Commands for financial data management and portfolio optimization

Carlos Alberto Dorantes Dosamantes
Graduate School of Business, Querétaro Campus
Assistant Professor

2013 Mexican Stata Users Group Meeting
May 3, 2013

Abstract

- Commands for online market data collection, data manipulation, and basic financial analysis for portfolio optimization are presented. When allowing for short sales, the minimum variance portfolio (MVP) is analytically estimated using the Lagrange multiplier method. When short sales are not allowed there is no analytical solution, so the literature recommends using quadratic programming to find the MVP. Instead of following this numeric algorithm to estimate the MVP, I designed an algorithm based on iterations and the analytical formula derived from the Lagrange multiplier method.

Commands - description

Command (tentative name)	Description
retornosyh	Downloads online stock/index prices, cleans and formats data, calculates returns (r & R), and merges the series into one ending file ready for analysis
meanrets	Estimates the expected simple returns (R's) of a series of continuously compounded returns (r's)
gmvport	Estimates the Global Minimum Variance Portfolio allowing or not for short sales. If the option **noshort** is specified the non-negative constrains for portfolio weighs are considered
mvport	Estimates the Minimum Variance Portfolio given a specific required/desired return. If the noshort option is specified, short sales are not allowed.
ovport	Estimates the Optimal Variance Portfolio given a specific risk-free rate. This is also known as the "tangent" portfolio

... commands - description

Command (tentative name)	Description
simporta	Simulate portfolios in the feasible region given a set of returns without allowing for short sales
efrontier	Estimates the efficient frontier with or without allowing for short sales
cmline	Estimates the efficient frontier without allowing for short sales along with the tangent portfolio and the capital market line

retornosyh command

- Input:

  - A list of stock symbols and/or indexes from Yahoo Finance

  - Range dates, frequency (d, m, w), price to download (close, adjclose, low, high, open)

  - Merge type (casewise deletion)

- Output:

  - Stata time-series file with price, volume, simple return, and continuously compounded return of each stock/index

Quick review of portfolio theory

- Estimation of stock expected returns

- Estimation of gmvp allowing for short sales

- Estimation of mvp allowing for short sales

- Estimation of gmvp without allowing for short sales

- Estimation of mvp without allowing for short sales

meanrets command: Expected Returns

$$E[R_i] = \prod_{j=1}^{N}\left(1 + R_{i,j}\right) - 1 = e^{1/N \sum_{j=1}^{N} r_{i,j}} - 1$$

In matrix form:

$$E[\boldsymbol{R}] = e^{\left(\frac{1}{N}r'\boldsymbol{1}\right)} - 1 = \mathbf{MR}$$

Where:

$E[R_i]$= Expected Return of the stock/index i

$R_{i,j}$ = Historical simple return for stock/index i for the period j. Periods can be days, weeks or months

$r_{i,j}$ = Historical continuously compounded return for stock/index i for the period j.

$E[\boldsymbol{R}]$ = MR= vector of Expected Simple Returns or vector of mean returns

Expected Portfolio Return, Variance and Risk

According to Portfolio Theory, the expected return of the portfolio of N stocks is estimated as:

$$E[P] = \sum_{i=1}^{N} w_i * E[R_i]$$

In matrix form:

$$E[P] = \mathbf{w}' * \mathbf{MR} = \mathbf{cross(w, MR)}$$

The expected variance of the portfolio is estimated as:

$$Var[P] = \sum_{i=1}^{N} \sum_{j=1}^{N} w_i * w_j * Cov(r_i, r_j)$$

In matrix form:

$$Var[P] = \mathbf{w}' * \mathbf{COV} * \mathbf{w} = \mathbf{cross(w, COV)} * \mathbf{w}$$

Portfolio risk is estimated as:

$$\sqrt{\mathbf{w}' * \mathbf{COV} * \mathbf{w}}$$

w_i = weight or proportion invested in stock/index i; **w** = weight Vector

$Cov(r_i, r_j)$ = Covariance of cc returns of stocks i and j; **COV** = Var-Cov Matrix

gmvport command: Global Minimum Variance Portfolio

Minimize Portfolio Standard Deviation:

$$\sqrt{\mathbf{w}' * \mathbf{COV} * \mathbf{w}}$$

Subject to:

$$\mathbf{w}' * \mathbf{1} = 1$$

Solving this optimization problem with Lagrange multipliers, the Portfolio with the minimum variance (GMVP) is:

$$\mathbf{w_{gmv}} = \frac{\mathbf{COV}^{-1} * \mathbf{1}}{\mathbf{1}' * \mathbf{COV}^{-1} * \mathbf{1}}$$

An equivalent but more efficient computation of this vector is:

$$\mathbf{w_{gmv}} = \frac{rowsum(\mathbf{COV}^{-1})}{sum(\mathbf{COV}^{-1})}$$

$\mathbf{COV}^{-1}$ **=** Inverse of the Variance-Covariance matrix of cc returns

rowsum is a Mata function that computes the sum of rows of a matrix

sum is a Mata function that computes the sum of all elements of a matrix

mvport command: Minimum Variance Portfolio

If we want a required rate of return the optimization problem would be:

Minimize Portfolio Standard Deviation:

$$\sqrt{\mathbf{w}' * \mathbf{COV} * \mathbf{w}}$$

Subject to:

$$\mathbf{w}' * \mathbf{1} = 1$$
$$\mathbf{w}' * \mathbf{MR} = MR_*$$

$\mathbf{MR}$ = Vector of Expected Returns; MR_* = Required expected return

Solving this optimization problem with Lagrange multipliers:

$$w_{mv} = \frac{\left\{\mathbf{B} * \mathbf{COV}^{-1} * \mathbf{1} - \mathbf{A} * \mathbf{COV}^{-1} * \mathbf{MR} + MR_* * \left(C * \mathbf{COV}^{-1} * \mathbf{MR} - A * \mathbf{COV}^{-1} * \mathbf{1}\right)\right\}}{D}$$

$$w_{mv} = \frac{\left\{\mathbf{B} * \mathbf{RS} - \mathbf{A} * \mathbf{COV}^{-1} * \mathbf{MR} + MR_* * \left(C * \mathbf{COV}^{-1} * \mathbf{MR} - A * \mathbf{RS}\right)\right\}}{D}$$

$$\mathbf{RS} = \text{rowsum}(\mathbf{COV}^{-1})$$
$$A = \mathbf{MR}' * \mathbf{COV}^{-1} * \mathbf{1} = \mathbf{MR}' * \mathbf{RS}$$
$$B = \mathbf{MR}' * \mathbf{COV}^{-1} * \mathbf{MR}$$
$$C = \mathbf{1}' * \mathbf{COV}^{-1} * \mathbf{1} = \text{sum}(\boldsymbol{COV}^{-1})$$
$$D = B * C - A^2$$

gmvport, noshort

- GMV Portfolio without allowing for short sales

- Quadratic programming is needed since inequality constraints are added to the optimization problem (non-negative weights)

- However, the following algorithm was designed as alternative:

  1) T=N=Total Financial Assets

  2) The GMVP is estimated with gmvport with T assets allowing for short sales

  3) Weights are sorted. If the minimum weight is non-negative, finish the algorithm. If not, assign zero to the more negative weight and take its corresponding asset out of the portfolio, so now T=T-1

  4) If T=1, finish the algorithm. If not, go to 2)

mvport, noshort

- Minimum variance portfolio given a specific required rate of return and without allowing for short sales

- Instead of Quadratic programming method, an algorithm was designed

- This algorithm is basically the same as gmvport, noshort, but instead of using gmvport, the mvport command is used

simporta command: Simulation/creation of portfolios in the feasible region

- This command estimates feasible portfolios using random combination of weights and also using an algorithm (iterations) to create all possible portfolios up to maximum 5 stocks moving each weight from 0 to 1 with a specific delta.

- If the option delta is small, then many portfolios will be created, and the border of the efficient and non-efficient portfolios will be easier to observe

- Command mainly for teaching purposes

efrontier command: Efficient Frontier

- 2 Efficient Frontier are estimated: 1) allowing short sales and 2) not allowing short sales

- Algorithm designed:

    – Frontier with short sales: a loop is used to estimate the portfolio risk. Loop goes from the GMVP return to the maximum return of the stocks, and for each specific return, the mvport command is used to estimate the portfolio risk

    – Frontier without short sales: same as above, but using the gmvport and the mvport with the option **noshort**

ovport command: optimal portfolio

Given a specific risk-free rate the optimal or tangency portfolio is estimated finding the highest Sharpe ratio

Maximize the ratio:

$$\frac{w' * MR - R_f}{\sqrt{\mathbf{w}' * \mathbf{COV} * \mathbf{w}}}$$

Subject to:

$$\mathbf{w}' * \mathbf{1} = 1$$

Solving this optimization problem with Lagrange multipliers:

$$\mathbf{w_{ov}} = \frac{\left\{\mathbf{COV^{-1}} * \left(\mathbf{MR} - R_f * \mathbf{1}\right)\right\}}{B - C * R_f}$$

$$B = \mathbf{MR}' * \mathbf{COV^{-1}} * \mathbf{1} = \mathbf{MR}' * rowsum(\mathbf{COV^{-1}})$$

$$C = \mathbf{1}' * \mathbf{COV^{-1}} * \mathbf{1} = sum(\mathbf{COV^{-1}})$$

cmline command: Capital Market Line

- Based on Portfolio theory and the Capital Asset Pricing model, the CML becomes the efficient frontier when a risk-free rate is included in the portfolio

- This command estimates the CML without allowing for short sales (for the non risk-free assets)

- Designed algorithm:
  - 1) The efrontier, noshort is estimated
  - 2) Sharpe ratios for all efficient portfolios are computed, and sorted
  - 3) Identify the efficient portfolio with the maximum Sharpe ratio
  - 4) Having the portfolio risk as the X axis and portfolio return as the Y axis, estimate the equation of the line from the Risk-free rate point (0,Rf) to the optimal or tangency portfolio

Characteristics of the commands

- All commands are rclass commands

- The Stata **if** and **in** conditionals are allowed

- According to the command, scalars, vectors and matrices are returned. For example, the efrontier command return the vector of efficient frontier portfolios (w and w/out short sales), the gmvp, etc

- All commands use Mata code for most of the computations

Status of the commands

- Testing of commands is in process. R is being used as a benchmark. Until now, all tests have passed

- Help of commands will be created son

- Hopefully, I will offer the commands for the scc repository, so anybody will be able to use them

- I appreciate your suggestions and help!

Examples, using some commands

retornosyh AMXL.MX BIMBOA.MX CEMEXCPO.MX ELEKTRA.MX ALFAA.MX, ///

 fm(1) fd(1) fy(2006) lm(11) ld(30) ly(2012) frequency(m) price(close)

The screen output of this command is:

Means and correlations of Continuously Compounded returns r:

(obs=68)

```
   Variable |      Mean    Std. Dev.      Min        Max
-------------+-----------------------------------------------------
   r_AMXL_MX |   .0026247    .062295   -.2339502    .1560733
 r_BIMBOA_MX |   .0156082   .0781074   -.1775083    .2005862
r_CEMEXCPO~X |   -.0103035  .1581569   -.6683896    .3523211
 r_ELEKTRA_MX|   .0198862   .1719005   -.5827958    .4376447
   r_ALFAA_MX |   .023345   .1318409   -.6611445    .3928333
```

```
        | r_AMXL~X r_BIMB~X r_CEME~X r_ELEK~X r_ALFA~X
-------------+----------------------------------------------
   r_AMXL_MX |   1.0000
 r_BIMBOA_MX |   0.3554   1.0000
r_CEMEXCPO~X |   0.5650   0.3279   1.0000
 r_ELEKTRA_MX|   0.1469   0.3646   0.1133   1.0000
   r_ALFAA_MX |   0.6023   0.4614   0.6563   0.2391   1.0000
```
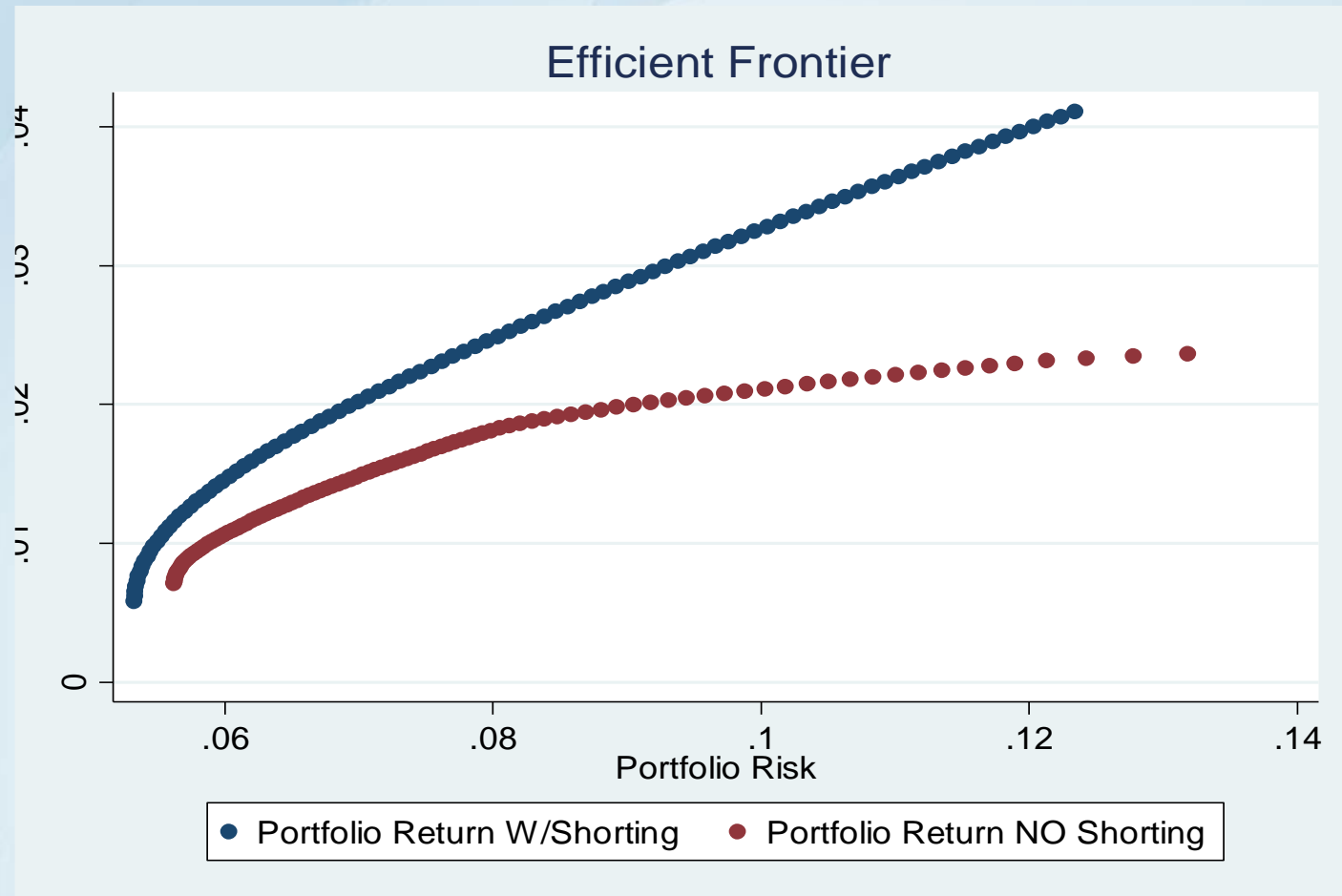
5 stocks from Yahoo Finanzas were downloaded; 83 periods in the time series

Number of observations with valid continuously compounded returns for all stocks: 68 out of 83

... Examples

retornosyh AMXL.MX BIMBOA.MX CEMEXCPO.MX ELEKTRA.MX ALFAA.MX, ///
 fm(1) fd(1) fy(2006) lm(11) ld(30) ly(2012) frequency(m) price(close)
efrontier r_*, nport(100)

Efficient Frontier

Commands for Financial data management
and Portfolio Optimization

… Examples

retornosyh AMXL.MX BIMBOA.MX CEMEXCPO.MX ELEKTRA.MX ALFAA.MX, ///
 fm(1) fd(1) fy(2006) lm(11) ld(30) ly(2012) frequency(m) price(close)

efrontier r_*, nport(100)

return list

matrices:

      r(exprets) : 5 x 1 (vector of expected stock returns)

       r(cov) : 5 x 5 (Var-Covariance matrix)

     r(sdefws) : 100 x 1 (vector of port sd of efficient portf with short)

     r(sdefwos) : 100 x 1 (vector of port sd of efficient port w/o short)

      r(refws) : 100 x 1 (vector of port ret of efficient port with short)

     r(refwos) : 100 x 1 (vector of port ret of efficient port w/o short)

     r(wefws) : 100 x 68 (set of weights vectors of the eff port with shorting)

     r(wefwos) : 100 x 68 (set of weights vectors of the eff port w/o short)

… Examples

retornosyh AMXL.MX BIMBOA.MX CEMEXCPO.MX ELEKTRA.MX, ///
 fm(1) fd(1) fy(2006) lm(11) ld(30) ly(2012) frequency(m) price(close)
cmline r_* if periodo>=tm(2009m1), nport(100) rfrate(0.0015)

Efficient Frontier & Capital Market Line, Risk-free rate=.0015
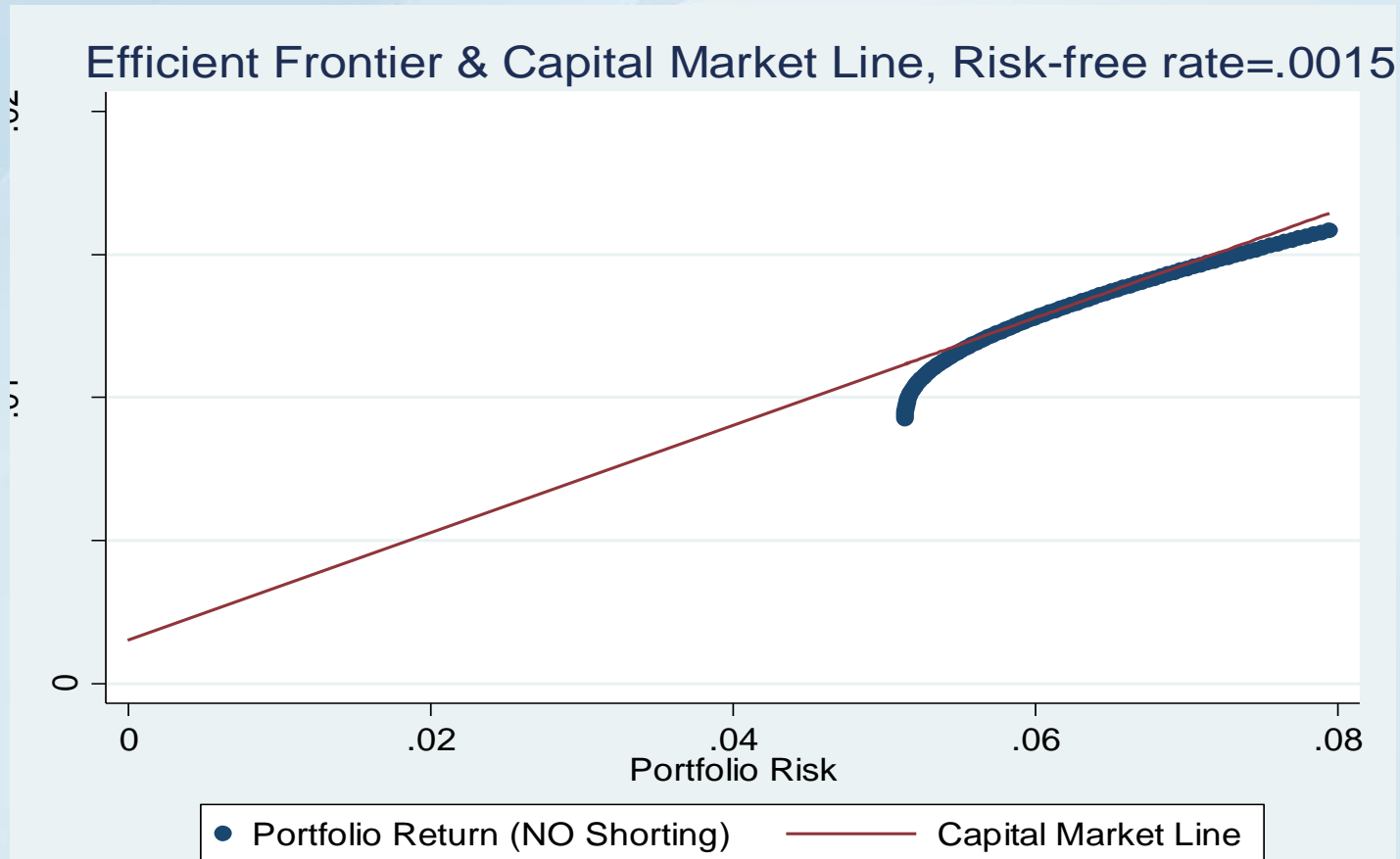
Thanks!

You can reach me at:

cdorante@itesm.mx

Alberto Dorantes Dosamantes

Professor at Monterrey Tech, Graduate School of Business, Querétaro Campus