Empirical Study of Ten Internet&Software Stocks in S&P500

Yue Wang, Wei Zhang
Fan Song, Jinhua Yang

Ladislaus von Bortkiewicz Chair of Statistics
Humboldt–Universität zu Berlin
http://lvb.wiwi.hu-berlin.de

Basics

- Data Source:
  1. http://finance.yahoo.com/
  2. https://www.treasury.gov/resource-center/data-chart-center/interest-rates/Pages/TextView.aspx?data=yield
- Time Period: 2015/05/01-2016/04/25
- $R_{t+1} = ln(P_{t+1}/P_t)$
- R_m: Mark return(Return of S&P 500 index)
- R_f: Risk free return(treasury bill rate)
- Panel analysis: Weekly return of R_i and R_m
- CAMP test: 13-week compounded return for R_i and R_m
- VaR: Daily Return R_i

Basics

Name of the company	Ticker
Google	GOOGL
Ebay	EBAY
Facebook	FB
Yahoo	YHOO
Wester Union	WU
Verisign Inc	VRSN
Netflix Inc.	NFLX
Total System Service	TSS
Fidelity National Information Services	FIS
Automatic Data Processing	ADP

Table 1: Stock List

Outline

1. Basics ✓
2. Data Description
3. Panel Data Analysis
4. CAPM Regression
5. Estimation of Portfolio VaR

Data Summary

```
setwd("d:/data")
Paneldata = read.csv("Panel.csv")
# *data description analysis *plot graphics (use different color for different
# companies)
Ri = Paneldata$Ri
Rm = Paneldata$Rm
summary(Paneldata$Ri)
summary(Paneldata$Rm)
library("stargazer")
stargazer(Paneldata, align = T)
```

- Balanced Panel Data N=10, T=52
- MP: Market Price - S&P 500 index
- ACP: Adjusted Closing Prices
- Rm: Market Return
- Ri: Stock Return for Each Company
- ari: Averaged Stock Return
- arm: Averaged Market Return
- amer: Averaged Market Excess Return
- aser: Averaged Stock Excess Return

Statistic	Mean	St. Dev.	Min	Max
Rm	−0.0002	0.02	−0.06	0.03
Ri	0.003	0.04	−0.16	0.27
Rf	0.15	0.12	0.01	0.35
ari	0.22	1.22	−0.90	21.22
arm	0.03	0.25	−0.55	0.52
amer	−0.12	0.28	−0.77	0.51
aser	0.08	1.24	−1.11	21.20

Table 2: Summary statistics of Panel Data Set

Heterogeneity across Companies and Time

```r
library(gplots)
plotmeans(Ri ~ company, main = "Heterogeineity across companies", data = Paneldata)
plotmeans(Ri ~ date, n.label = FALSE, main = "Heterogeineity across time", data = Paneldata)
```

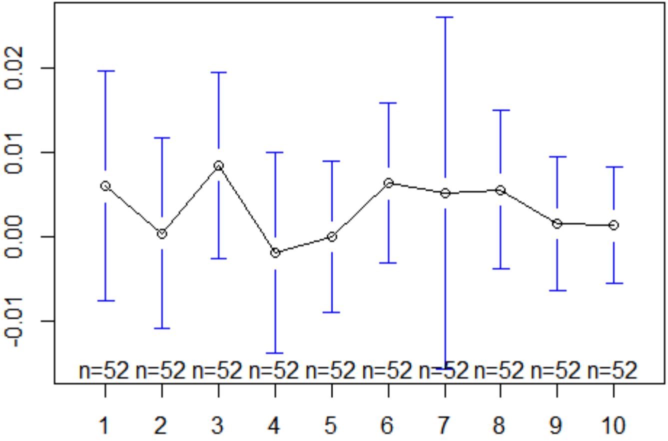

Figure 1: Heterogeneity across companies

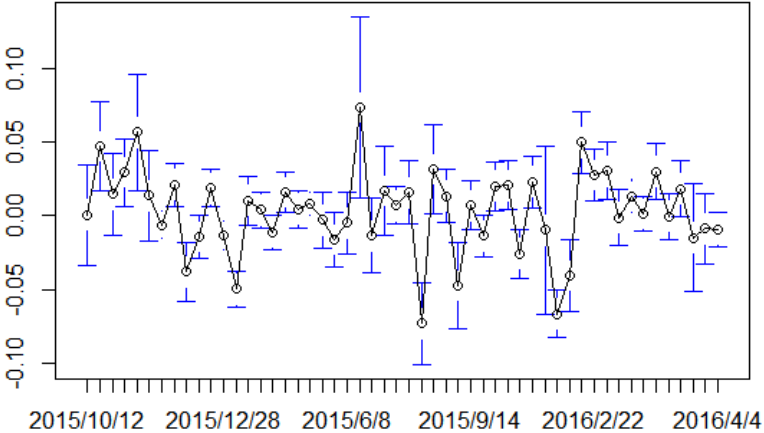

Figure 2: Heterogeneity across time

Relationship between market return and stock return

```
com = factor(Paneldata$company)
xyplot(Ri ~ Rm | com, data = Paneldata, panel =
  function(x, y) {
  panel.xyplot(x, y)
  panel.abline (h = median (y), lty = 2, col = "
    gray ")
  panel.lmline(x, y, col = "red")
}, xlab = "market return", ylab = "stock return",
  main = "relationship between market return and
  stock return")
```

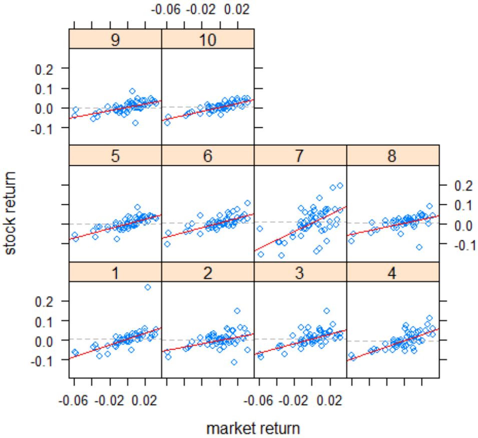

Figure 3: Relationship between market return and stock return

Outline

1. Basics ✓
2. Data Description ✓
3. Panel Data Analysis
4. CAPM Regression
5. Estimation of Portfolio VaR

Panel Analysis

```
# Pooled Regression Model Regression
library(plm)
pooling = plm(Ri ~ Rm, data = Paneldata, index = c("
  company", "date"), model = "pooling")
summary(pooling)

# Fixed Effects Model Regression
fixed = plm(Ri ~ Rm, data = Paneldata, index = c("
  company", "date"), model = "within")
summary(fixed)

# Random Effects Model Regression
random = plm(Ri ~ Rm, data = Paneldata, index = c("
  company", "date"), model = "random")
summary(random)
```

Pooled Regression Results

	Dependent variable:
	Ri
Rm	1.235***
	(0.071)
Constant	0.003**
	(0.001)
Observations	520
R^2	0.368
Adjusted R^2	0.366
F Statistic	301.162***

Table 3: Pooled Regression Results

Fixed Effects Model Regression Results

	Dependent variable:
	Ri
Rm	1.235***
	(0.071)
Observations	520
R^2	0.370
Adjusted R^2	0.362
F Statistic	298.774***

Table 4: Fixed Effects Regression Using plm

Random Effects Model Regression Result

	Dependent variable: Ri
Rm	1.235***
	(0.071)
Constant	0.003***
	(0.001)
Observations	520
R^2	0.366
Adjusted R^2	0.365
F Statistic	299.657***

Table 5: Random Effects Regression

Graphical Analysis of Fixed Effects

```
yhat = fixed.dum$fitted.values
library(car)
scatterplot(yhat ~ Rm | Paneldata$company, boxplots
  = FALSE, xlab = "Market price", ylab = "yhat",
smooth = FALSE)
abline(lm(Ri ~ Rm), lwd = 3, col = "Dark Blue")
```

Model Comparison

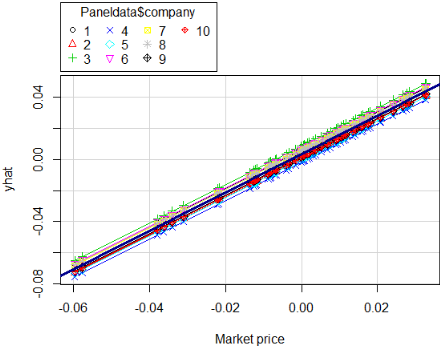

Figure 4: Significance of Fixed Effects Model

- Significance of Fixed Effects Model: Before testing for FE, first look at the graph. nt companies and the overall regression are not very large.
- The intercept differences among the regressions of different companies and the overall regression are not very large.

Tests to Compare Different Models

```
# Fixed Effects Model V.S. Pooled Regression
pFtest(fixed, pooling)

#Fixed Effects Model V.S. Random Effects Model
phtest(fixed, random)

#Random Effects Model V.S. Pooled Regression
plmtest(pooling)
```

Tests	F-test	Hausman Test	B-P LM test
Models	FE vs. Pooled	RE vs. FE	RE vs. Pooled
Statistics	0.54	$4.72e-25$	-1.14
P-value	0.84	1.00	0.25
H_0	$\mu_i = 0$	$E(\epsilon \mid x) = 0$	$\sigma_\mu^2 = 0$
Conclusion	Pooled	RE	Pooled

Table 6: Tests for 3 Models

- **Conclusion**: pooled regression model is the best one compared to fixed-effects model and random-effects model.

Outline

1. Basics ✓
2. Data Description ✓
3. Panel Data Analysis ✓
4. CAPM Regression
5. Estimation of Portfolio VaR

CAPM Regression

- CAMP Model:

$$E(R_i) - R_f = \beta_i[E(R_m) - R_f]$$

- GRS test statistic:

$$\frac{T(T-N-1)}{N(T-2)} \cdot \frac{\hat{\alpha}' \cdot \Sigma^{-1} \cdot \hat{\alpha}}{1+S_m^2} \overset{H_0}{\sim} F_{N,T-N-1}$$

- Sharp Ratio:

$$S_m = \frac{E(R_m - R_f)}{\sigma_m}$$

```r
xyplot(aser ~ amer | com, data = Paneldata, panel = function(x, y) {
panel.xyplot(x, y)
panel.abline(h = median(y), lty = 2, col = "gray")
panel.lmline(x, y, col = "red")
}, xlab = "average excess market return", ylab = "average excess stock return", xlim = c(0,
2), ylim = c(0, 5), main = "Excess market return and excess stock return")
```

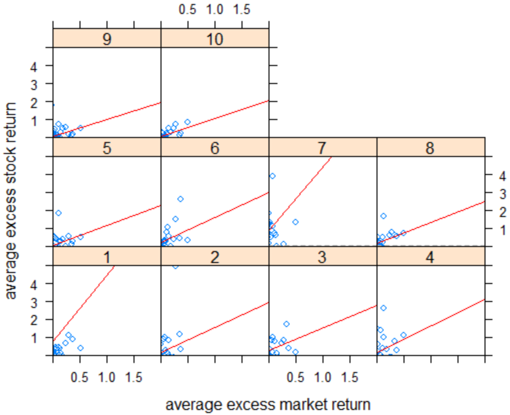

Figure 5: Excess market return and excess stock return

Seeming Unrelated Regression

```
Capm = read.csv("CAPM.csv")
Google = Capm$Google
Ebay = Capm$Ebay ...
eq1 = Google ~ amer eq2 = Ebay ~ amer... eq10 = ADP ~ amer
system = list(eq1 = eq1, eq2 = eq2, eq3 = eq3, eq4 = eq4, eq5 = eq5, eq6 = eq6, eq8 = eq8,
eq9 = eq9, eq10 = eq10)
library(Matrix)
library(zoo)
library(lmtest)
library(systemfit)
library(car)
sur = systemfit(system, method = "SUR", data = Capm)
summary(sur)
```

GRS Test for CAPM

```
Capm1 = read.csv("Capm1.csv")
sur1 = matrix(sur$coefficients, nrow = 2, ncol = 9)
alphas = sur1[1, ]
sig.hat = sur$residCov
sample.sr = function(x) {
    mu = mean(x)
    sg = sd(x)
    return(mu/sg)
}
```

```
amer = Capm1$amer
mkt.sr = sample.sr(amer)
GRS.stat = t(alphas) %*% solve(sig.hat, (alphas))/(1
    + mkt.sr^2)
T = 52
N = 9
F.stat = T * (T - N - 1) * GRS.stat/(N * (T - 2))
p.val = pf(F.stat, N, T - N - 1, 0, lower.tail = F)
p.val
```

- P-value=0.0027
- We can reject the null hypothesis that the intercepts are 0. However, this does not necessarily mean that CAPM is a wrong model.

Conclusion

Why can't we get 0 intercepts?

- This is not a well-diversified portfolio.
- Technical Explanations: Roll's Critique, data snooping, out-of-sample performance, measurement error, etc.
- Multiple risk factors:The CAPM suffers from omitted variables
- Irrational investor behavior: Investors overreact to news, etc.

Outline

1. Basics ✓
2. Data Description ✓
3. Panel Data Analysis ✓
4. CAPM Regression ✓
5. Estimation of Portfolio VaR

Portfolio Construction

- The portfolio is a price weighted portfolio, containing 1 unit of each stock during the whole period. The initial value of the portfolio is the sum of the prices of these 10 stocks at 2015/05/01.

VaR Estimation

$$VaR = (E(r_i + \Phi^{-1}(1-\alpha)\sigma_i) * Initial\ Value\ of\ Portfolio$$

- $Porfolio\ VaR \Leftarrow Portfolio\ Variance \Leftarrow$ $Variance\ covariance\ matrix\ of\ return\ of\ stocks$
- Assumption
  - Every stock has autocorrelation
  - Every stock has ARCH effect
- Covariance Estimation
  - Multivariate GARCH models suffer from huge parameter spaces. Correlation models have problem with higher dimensional systems.
  - Dynamic conditional correlation (DCC) model solves the trade off between model feasibility and flexibility.

DCC Model (Dynamic Conditional Correlation)

A vetor GARCH(1, 1) equation:

$h_t = a + A\varepsilon_{t-1}^{(2)} + Bh_{t-1},\ \varepsilon_{i,t} = h_{i,t}^{1/2} z_{i,t},\ z_t \sim ID(0, P_t)$

DCC of Engle (2002) and Engle and Sheppard (2001)

$$P_t = (Q_t \odot I_N)^{-1/2} Q_t (Q_t \odot I_N)^{-1/2}$$

$$Q_t = (1 - \alpha - \beta)Q + \alpha z_{t-1} z'_{t-1} + \beta Q_{t-1}$$

$$\alpha + \beta < 1 \text{ and } \alpha, \beta > 0$$

where Q is a sample covariance matrix of z_t.

Core Packages and Functions

- Package
  - tseries
  - ccgarch

- Function
  - arma
  - garch
  - dcc.estimation

```
for (i in 1:n) {
residual[, i] = matrix(residuals(arma(ts[, i], order = c(1, 0))))
coef[i, ] = matrix(coef(garch(residual[, i], order = c(1, 1), series = NULL)))
}
results = dcc.estimation(inia = a, iniA = A, iniB = B, ini.dcc = dcc.para, dvar = residual, model = "diagonal")
h = results$h
dcc = results$DCC
v = sqrt(diag(h[t - 1, ]))
R = matrix(data = dcc[1, ], nrow = n, ncol = n)
H = v %*% R %*% v
VaR95[j, ] = sqrt(t(W) %*% H %*% W) * alpha95 * value[j]
```

Estimation Settings

- Time period: 248 days
- Rolling screen: 200 days
- Number of VaR estimated: 48
- VaR level: 95%VaR, 99%VaR

Estimation results

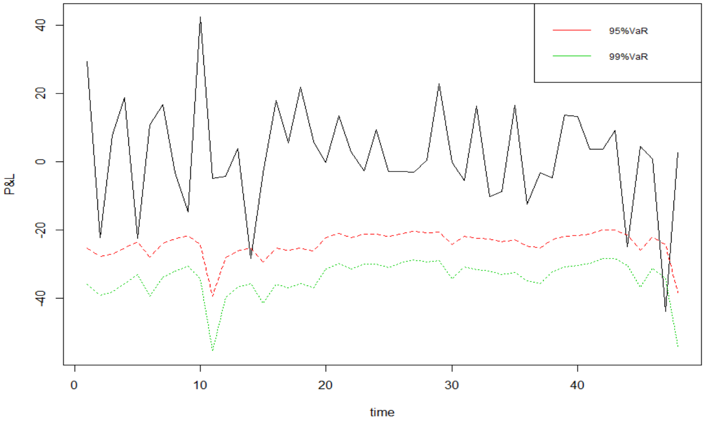

Figure 6: Portfolio P&L and estimated VaR

DW-test

If e_t is the residual associated with the observation at time t, then the test statistic is

$$d = \frac{\sum_{t=2}^{T}(e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2}$$

Engle's ARCH Test

Engle's ARCH test is a Lagrange multiplier test to assess the significance of ARCH effects

$$y_t = \mu_t + \varepsilon_t$$

where μ_t is the conditional mean of the process, and ε_t is an innovation process with mean zero.

Suppose the innovations are generated as

$$\varepsilon_t = \sigma_t z_t$$

where z_t is an independent and identically distributed process with mean 0 and variance 1. Thus,

$$E(\varepsilon_t \varepsilon_{t+h}) = 0$$

for all lags $h \neq 0$ and the innovations are uncorrelated.

Let H_t denote the history of the process available at time t. The conditional variance of y_t is

$$Var(y_t \mid H_{t-1}) = Var(\varepsilon_t \mid H_{t-1}) = E(\varepsilon_t^2 \mid H_{t-1}) = \sigma_t^2$$

Thus, conditional heteroscedasticity in the variance process is equivalent to autocorrelation in the squared innovation process.

Define the residual series $\qquad e_t = y_t - \mu_t$

The alternative hypothesis for Engle's ARCH test is autocorrelation in the squared residuals, given by the regression

$$H_a : e_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + \ldots + \alpha_m e_{t-m}^2 + \mu_t,$$

where μ_t is a white noise error process. The null hypothesis is

$$H_0 : \alpha_0 = \alpha_1 = \ldots = \alpha_m.$$

Core Packages and Functions

- DW-Test
  - Package: lmtest
  - Function: dwtest

- Engle's ARCH Test
  - Package: FinTS
  - Function: ArchTest

```
pvalue_dwtest = matrix(rep(0), ncol = 10)
dw_stat = matrix(rep(0), ncol = 10)
for (i in 1:10) {
    dw = dwtest(return[2:247, i] ~ return[1:246, i])
    pvalue_dwtest[, i] = dw$p.value
    dw_stat[, i] = dw$statistic
}

return1 = as.list(return)
archtest = lapply(return1, ArchTest, lags = 8,
  demean = TRUE)
a = as.data.frame(matrix((unlist(archtest)), c(5,
  10)))
LM_test_ARCH = rbind(a[1, ], a[3, ])}
```

Test Result

	GOOGL	EBAY	FB	YHOO	WU
dw_stat	1.98	1.99	1.99	2.00	1.99
p_value	0.44	0.47	0.48	0.52	0.49

	VRSN	NFLX	TSS	FIS	ADP
dw_stat	1.98	1.99	1.98	2.02	1.99
p_value	0.44	0.47	0.46	0.57	0.49

Table 7: DW-Test

Test Result

	GOOGL	EBAY	FB	YHOO	WU
LM_stat	1.33	0.42	8.18	5.12	1.19
p_value	0.99	0.99	0.41	0.74	0.99

	VRSN	NFLX	TSS	FIS	ADP
LM_stat	10.21	2.06	2.96	0.92	39.31
P_value	0.25	0.97	0.93	0.99	4.30E-06

Table 8: Engle's ARCH Test

Conclusion

- Not every stock has the characteristics which we assumed before. The VaR estimation might be problematic.
- Reason:
  - limitation of observations
  - only stocks from one industry, too similar
- Improvement:
  - enlarge time span of observations
  - more industry
  - change the length of rolling screen