Lecture 5

Jussi Klemelä

February 17, 2009

1 Nonparametric portfolio selection

1.1 The Approach

We have available d assets and we want to choose an optimal allocation of the wealth among these assets. Let us denote with T the current time. We want to choose the portfolio so that the wealth is maximized at time $T+1$. We have available the previous asset prices $S_t \in (0,\infty)^d$, $t = 0,1,\ldots,T$, which can be used in portfolio selection. We denote $S_t = (S_t^1,\ldots,S_t^d)$.

A portfolio vector $b \in \mathbf{R}^d$ satisfies $b_i \geq 0$ and $\sum_{i=1}^d b_i = 1$. The value b_i gives the proportion of wealth which is invested in asset S_i at time T. Let W_T be the wealth available at time T. When the portfolio vector is b, then the wealth at time $T+1$ is

$$W_{T+1} = W_T \cdot b^T(S_{T+1}/S_T).$$

We shall choose b so that the regression function

$$f(x) = E\left(Y \mid X_T = x\right), \qquad Y = u(b^T(S_{T+1}/S_T))$$

is maximized, where X_T is the relevant information available at time T and $u : (0,\infty) \to \mathbf{R}$ is a utility function. We discuss various utility functions in Section 1.2.

Response variables. Let

$$U_t = S_t/S_{t-1} = (S_t^1/S_{t-1}^1,\ldots,S_t^d/S_{t-1}^d), \qquad t = 1,2,\ldots,$$

be the price relatives. We assume that the U_t are identically distributed random variables. Let

$$Y_t = u\left(b^T U_{t+1}\right) \in \mathbf{R}, \qquad t = 0,\ldots,T-1,$$

be the realizations of the response variable.

Explanatory variables. We denote the explanatory data by

$$X_t \in \mathbf{R}^p, \qquad t = t_0, \ldots, T,$$

and this data can consist of technical or fundamental information. For example, we can use the past observations to choose the optimal portfolio. Then

$$X_t = (U_{t-k+1}, \ldots, U_t) \in \mathbf{R}^{dk}, \qquad t = k, \ldots, T,$$

where $k \geq 1$ is an integer.

Portfolio selection. Our regression data is

$$(X_t, Y_t), \qquad t = t_0, \ldots, T-1.$$

We can assume that the regression data is identically distributed and denote with (X, Y) a random variable distributed as (X_t, Y_t). Note that Y depends on b. We estimate the regression function

$$f(x; b) = E(Y \mid X = x), \qquad x \in \mathbf{R}^p$$

with the estimator $\hat{f}(x; b)$. We choose the optimal portfolio vector $\hat{b}$ as

$$\hat{b} = \mathrm{argmax}_{b \in S_d} \hat{f}(X_T; b),$$

where

$$S_d = \left\{ b \in \mathbf{R}^d : b_i \geq 0, \ \sum_{i=1}^{d} b_i = 1 \right\}.$$

1.2 Utility functions

Continued from Lecture 4.

Markowitz utility. Portfolio choice with mean-variance preferences ranks the distributions according to

$$EU(b) - \frac{\gamma}{2} \mathrm{Var}[U(b)],$$

where $\gamma \geq 0$ is the coefficient of absolute risk aversion.

When we consider the one step long only portfolio choice, then $U(b) = b^T U_{T+1}$ and we can solve the portfolio weights explicitely:

$$b = \Sigma^{-1} 1_d \frac{\gamma W_t - 1_d \cdot \Sigma^{-1} \mu}{\gamma W_t 1_d \cdot \Sigma^{-1} 1_d} + \frac{\Sigma^{-1} \mu}{\gamma W_t},$$

where $\mu = EU_{T+1}$, $\Sigma = \mathrm{Var}[U_{T+1}]$, and $1_d = (1, \ldots, 1) \in \mathbf{R}^d$.

1.3 Case study.

Aït-Sahalia and Brandt (2001) choose the portfolio from SP500 stock index, 10 year treasury bond, and 1 month treasury bill. The portfolio is rebalanced once in month. They use DRI/Citibase database and choose the following explanatory variables.

1. the default spread (yield difference between Moody's Baa- and Aaa-rated corporate bonds),
2. the log dividend-to-price ratio (the sum of dividends paid on the S&P index over the past 12 months divided by the current level of the index),
3. the term spread (the yield difference between the 10- and 1-year government bonds),
4. index trend (momentum, the difference between the log of the current S&P index level and the log of the average index level over the previous 12 months).

2 Risk measures

Centers of distribution. The mean of a real valued random variable is a real valued number charactericing the center of a distribution. The mean of random variable $Y \in \mathbf{R}$ is defined by

$$\int_{-\infty}^{\infty} y f_Y(y)\, dy, \tag{1}$$

where $f_Y : \mathbf{R} \to \mathbf{R}$ is the density function of Y. The regression function has been defined as the conditional mean of Y. An other important measure of the center of a distribution is the median, defined in the case of continuous distribution function of a random variable $Y \in \mathbf{R}$ as the number $\mathrm{med}(Y) \in \mathbf{R}$ satisfying

$$P(Y \leq \mathrm{med}(Y)) = 0.5.$$

In general, covering also the case of discrete distributions, we can define the median uniquely as the generalized inverse of the distribution function:

$$\mathrm{med}(Y) = \inf\{y : P(Y \leq y) \geq 0.5\}. \tag{2}$$

Similarly as in the case of the conditional mean, we can also be interested in the estimation of the conditional median:

$$\mathrm{med}(Y \,|\, X = x) = \inf\{y : P(Y \leq y \,|\, X = x) \geq 0.5\}, \qquad x \in \mathbf{R}^d.$$

A third characerization of the center of a distribution is the mode, which is defined as an argument maximizing the density function of a random variable:

$$\text{mode}(Y) = \text{argmax}_{y \in \mathbf{R}} f_Y(y), \tag{3}$$

where $f_Y : \mathbf{R} \to \mathbf{R}$ is the density function of Y. The mode is in general not uniquely defined, and its use seems to be interesting only in cases where the density function is unimodal (has a unique maximum). The conditional mode can also be defined, as an argument maximizing the conditional density.

2.1 Characterizing the deviance

We have defined the mean, median, and the mode in (1), (2), and (3) as real valued characterizations of the center of the distribution. We need also characterizations of the deviance of the distribution, for example, in predicting the risk of a portfolio. We shall define the variance as a deviance measure related to the mean and the quantiles as deviance measures extending the median.

- The variance of random variable $Y \in \mathbf{R}$ is defined as

  $$\text{Var}(Y) = E(Y - EY)^2 = EY^2 - (EY)^2.$$

  The variance can be extended to other centered moments

  $$E|Y - EY|^k,$$

  for $k = 1, 2, \ldots$. The centered moments take contribution from the left and the right tail of the distribution. When we are interested to control only the left tail (the losses), then we can use the the lower partial moments:

  $$\text{LPM}_{\tau,k}(Y) = E\left[(\tau - Y)^k I_{[0,\infty)}(\tau - Y)\right] = \int_{-\infty}^{\tau} (\tau - Y)^k f_Y(y)\, dy,$$

  where $k = 0, 1, 2, \ldots$ and $\tau \in \mathbf{R}$ is a target rate. For example, when $k = 0$, then

  $$\text{LPM}_{\tau,0}(Y) = P(Y \le \tau)$$

  is the probability that Y is smaller than τ. The upper partial moment is

  $$\text{UPM}_{\tau,k}(Y) = E\left[(Y - \tau)^k I_{[0,\infty)}(Y - \tau)\right].$$

  We can define conditional versions of all these concepts. For example, the conditional variance function is defined as

  $$\text{Var}(Y \mid X = x) = E(Y^2 \mid X = x) - [E(Y \mid X = x)]^2, \qquad x \in \mathbf{R}^d$$

- The median can be extended to other quantiles:
$$\mathrm{Q}_p(Y) = \inf\{y : P(Y \le y) \ge p\},$$
where $0 < p < 1$. In the case of a continuous distribution function
$$P\left(Y \le \mathrm{Q}_p(Y)\right) = p.$$
The conditional quantile is defined as
$$Q_p(Y \,|\, X = x) = \inf\{y : P(Y \le y \,|\, X = x) \ge p\}, \qquad x \in \mathbf{R}^d,$$
where $0 < p < 1$. Conditional quantile estimation has been considered in Koenker (2005) and Koenker and Bassett (1978).

- The shortfall is defined as a risk measure where the expectation is taken only over the left tail, when the left tail is defined as the region which is to the left of a quantile of the distribution. Thus the shortfall is defined as
$$Q_p(Y) - E[Y \,|\, Y \le Q_p(Y)]$$
and the absolute shortfall is defined as
$$-E[Y \,|\, Y \le Q_p(Y)].$$

A reasonable risk measure satisfies the axioms of the coherent risk measure. Coherent risk measures ρ satisfy

1. Monotonicity: if $Y \ge X$, then $\rho(Y) \ge \rho(X)$.

2. Subadditivity: $\rho(X + Y) \le \rho(X) + \rho(Y)$.

3. Positive homogeneity: for $\lambda \ge 0$, $\rho(\lambda Y) = \lambda \rho(Y)$.

4. Transaltion invariance: for $a \in \mathbf{R}$, $\rho(Y + a) = \rho(Y) + a$.

The quantiles do not satisfy the subadditivity, but the expected shortfall is a coherent risk measure.

2.2 Conditional variance estimation

The kernel estimator of the conditional variance is defined as
$$\widehat{\mathrm{Var}}(Y \,|\, X = x) = \sum_{i=1}^n p_i(x) Y_i^2 - \left(\sum_{i=1}^n p_i(x) Y_i \right)^2,$$
where
$$p_i(x) = \frac{K_h(x - X_i)}{\sum_{i=1}^n K_h(x - X_i)}, \qquad i = 1, \ldots, n, \tag{4}$$
$K : \mathbf{R}^d \to \mathbf{R}$ is the kernel function, $K_h(x) = K(x/h)/h^d$, and $h > 0$ is the smoothing parameter.

2.3 Conditional quantile estimation

An estimator of a quantile can be defined with the help of the empirical distribution function

$$\hat{F}(y) = \frac{1}{n}\sum_{i=1}^{n} I_{(-\infty,y]}(Y_i), \qquad y \in \mathbf{R},$$

as

$$\hat{Q}_p(Y) = \inf\{y : \hat{F}(y) \geq p\}.$$

We can define the kernel estimator of the conditional distribution function as

$$\hat{F}(y \,|\, X = x) = \sum_{i=1}^{n} p_i(x) I_{(-\infty,y]}(Y_i),$$

where $p_i(x)$ are the kernel weights defined in (4). Thus the kernel estimator of a conditional quantile is

$$\hat{Q}_p(Y \,|\, X = x) = \inf\{y : \hat{F}(y \,|\, X = x) \geq p\}.$$

We have

$$\hat{F}\left(\hat{Q}_p(Y \,|\, X = x) \,|\, X = x\right) \approx p.$$

3 Illustrations

We look at the following code in

http://cc.oulu.fi/~jklemela/finatool/

```
# portfolio selection

# first we download data

ticker<-c("^GDAXI","^MDAXI")
destfile<-"~/pois"
# destfile<-"C:\\Documents and Settings\\user\\Desktop\\pois"
ry<-read.yahoo(ticker, source="web", destfile=destfile)

#save(file="/home/jsk/Arti/statfina/var/DaxMdax.var",list=c("ry"))
#save(file="/Users/jsk/Karhu/Arti/statfina/var-ada/DaxMdax.var",list=c("ry"))
#load(file="/Users/jsk/Karhu/Arti/statfina/var-ada/DaxMdax.var")
```

```
dm<-data.manip(ry,ticker)

method<-rep("price",length(ticker))
dfs<-data.final(dm,ticker,method=method)
plot(dfs[,1],type="l")

plot(dfs[,2],type="l")

method<-rep("relative",length(ticker))
df<-data.final(dm,ticker,method=method)
plot(df)

# we shall use the previous price relatives to predict future price relatives

d<-length(ticker)
n<-dim(df)[1]
U<-matrix(0,n,d)
for (i in 1:d) U[,i]<-df[1:n,i]

k<-4
marginal<-"gauss"
mp<-make.portdat(U,k,marginal=marginal,rate=0.02/360)


# we study the historical performance of the nearest neighborhood method

estimator<-"nn"
m<-15
gamma<-15
pfseq<-pf.seq(mp$Z,mp$X,estimator=estimator,m=m,gamma=gamma)
#,quantile=TRUE,alpha=0.05)
#,markow=TRUE)

end<-length(pfseq$wealth)
start<-1 #end-round(n/2)
plot(pfseq$wealth[start:end]/pfseq$wealth[start],type="l")

plot(pfseq$port[start:end,1])

plot(pfseq$return[start:end])
```

```
# we compare the nn-portfolio choice to the equally weighted portfolio

method<-rep("price",length(ticker))
dp<-data.final(dm,ticker,method=method)
dpmean<-(dp[,1]+dp[,2])/2

mata<-matrix(0,end-start+1,2)
mata[,1]<-dpmean[start:end]/dpmean[start]
mata[,2]<-pfseq$wealth[start:end]/pfseq$wealth[start]
matplot(mata,type="l",xlab="time",ylab="wealth")
```

4 Examination

A possible questions in the examination:

4) Define variance, quantiles, lower partial moments, and shortfall.
5) Explain how conditional variance and conditional quantiles can be estimated with a kernel method.

References

Aït-Sahalia, Y. and Brandt, M. W. (2001), 'Variable selection for portfolio choice', *J. Finance* **56**(4), 1297–1351.

Koenker, R. (2005), *Quantile Regression*, Cambridge University Press, Cambridge.

Koenker, R. and Bassett, G. (1978), 'Regression quantiles', *Econometrica* **46**, 33–50.