MS&E 349: Homework 2

Markus Pelger

Due 24pm, February 17th 2021. Submit on Canvas

Please submit one homework assignment per group. The homework solution should be submitted online to Canvas. Please indicate clearly the names of all group members. I prefer that solutions are typed in Latex, but it is also fine to submit scanned copies of handwritten solutions. Include the commented code in an appendix section. Please also submit the executable and commented code.

Theoretical Questions

Question 1 Method of Moments

Consider a random variable X that has an Poisson distribution with parameter λ, i.e. the probability of X is

$$P(X = k) = e^{-\lambda}\frac{\lambda^k}{k!} \qquad \text{for } k \geq 0 \qquad \text{and} \qquad P(X = k) = 0 \qquad \text{for } k < 0$$

Assume you observe a sequence of i.i.d. observations of random variables with Poisson distribution $\{X_t : 1 \leq t \leq T\}$. The first four moments are given by

$$E[X] = \lambda \qquad E[X^2] = \lambda^2 + \lambda \qquad E[X^3] = \lambda^3 + 3\lambda^2 + \lambda \qquad E[X^4] = \lambda^4 + 6\lambda^3 + 7\lambda^2 + \lambda.$$

1. **(Bonus question: not required and no additional points)**
   Calculate the first four moments of X.
   (Hint: you can use $e^\lambda = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!}$, $k^2 = k(k-1) + k$, $k^3 = k(k-1)(k-2) + 3k(k-1) + k$ and $k^4 = k(k-1)(k-2)(k-3) + 6k(k-1)(k-2) + 7k(k-1) + k$ to calculate the moments.)
2. Derive the method of moment estimator for λ using the first moment.
3. Derive the method of moment estimator for λ using the second moment.
4. Now consider the GMM estimator for λ using the first and second moment. Calculate the asymptotic distribution of $\hat{\lambda}_{GMM}$. What is the optimal weighting matrix for the GMM estimator? How does the optimal GMM estimator compare to the estimators based only on one moment.

. . . .

Question 2 GARCH

Consider the GJR-GARCH model

$$\begin{aligned} r_t &= \rho r_{t-1} + \epsilon_t \\ \epsilon_t &= \sigma_t e_t \\ \sigma_t^2 &= \omega + \beta\sigma_{t-1}^2 + \alpha\epsilon_{t-1}^2 + \gamma\epsilon_{t-1}^2 I_{\{\epsilon_{t-1}<0\}} \\ e_t &\sim i.i.d.\ N(0,1) \end{aligned}$$

Assume that the process is covariance stationary.

1. What is $E[r_{t+1}]$?
2. What is $E_t[r_{t+1}]$?

3. What is $Var(r_{t+1})$?

4. What is $Var_t(r_{t+1})$?

5. What is $Var_t(r_{t+2})$?

. . . .

Empirical Questions

The data that you need for this exercise can be found in the CRSP data base of WRDS. You have free access to CRSP with your Stanford ID. Please use the Stanford library link: http://libguides.stanford.edu/databases/crsp.

Please download the following data:

- Monthly and daily returns of the S&P 500 Composite Index from 07/01/1963 to 03/31/2017. (Look for Stock Market Indices in CRSP).

- Daily yields of 1-month interest rates from 01/02/1997 to 01/3/2009. (Look for Federal Reserve Bank/Interest Rates/Data/1-month).

Question 3 Volatility

In this question we want to estimate the volatility of the S&P 500 index. Consider a GARCH(1,1) model of the form

$$
\begin{aligned}
r_t &= \mu + \epsilon_t \\
\epsilon_t &= \sigma_t e_t \\
\sigma_t^2 &= \omega + \alpha \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2 \\
e_t &\sim N(0,1)
\end{aligned}
$$

1. Simulate a time series of daily log returns for $T_{day} = 13,200$ days of a GARCH(1,1) model with $\mu = 0.01, \omega = 0.0073, \alpha = 0.93$ and $\beta = 0.06$. Aggregate the daily log returns to $T_{month} = 600$ monthly returns by adding up 22 daily returns. This will be your simulated daily and monthly data set. The S&P 500 daily and monthly returns are the second data set.

2. Use the monthly simulated and return data of the S&P 500 index to estimate a GARCH(1,1) for monthly volatility. What features of the model and/or data are noteworthy? Report your estimates. You will need to use a recursive formulation in your code. Assume that all observations needed for starting the recursion are equal to their unconditional expectation.

3. Use the daily data (simulated and S&P 500 returns) to calculate a monthly time-series of realized volatilities (based on quadratic variations). (Assume that a month has 22 trading days.)

4. Compare your realized volatilities with your estimates based on the GARCH(1,1) model (for simulated and S&P 500 returns). Plot both time-series into the same graph.

5. Estimate an AR(1) model for realized volatility (for simulated and S&P 500 returns). What features of the model and/or data are noteworthy?

6. Use the estimated monthly realized volatilities and the estimated AR(1) model to produce one-month-ahead forecasts of the volatilities. Plot realized volatility and forecasted volatility.

7. Compute the Mean-Square-Error (MSE) of the AR(1) forecasts, defined as $E[(\sigma_{t+1} - \hat{\sigma}_{t+1|t})^2]$, where σ_{t+1} is the realized value of the volatility at $t+1$ and $\hat{\sigma}_{t+1|t}$ is its forecast given information at time t.

8. Use the estimated monthly volatilities and the estimated GARCH(1,1) model to produce one-month-ahead forecasts of the volatilities. Plot volatility as estimated by the GARCH(1,1) and forecasted volatility based on the GARCH(1,1).

9. Compute the Mean-Square-Error (MSE) of the GARCH(1,1) forecasts, defined as $E[(\sigma_{t+1} - \hat{\sigma}_{t+1|t})^2]$, where σ_{t+1} is the realized volatility based on the GARCH(1,1) model at $t+1$ and $\hat{\sigma}_{t+1|t}$ is its forecast given information at time t based on the GARCH(1,1).

....

Question 4 Parameter Estimation of SDEs

In this problem we estimate the parameters of the Cox- Ingersoll-Ross (CIR) process, which is often used to model interest rates and is defined as:

$$dr_t = \kappa(\theta - r_t)dt + \sigma\sqrt{r_t}dW_t \tag{1}$$

The model can also be used to model the volatility in stochastic variance models, since it always remain positive. The estimation will be applied to two data sets: (1) simulated data and (2) interest rate data.

Use an Euler approximation to obtain the following discrete time model of the CIR:

$$r_{k+1} = r_k + \kappa(\theta - r_k)\Delta_k + \epsilon_{k+1} \tag{2}$$

where $\Delta_k = t_{k+1} - t_k$ and ϵ_{k+1} is normally distributed with

$$E[\epsilon_{k+1}|F_{t_k}] = 0$$
$$E[\epsilon_{k+1}^2|F_{t_k}] = \sigma^2 r_k \Delta_k$$

and F_{t_k} denotes the information set at time t_k.

Simulate 100,000 steps with the Euler discretization and take every 250 observation of the process to obtain a simulated CIR model with 400 observations. Use the parameters $\kappa = 0.234, \theta = 0.081$ and $\sigma^2 = 0.0073$. Use the same simulated time series with 400 observations for all the estimations. The second time series is based on the monthly treasury yields. Report all your estimation results for both time series. The following questions will ask you to estimate the parameters $\gamma = \begin{pmatrix} \kappa & \theta & \sigma \end{pmatrix}'$ with different estimation approaches and to report the asymptotic covariance matrix Σ from $\sqrt{T}(\hat{\gamma} - \gamma) \xrightarrow{d} N(0, \Sigma)$. Please report also the standard errors for each estimate $\hat{\gamma}_i$, i.e. $\frac{1}{\sqrt{T}}\sqrt{\Sigma_{i,i}}$. (Hint: You can calculate derivatives numerically.)

1. **GMM:** Estimate the parameters in the CIR model and the covariance matrices of the estimators by using GMM. For this purpose we approximate the stochastic differential equation with an Euler scheme (higher order methods would work even better). The advantage of using the discretized process is that we can estimate complex models without much problem.

   Use the following four moment restrictions:

$$\epsilon_{k+1} - E[\epsilon_{k+1}|F_k]$$
$$(\epsilon_{k+1} - E[\epsilon_{k+1}|F_k])r_k$$
$$\epsilon_{k+1}^2 - E[\epsilon_{k+1}^2|F_k]$$
$$(\epsilon_{k+1}^2 - E[\epsilon_{k+1}^2|F_k])r_k$$

   In order to obtain an efficient GMM estimation, you can first estimate the parameters with the identity matrix as weighting matrix. Use these parameters to calculate the optimal weight matrix, chosen as the inverse of the covariance matrix for the moment restrictions. Estimate the new parameters with the new weight matrix.

2. **QML:** Estimate the parameters and their covariance matrices in the models above by using approximate Quasi Maximum Likelihood (by discretizing the model and using the likelihood generated by the approximate model).

3. **Transformed approximated likelihood:** Use a transformation of the data for the CIR-model and the Shoji and Ozaki approximated likelihood. Using the transformation $y_t = 2\sqrt{r_t}$

we obtain a state independent diffusion term. The Ito formula yields

$$
\begin{aligned}
dy_t &= \left(\frac{1}{\sqrt{r_t}} \kappa(\theta - r_t) \frac{-1}{2r_t^{3/2}} \frac{1}{2}\sigma^2 r_t \right) dt + \frac{1}{\sqrt{r_t}} \sqrt{r_t} dW_r \\
&= \left(\frac{2}{y_t} \kappa \left(\theta - \frac{y_t^2}{4} \right) + \frac{-4}{y_t^3} \sigma^2 \frac{y_t^2}{8} \right) dt + \sigma dW_t \\
&= \left(\frac{2}{y_t} \kappa\theta - \kappa \frac{y_t}{2} + \frac{-2}{y_t} \frac{\sigma^2}{4} \right) dt + \sigma dW_t \\
&= \left(\frac{2}{y_t} \left(\kappa\theta - \frac{\sigma^2}{4} \right) - \kappa \frac{y_t}{2} \right) dt + \sigma dW_t \\
&= \mu(y_t) dt + \sigma dW_t
\end{aligned}
$$

We can now use Shoji and Ozaki's approach[1] to approximate the likelihood with a Gaussian likelihood:

$$
\sum_{k=2}^{T} log(f(y_{t_k} | y_{t_{k-1}})) \approx \sum_{k=2}^{T} \left(\frac{-(y_{t_k} - m_k)^2}{2v_k} - \frac{1}{2} \log(2\pi v_k) \right)
$$

$$
\begin{aligned}
m_k &= y_{t_{k-1}} + \frac{a_k}{b_k} K_k + \frac{\sigma^2 c_k}{2b_k^2} (K_k - b_k \Delta_k) \\
v_k &= \frac{\sigma^2}{2b_k} \left(\exp(2b_k \Delta_k) - 1 \right) \\
K_k &= \exp(b_k \Delta_k) - 1 \\
\Delta_k &= t_k - t_{k-1} \\
a_k &= \mu(y_{t_{k-1}}) = \frac{2}{y_{t_{k-1}}} \left(\kappa\theta - \frac{\sigma^2}{4} \right) - \kappa \frac{y_{t_{k-1}}}{2} \\
b_k &= \mu'(y_{t_{k-1}}) = \frac{-2}{y_{t_{k-1}}^2} \left(\kappa\theta - \frac{\sigma^2}{4} \right) - \frac{\kappa}{2} \\
c_k &= \mu''(y_{t_{k-1}}) = \frac{4}{y_{t_{k-1}}^3} \left(\kappa\theta - \frac{\sigma^2}{4} \right)
\end{aligned}
$$

4. **Exact likelihood for the CIR model:** The CIR process has a known transition density. If we let $\Delta_k = t_k - t_{k-1}, c = \frac{2\kappa}{\sigma^2(1-\exp(-\kappa\Delta_k))}$, and $Y = 2cX$, then $Y_{t_k}|Y_{t_{k-1}}$ is distributed as a noncentral chi-squared with $2\kappa\theta/\sigma^2$ degrees of freedom and noncentrality parameter $Y_{t_{k-1}} \exp(\kappa\Delta_k)$. Using this we can write up the density as

$$
f_{X_{t_k}|X_{t_{k-1}}}(x_{t_k}|x_{t_{k-1}}) = ce^{-u-v} \left(\frac{v}{u} \right)^{q/2} I_q(2\sqrt{uv})
$$

where $u = cx_{t_{k-1}} \exp(-\kappa\Delta_k), v = cx_{t_k}, q = 2\kappa\theta/\sigma^2 - 1$ and $I_q(z)$ is the modified Bessel

[1] Shoji, I. and Ozaki, T. (1998). Estimation for nonlinear stochastic differential equations by a local linearization method, Stochastic Analysis and Applications, 16, 733?752.

function of the first kind of order q ($iv(v,z)$ in Python or $besseli(q,z)$ in Matlab).

Estimate the parameters and their covariance matrix for the Cox-Ingersoll-Ross model by using the exact likelihood.

5. **Simulated Maximum Likelihood (Bonus question: Not required and no additional points)**

Sometimes it is possible to use Maximum Likelihood on a simulated likelihood in stochastic differential equations. Assume that we have observations $y_n, n = 1, ..., N$ from some model. Also assume that the sampling interval is Δ. We now want to find the likelihood

$$L(\theta) = \prod_{i=1}^{N} f(y_i|y_{i-1}, \theta).$$

If this is not available in closed form, one might approximate it using simulations. First we discretize the dynamics of the stochastic differential equation using a scheme that gives a simple transition probability, preferably Euler. Then, divide the interval Δ into M subintervals of length $\delta = \Delta/M$. The idea is the to simulate K trajectories on a grid of size δ up to subinterval $M-1$ starting in y_n. Do this for every time n, resulting in K samples at every n. Under the Euler discretization, the transition density from $M-1$ to M is Gaussian with mean $\mu_{n,k}$ and standard deviation $\sigma_{n,k}$, both given by the model in question. Then we approximate the likelihood by the following:

$$\begin{aligned} L(\theta) &= \prod_{n=1}^{N} f(y_{n+1}|y_n, \theta) \\ &\approx \prod_{n=1}^{N} \left(\frac{1}{K} \sum_{k=1}^{K} \phi(y_{n+1}, \mu_{n,k|\theta}, \sigma_{n,k|\theta}) \right) \end{aligned}$$

where $\phi(y, \mu, \sigma)$ is the density of the Normal distribution with mean μ and standard deviation σ. The log likelihood is given by

$$l(\theta) = \log(L(\theta)) = \sum_{n=1}^{N} \log\left(f(y_{n+1}|y_n, \theta)\right) \approx \sum_{n=1}^{N} \left(\log\left(\frac{1}{K} \sum_{k=1}^{K} \phi(y_{n+1}, \mu_{n,k|\theta}, \sigma_{n,k|\theta}) \right) \right)$$

We then use numerical optimization techniques to maximize the loglikelihood. In this exercise we keep the number of intermediate steps fairly small, say $M = 2$ or 3. If M is too big, the variance of the sample will become large and make the estimation difficult. The remedy for this is to use an importance sampler.

It is usually a good idea to use the same sequence of random numbers each time the likelihood is evaluated. This is referred to as Common Random Numbers and is a way to avoid the Monte Carlo error in the minimization. Practically this is done by drawing a large number of random numbers, in this case $N \cdot (M-1) \cdot K$ and use these as input to the likelihood.

Estimate the parameters and their covariance matrices in the models above by using Simulated Maximum Likelihood. Choose an appropriate number of samples K based on your estimation results.

6. Compare your results for the different estimators.

. . . .