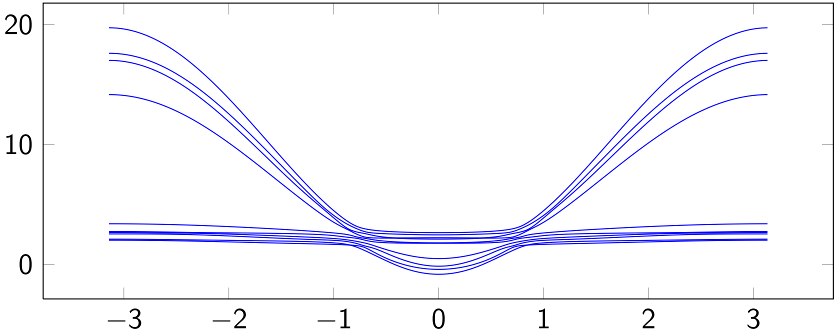
- If $P(\lambda) = \Gamma_1^T \lambda^{-1} + \Gamma_0 + \Gamma_1 \lambda \succeq 0$ on the unit circle, a unique stable B ($\rho(B) \leq 1$) exists
- Otherwise, no pair (B, Σ) with $\Sigma \succeq 0$ exists

If our model holds, B exists (but often eigenvalues close to the unit circle).

Problem

Γ_0, Γ_1 come from the data and must be estimated.
If large errors, (P) and (R) are perturbed to unsolvable equations

Example



Work in progress (with T. Brüll, C. Schröder): find small perturbations to Γ_0, Γ_1 which "move the lines up".

So, does it really work?

Yes, but still not as accurate as MLE. Weak point: the convergence $\hat{M}_k \to M_k$ is not that fast

However, much faster than MLE. Suggested uses:

- use it as a starting value for optimization in MLE...
- ...or use an iterative feasible least-squares procedure to refine it
  Almost finished: asymptotic consistency/normality properties of LS

Idea of feasible GLS

Method to refine an estimate $\hat{c}, \hat{A}, \hat{B}$
$x_t = h_t + (\text{error})$; $h_t = f(c, A, B, x_{t-1}, h_{t-1})$ linear in c, A, B

$$\min_{c,A,B} \sum \| f(c, A, B, x_{t-1}, h_{t-1}) - x_t \|^2$$

Problems:

- We do not know h_{t-1} in the formula
  Replace it by $\hat{h}_{t-1}$
- Variance of the error varies (it is h_t), least squares methods start from the assumption that it is fixed
  Rescale it with $\hat{h}_t$

We run a few iterations of this method and check which one has the largest likelihood

Simulated (Monte-Carlo) results

Only some preliminary results. . .

- 500, 1000 or 5000 observations
- $\rho(A + B)$ up to 0.9
- ☺ Speedup in QML around 30% for sufficiently hard experiments
- ☹ In some cases, convergence problems in QML with our starting values — we are trying to understand this
- ☺ Closed form estimator + 3–4 iterations of FGLS: accuracy on par with QML, much faster, simple to code.

Conclusions

- New application for matrix equations and structured linear algebra
- Correct estimations are crucial to assess risk in economic models

Conclusions

- New application for matrix equations and structured linear algebra
- Correct estimations are crucial to assess risk in economic models
- We help practitioners make the same errors, but much faster!

Thanks for your attention
And a happy retirement to Guy!

Conclusions

- New application for matrix equations and structured linear algebra
- Correct estimations are crucial to assess risk in economic models
- We help practitioners make the same errors, but much faster!

Thanks for your attention
And a happy retirement to Guy!

- ???
- Profit!