



**AgEcon** SEARCH  
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

*The World's Largest Open Access Agricultural & Applied Economics Digital Library*

**This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.**

**Help ensure our sustainability.**

Give to AgEcon Search

AgEcon Search  
<http://ageconsearch.umn.edu>  
[aesearch@umn.edu](mailto:aesearch@umn.edu)

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

# THE STATA JOURNAL

## Editors

H. JOSEPH NEWTON  
Department of Statistics  
Texas A&M University  
College Station, Texas  
editors@stata-journal.com

NICHOLAS J. COX  
Department of Geography  
Durham University  
Durham, UK  
editors@stata-journal.com

## Associate Editors

CHRISTOPHER F. BAUM, Boston College  
NATHANIEL BECK, New York University  
RINO BELLOCCO, Karolinska Institutet, Sweden, and  
University of Milano-Bicocca, Italy  
MAARTEN L. BUIS, WZB, Germany  
A. COLIN CAMERON, University of California–Davis  
MARIO A. CLEVES, University of Arkansas for  
Medical Sciences  
WILLIAM D. DUPONT, Vanderbilt University  
PHILIP ENDER, University of California–Los Angeles  
DAVID EPSTEIN, Columbia University  
ALLAN GREGORY, Queen's University  
JAMES HARDIN, University of South Carolina  
BEN JANN, University of Bern, Switzerland  
STEPHEN JENKINS, London School of Economics and  
Political Science  
ULRICH KOHLER, University of Potsdam, Germany

FRAUKE KREUTER, Univ. of Maryland–College Park  
PETER A. LACHENBRUCH, Oregon State University  
JENS LAURITSEN, Odense University Hospital  
STANLEY LEMESHOW, Ohio State University  
J. SCOTT LONG, Indiana University  
ROGER NEWSON, Imperial College, London  
AUSTIN NICHOLS, Urban Institute, Washington DC  
MARCELLO PAGANO, Harvard School of Public Health  
SOPHIA RABE-HESKETH, Univ. of California–Berkeley  
J. PATRICK ROYSTON, MRC Clinical Trials Unit,  
London  
PHILIP RYAN, University of Adelaide  
MARK E. SCHAFFER, Heriot-Watt Univ., Edinburgh  
JEROEN WEESIE, Utrecht University  
IAN WHITE, MRC Biostatistics Unit, Cambridge  
NICHOLAS J. G. WINTER, University of Virginia  
JEFFREY WOOLDRIDGE, Michigan State University

## Stata Press Editorial Manager

LISA GILMORE

## Stata Press Copy Editors

DAVID CULWELL and DEIRDRE SKAGGS

The *Stata Journal* publishes reviewed papers together with shorter notes or comments, regular columns, book reviews, and other material of interest to Stata users. Examples of the types of papers include 1) expository papers that link the use of Stata commands or programs to associated principles, such as those that will serve as tutorials for users first encountering a new field of statistics or a major new technique; 2) papers that go “beyond the Stata manual” in explaining key features or uses of Stata that are of interest to intermediate or advanced users of Stata; 3) papers that discuss new commands or Stata programs of interest either to a wide spectrum of users (e.g., in data management or graphics) or to some large segment of Stata users (e.g., in survey statistics, survival analysis, panel analysis, or limited dependent variable modeling); 4) papers analyzing the statistical properties of new or existing estimators and tests in Stata; 5) papers that could be of interest or usefulness to researchers, especially in fields that are of practical importance but are not often included in texts or other journals, such as the use of Stata in managing datasets, especially large datasets, with advice from hard-won experience; and 6) papers of interest to those who teach, including Stata with topics such as extended examples of techniques and interpretation of results, simulations of statistical concepts, and overviews of subject areas.

The *Stata Journal* is indexed and abstracted by *CompuMath Citation Index*, *Current Contents/Social and Behavioral Sciences*, *RePEc: Research Papers in Economics*, *Science Citation Index Expanded* (also known as *SciSearch*, *Scopus*, and *Social Sciences Citation Index*).

For more information on the *Stata Journal*, including information for authors, see the webpage

<http://www.stata-journal.com>

**Subscriptions** are available from StataCorp, 4905 Lakeway Drive, College Station, Texas 77845, telephone 979-696-4600 or 800-STATA-PC, fax 979-696-4601, or online at

<http://www.stata.com/bookstore/sj.html>

**Subscription rates** listed below include both a printed and an electronic copy unless otherwise mentioned.

U.S. and Canada		Elsewhere	
<b>Printed &amp; electronic</b>		<b>Printed &amp; electronic</b>	
1-year subscription	\$ 98	1-year subscription	\$138
2-year subscription	\$165	2-year subscription	\$245
3-year subscription	\$225	3-year subscription	\$345
1-year student subscription	\$ 75	1-year student subscription	\$ 99
1-year university library subscription	\$125	1-year university library subscription	\$165
2-year university library subscription	\$215	2-year university library subscription	\$295
3-year university library subscription	\$315	3-year university library subscription	\$435
1-year institutional subscription	\$245	1-year institutional subscription	\$285
2-year institutional subscription	\$445	2-year institutional subscription	\$525
3-year institutional subscription	\$645	3-year institutional subscription	\$765
<b>Electronic only</b>		<b>Electronic only</b>	
1-year subscription	\$ 75	1-year subscription	\$ 75
2-year subscription	\$125	2-year subscription	\$125
3-year subscription	\$165	3-year subscription	\$165
1-year student subscription	\$ 45	1-year student subscription	\$ 45

Back issues of the *Stata Journal* may be ordered online at

<http://www.stata.com/bookstore/sjj.html>

Individual articles three or more years old may be accessed online without charge. More recent articles may be ordered online.

<http://www.stata-journal.com/archives.html>

The *Stata Journal* is published quarterly by the Stata Press, College Station, Texas, USA.

Address changes should be sent to the *Stata Journal*, StataCorp, 4905 Lakeway Drive, College Station, TX 77845, USA, or emailed to [sj@stata.com](mailto:sj@stata.com).



Copyright © 2013 by StataCorp LP

**Copyright Statement:** The *Stata Journal* and the contents of the supporting files (programs, datasets, and help files) are copyright © by StataCorp LP. The contents of the supporting files (programs, datasets, and help files) may be copied or reproduced by any means whatsoever, in whole or in part, as long as any copy or reproduction includes attribution to both (1) the author and (2) the *Stata Journal*.

The articles appearing in the *Stata Journal* may be copied or reproduced as printed copies, in whole or in part, as long as any copy or reproduction includes attribution to both (1) the author and (2) the *Stata Journal*.

Written permission must be obtained from StataCorp if you wish to make electronic copies of the insertions. This precludes placing electronic copies of the *Stata Journal*, in whole or in part, on publicly accessible websites, file servers, or other locations where the copy may be accessed by anyone other than the subscriber.

Users of any of the software, ideas, data, or other materials published in the *Stata Journal* or the supporting files understand that such use is made without warranty of any kind, by either the *Stata Journal*, the author, or StataCorp. In particular, there is no warranty of fitness of purpose or merchantability, nor for special, incidental, or consequential damages such as loss of profits. The purpose of the *Stata Journal* is to promote free communication among Stata users.

The *Stata Journal* (ISSN 1536-867X) is a publication of Stata Press. Stata, **stata**, Stata Press, Mata, **mata**, and NetCourse are registered trademarks of StataCorp LP.

# Financial portfolio selection using the multifactor capital asset pricing model and imported options data

Mehmet F. Dicle  
Loyola University New Orleans  
New Orleans, LA  
mfdicle@loyno.edu

**Abstract.** Diversification and portfolio selection are integral parts of a finance curriculum. In this article, a multifactor capital asset pricing model is fit for components of the Dow Jones Composite Index using data from Yahoo! Finance. Along with the capital asset pricing model's Beta, other statistics that are common criteria for portfolio selection are calculated: historic standard deviation (total risk), total return, average daily return, and Sharpe and Treynor measures. Two new commands are introduced, `fetchcomponents` and `fetchportfolio`, that automate the entire process. A third new command, `fetchyahoooptions`, is provided to download and parse equity options data from Yahoo! Finance webpages and, optionally, to calculate the implied volatilities for the downloaded options.

**Keywords:** dm0070, `fetchcomponents`, `fetchportfolio`, `fetchyahoooptions`, finance, financial data, multifactor capital asset pricing model, Beta, diversification, portfolio selection, Sharpe, Treynor, options, implied volatility

## 1 Introduction

Diversification, portfolio selection, hedging, and the capital asset pricing model (CAPM) (Sharpe 1964; Lintner 1965) are integral parts of an undergraduate and a graduate finance curriculum. Most textbooks provide detailed explanations about how to diversify, how to evaluate different financial securities, and how to fit a CAPM. Considering the importance of “learning by doing” or a “hands-on” approach, financial calculations and CAPM estimations are performed in classrooms and given as assignments. There are abundant resources about these financial calculations and CAPM estimations mostly using Microsoft Excel. However, there seem to be limited resources for students to do financial calculations and CAPM estimations using an econometric software such as Stata. Considering the speed and ease of repeatability of Stata, this is a drawback in the finance curriculum: teaching should focus more on theory and interpretation of results than on long and tiresome steps of calculations and estimations.

Microsoft Excel is one of the primary software in the industry and an asset for finance students, who should be able to use it to fit a CAPM and other financial calculations. However, automated tasks such as fitting a CAPM for multiple stocks would require them to learn either Excel macro programming or some type of econometric software.

Recently, Stata has become the popular choice of academics and students, perhaps because of its user friendliness and abundant resources.

This article is of interest to finance instructors, students, and investors.<sup>1</sup> It entails lecture notes to teach different criteria for portfolio selection, diversification, and options. It shows how some of the most common statistics are calculated. It provides fast and easy commands to repeat these tasks during a study session or a lecture using Stata. It enables finance instructors to assign projects using real-life data and to spend time on interpretations and on methods proportional to their importance. The provided procedures are also useful for investors (for educational purposes): using real financial data, they can compare their investment choices to achieve portfolio objectives.

The initial step is to obtain a list of stocks that make up the Dow Jones Composite Index. For this task, a new Stata command is used: **fetchcomponents**. The second step is to download historic prices for these stocks using the Stata command **fetchyahooquotes** (Dicle and Levendis 2011). Using the same command, I download Fama–French factors to fit a multifactor CAPM (Fama and French 1992, 1993). The third step is to calculate the average daily returns, total returns, standard deviation of daily returns, Sharpe (1970) measure, and Treynor (1965) measure. In this step, multifactor CAPM following Fama and French (1992, 1993) is also estimated for each stock. Another new Stata command is used to automate this process: **fetchportfolio**. The fourth step is to interpret the results. The final step is to introduce the idea of hedging portfolio risk by using **fetchyahooptions** to download options data for a few stocks. Their implied volatilities are calculated and graphed.

## 2 Components of Dow Jones Composite Index

A list of stocks that make up the Dow Jones Composite Index (DJA) is provided by Yahoo! Finance.<sup>2</sup> A new command, **fetchcomponents**, downloads this list to a Stata variable that contains the ticker symbols.

---

1. This article and associated Stata codes are for educational use. There is no direct or implied financial advice. While every effort is made for accuracy and reliability, the data and the results may not be accurate or reliable.

2. Available through <http://finance.yahoo.com/q/cp?s=DJA+Components>.

```
. fetchcomponents, symbol(^DJA) page(1)
Page: 0
Page: 1
. describe
Contains data
  obs:          65
  vars:          1
  size:         260
```

---

variable name	storage type	display format	value label	variable label
Symbol	str4	%9s		

---

```
Sorted by: Symbol
Note: dataset has changed since last saved
```

## 2.1 Definition

`fetchcomponents` downloads the list of components for an index. A list of stocks and components is provided by Yahoo! Finance 50 at a time. Each page of 50 is defined by the option `page`.

## 2.2 Syntax

```
fetchcomponents, symbol(string) [page(#)]
```

## 2.3 Options

`symbol(string)` specifies the symbol of the index for which the user wants to download the components (that is, `^NYA`). There can be only one symbol defined, which must be an index. `symbol()` is required.

`page(#)` specifies the number of pages for the list of components (starts from zero). Because the list of components is provided in batches of 50, it may take several pages to download the entire list (that is, for `^NYA`, there are 0 to 38 pages).

## 3 Historic stock prices

Daily stock prices are downloaded from Yahoo! Finance.<sup>3</sup> Daily Fama–French factors are downloaded from Kenneth R. French’s website.<sup>4</sup> The Stata code for downloading and preparing the data for the study is `fetchyahooquotes`.

3. Available through <http://finance.yahoo.com/>.

4. Available through [http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data\\_library.html](http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html).

```
. levelsof Symbol, local(tickers) clean
AA AEP AES ALK AXP BA BAC CAT CHRW CNP CNW CSCO CSX CVX D DAL DD DIS DUK ED EIX
> EXC EXPD FDX FE GE GMT HD HPQ IBM INTC JBHT JBLU JNJ JPM KEX KO KSU LSTR LUV
> MATX MCD MMM MRK MSFT NEE NI NSC PCG PEG PFE PG R SO T TRV UAL UNH UNP UPS
> UTX VZ WMB WMT XOM

. fetchyahoquotes `tickers`, freq(d) chg(per) start("01jan2010")
> end("31dec2011") ff3
Fama/French daily factors are downloaded from `Kenneth R. French - Data Library`.
AA is downloaded.
AEP is downloaded.
AES is downloaded.

(output omitted)

WMB is downloaded.
WMT is downloaded.
XOM is downloaded.

time variable: date, 04jan2010 to 30dec2011, but with gaps
delta: 1 day
(0 observations deleted)
(0 observations deleted)

. describe
```

Contains data from temporary\_AA.dta

```
obs:      504
vars:      133                      25 Apr 2013 10:41
size:      278,208
```

variable name	storage type	display format	value label	variable label
date	double	%td		
adjclose_AA	float	%9.0g		Adj Close
per_AA	float	%9.0g		Percentage change for AA
adjclose_AEP	float	%9.0g		Adj Close
per_AEP	float	%9.0g		Percentage change for AEP
adjclose_AES	float	%9.0g		Adj Close
per_AES	float	%9.0g		Percentage change for AES
(output omitted)				
adjclose_WMB	float	%9.0g		Adj Close
per_WMB	float	%9.0g		Percentage change for WMB
adjclose_WMT	float	%9.0g		Adj Close
per_WMT	float	%9.0g		Percentage change for WMT
adjclose_XOM	float	%9.0g		Adj Close
per_XOM	float	%9.0g		Percentage change for XOM
ff3_Mkt_RF	double	%10.0g		
ff3_SMB	double	%10.0g		
ff3_HML	double	%10.0g		
ff3_RF	double	%10.0g		

Sorted by: date

Note: dataset has changed since last saved

## 4 Calculating financial statistics and fitting a multifactor CAPM

The formula for fitting a multifactor CAPM (Fama and French 1992, 1993) is as follows:

$$r_{i,t} - r_{f,t} = \alpha + \beta(r_{m,t} - r_{f,t}) + \gamma(SMB_t) + \theta(HML_t) + \epsilon_t \quad (1)$$

A new command, `fetchportfolio`, fits the multifactor CAPM and other financial statistics for all the stocks that constitute the DJIA. CAPM estimations are based on the daily percentage change of dividend- and split-adjusted closing prices.

```
. fetchportfolio `tickers`, year(2010 2011)
Fama/French daily factors are downloaded from `Kenneth R. French - Data Library`.
AA is downloaded.
AEP is downloaded.
AES is downloaded.
(output omitted)
WMB is downloaded.
WMT is downloaded.
XOM is downloaded.
      time variable:  date, 04jan2010 to 30dec2011, but with gaps
                   delta:  1 day
(3965 observations deleted)
(140 observations deleted)
. describe
Contains data from temp_00001_2010.dta
   obs:           64
  vars:           27
 size:          6,912
25 Apr 2013 10:49
```

variable name	storage type	display format	value label	variable label
Symbol	str4	%9s		
Beta_2010	float	%9.0g		Multifactor CAPM Beta
R2_2010	float	%9.0g		Multifactor CAPM R-squared
Sum_2010	float	%9.0g		Total return for the year (includes dividend yield).
Sd_2010	float	%9.0g		Standard deviation of daily returns for the year.
Mean_2010	float	%9.0g		Average daily return for the year.
Mean_rf_2010	float	%9.0g		Average daily return for the year minus average risk-free rate.
Dividends_2010	float	%9.0g		Dividend yield for the year.
Beta_2011	float	%9.0g		Multifactor CAPM Beta
R2_2011	float	%9.0g		Multifactor CAPM R-squared
Sum_2011	float	%9.0g		Total return for the year (includes dividend yield).
Sd_2011	float	%9.0g		Standard deviation of daily returns for the year.
Mean_2011	float	%9.0g		Average daily return for the year.
Mean_rf_2011	float	%9.0g		Average daily return for the year minus average risk-free rate.
Dividends_2011	float	%9.0g		Dividend yield for the year.
Mean_sd_2010	float	%9.0g		Mean daily return / Standard



Sharpe_2010	float	%9.0g	deviation of daily returns (Mean daily return - mean risk-free rate) / Standard deviation of daily returns
Sum_sd_2010	float	%9.0g	Total return (includes dividend yield) / Standard deviation of daily returns
Mean_Beta_2010	float	%9.0g	Mean daily return / Multifactor CAPM Beta
Treynor_2010	float	%9.0g	(Mean daily return - mean risk-free rate) / Multifactor CAPM Beta
Sum_Beta_2010	float	%9.0g	Total return (includes dividend yield) / Multifactor CAPM Beta
Mean_sd_2011	float	%9.0g	Mean daily return / Standard deviation of daily returns
Sharpe_2011	float	%9.0g	(Mean daily return - mean risk-free rate) / Standard deviation of daily returns
Sum_sd_2011	float	%9.0g	Total return (includes dividend yield) / Standard deviation of daily returns
Mean_Beta_2011	float	%9.0g	Mean daily return / Multifactor CAPM Beta
Treynor_2011	float	%9.0g	(Mean daily return - mean risk-free rate) / Multifactor CAPM Beta
Sum_Beta_2011	float	%9.0g	Total return (includes dividend yield) / Multifactor CAPM Beta

---

Sorted by: Symbol

Note: dataset has changed since last saved

## 4.1 Definition

`fetchportfolio` estimates and calculates financial statistics to compare financial securities for portfolio selection. `fetchyahooquotes` is needed for `fetchportfolio` to run.

## 4.2 Syntax

`fetchportfolio namelist, year(numlist)`

*namelist* is a list of ticker symbols for which the statistics are calculated and a CAPM is fit. Symbols are separated by spaces.

## 4.3 Option

`year(numlist)` is a list of years for which the statistics are calculated and a CAPM is fit. Years are separated by spaces.

## 5 Interpreting results

### 5.1 Multifactor CAPM Beta

```
. list Symbol Beta* R2*
```

	Symbol	Bet-2010	Bet-2011	R2_2010	R2_2011
1.	AA	1.405193	1.545471	.5650427	.7981184
2.	AEP	.7943012	.6373942	.5562428	.5718821
3.	AES	1.257982	1.262856	.5833274	.7186341
4.	ALK	1.468361	.9333242	.4253781	.5818244
5.	AXP	1.250965	1.164087	.6207132	.7460121

(output omitted)

61.	VZ	.6491743	.6290471	.3628424	.5334482
62.	WMB	1.230148	1.400854	.6391539	.6731739
63.	WMT	.6216294	.5254323	.3249495	.4530556
64.	XOM	.9589896	.9934663	.7284095	.7589387

The analysis is performed on the list of stocks that make up the DJIA. There are a total of 65 stocks. For the purpose of compactness, results are provided for 10 of these 65 stocks. In the table above, there are two columns for each year: **Beta** and **R2**. **Beta** is the Beta estimated with (1). **R2** refers to the  $R$ -squared for the same regression. A CAPM is fit for each stock separately for 2010 and 2011. Higher Beta<sup>5</sup> means higher market risk for each stock. Higher  $R$ -squared means more of the variation of daily stock returns is explained by the controlled independent variables: market risk, small-minus-big, and high-minus-low. As Beta increased from 2010 to 2011, we can conclude that the stocks' market risk increased. As  $R$ -squared increased from 2010 to 2011, we can conclude that the uncontrolled factors (that is, company-specific risks) became less important factors to explain the variation of daily returns.

---

5. Absolute value of the Beta.

## 5.2 Total return

```
. list Symbol Sum_20*
```

	Symbol	Sum_2010	Sum_2011
1.	AA	-5.674569	-43.91911
2.	AEP	13.75836	25.34019
3.	AES	-10.90641	-5.622488
4.	ALK	62.59323	29.71666
5.	AXP	8.655738	12.01655

(output omitted)

61.	VZ	21.74054	21.91481
62.	WMB	20.21237	40.64208
63.	WMT	4.188481	15.37066
64.	XOM	11.46432	18.8829

Total return is the percentage change in split- and dividend-adjusted closing prices from the first day to the last day of the period plus the dividend yield for the year. The figures are in percentages (that is, AA's return for 2010 was  $-5.67\%$ , whereas it was  $-43.92\%$  for 2011).

## 5.3 Total risk

```
. list Symbol Sd_20*
```

	Symbol	Sd_2010	Sd_2011
1.	AA	2.349188	2.711182
2.	AEP	1.045383	1.196054
3.	AES	2.305197	2.168423
4.	ALK	2.611011	2.442996
5.	AXP	2.066332	1.918625

(output omitted)

61.	VZ	1.043751	1.217522
62.	WMB	2.08644	2.435364
63.	WMT	.8757704	1.045985
64.	XOM	1.132489	1.597168

Total risk is the standard deviation of daily returns. Standard deviation of daily returns is based on the daily percentage change of dividend- and split-adjusted closing prices. It is interpreted as the combination of systemic and unsystemic risks.

## 5.4 Sharpe measure

```
. list Symbol Sharpe_20*
```

	Symbol	Shar-2010	Shar-2011
1.	AA	-.0001682	-.0693872
2.	AEP	.0343362	.0683895
3.	AES	-.008775	.0056541
4.	ALK	.0870246	.0578227
5.	AXP	.0226696	.0323365

(output omitted)

61.	VZ	.0560282	.0605425
62.	WMB	.0400964	.0636963
63.	WMT	.0114429	.0543649
64.	XOM	.0340558	.0504581

The formula for the Sharpe measure is as follows:  $\text{Sharpe} = (\text{mean daily return} - \text{mean risk-free rate}) / \text{standard deviation of daily returns}$ . Mean daily returns are based on the daily percentage change of dividend- and split-adjusted closing prices. Therefore, the higher the ratio, the higher the mean return per level of standard deviation (total risk).

## 5.5 Treynor measure

```
. list Symbol Treynor_20*
```

	Symbol	Trey-2010	Trey-2011
1.	AA	-.0002812	-.1217243
2.	AEP	.04519	.1283311
3.	AES	-.0160798	.0097086
4.	ALK	.1547454	.1513523
5.	AXP	.0374454	.0532964

(output omitted)

61.	VZ	.0900828	.1171802
62.	WMB	.068007	.1107351
63.	WMT	.0161211	.1082249
64.	XOM	.0402171	.0811201

The formula for the Treynor measure is as follows: Treynor = (mean daily return – mean risk-free rate)/multifactor CAPM Beta. As in the Sharpe measure, the higher the ratio, the higher the mean return for the level of Beta (market risk).

## 6 Importing options data

Yahoo! Finance provides free financial data for public use. While historical prices for most financial assets, as well as some important statistics, can be downloaded using Yahoo! Finance’s application programming interface,<sup>6</sup> some data are only available through webpages. These important data can be accessed via web browsers. However, to access the data as Stata-usable data, one needs to parse these webpages.<sup>7</sup> Stata has a powerful and fast programming language: Mata. Even though it is intended as a matrix programming language, it has hypertext transfer protocol support, regular expressions, and extensive string functions. It allows the newly introduced command **fetchyahooptions** to fetch the Yahoo! Finance options page, to parse the page, and to process its contents to make them usable as Stata data. **fetchyahooptions** also calculates implied volatility for downloaded options using Black and Scholes’s (1973) option pricing formula following (2), (3), (4), and (5).

Unlike equity prices, options data are not easily accessible for everyone as usable data. They are an important free service that Yahoo! Finance offers for public use. Options data are important for finance lecturers to use in their classes and teaching notes. They are important for researchers of financial derivatives. Although options data are only current and historical time series are not available, they can be accessed daily and stored to create a time series. Options data are also important for investors (for educational purposes). The implied volatility allows investors to have a sense of expected volatility in the market and may allow investors to predict market direction.

### 6.1 Syntax

```
fetchyahooptions namelist, m(string) [iv(real) ]
```

*namelist* specifies a list of ticker symbols for which the options are parsed and downloaded from Yahoo! Finance’s options webpage. Symbols are separated by spaces.

### 6.2 Options

**m**(*string*) specifies the month in which the options expire (that is, 2011–12). Multiple maturity months can be included (that is, 2011–12 2012–01). **m**() is required.

---

6. Stata commands **fetchyahoquotes** and **fetchyahokeystats** use Yahoo! Finance’s application programming interface to download historical prices and key statistics.

7. Parsing Hypertext Markup Language (HTML) pages is a common practice; some programming languages, such as PHP and Java script, provide extensive language support. There is a regular expressions parsing language that is adopted by most web programming languages.

`iv(real)` specifies the calculated implied volatility using Black and Scholes's (1973) option pricing formula (that is, `iv(0.0001)`) following (2), (3), (4), and (5). It uses a trial-and-error method to loop through levels of volatilities to calculate a call or put option price that matches the ask price. User-supplied increment is used for the loop. As the increment gets smaller, the loop will take longer. Implied volatility is calculated separately for each strike price.

$$d_1 = \frac{\ln(S_0/K) + \left\{ (r_f + \frac{\sigma^2}{2})T \right\}}{\sigma\sqrt{T}} \quad (2)$$

$$d_2 = d_1 - \sigma\sqrt{T} \quad (3)$$

$$c = S_0 N(d_1) - K e^{-r_f T} N(d_2) \quad (4)$$

$$p = K e^{-r_f T} N(-d_2) - S_0 N(-d_1) \quad (5)$$

In these equations,  $S_0$  refers to the spot price of the underlying security,  $K$  to the strike price,  $r_f$  to the risk-free rate,<sup>8</sup>  $\sigma$  to the standard deviation, and  $T$  to the years to maturity.

### 6.3 HTML source code to Stata data transformation

Yahoo! Finance<sup>9</sup> provides current prices of options for individual stocks through HTML pages. `fetchyahoooptions` uses Mata to access Yahoo! Finance options pages, parse them into string variables, and turn them into usable Stata data. The following are some of the processes that `fetchyahoooptions` uses.

The following Mata function is used to get the HTML source code from the web as a string.

```
string file_get_contents (string scalar raw)
{
    fh = fopen(raw, "r")
    raw=""
    while ((line=fget(fh))!=J(0,0,"")) {
        raw=raw+line
    }
    fclose(fh)
    return (raw)
}
```

---

8. For the risk-free rate, `fetchyahoooptions` downloads the current `^IRX`, the 13-week U.S. Treasury Bill yield index, from Yahoo! Finance.

9. For example, <http://finance.yahoo.com/q/op?s=IBM+Options>.

The following Mata function parses the current price of the underlying asset for the option.

```

string get_price (string scalar raw, string scalar ticker)
{
    bas_pos = strpos(strlower(raw), "<span id=" + char(34) + "yfs_l10_"
+ strlower(ticker) + char(34) + ">") + strlen(ticker) + 20

    if (bas_pos<100) bas_pos = strpos(strlower(raw), "<span id=" + char(34)
+ "yfs_l84_" + strlower(ticker) + char(34) + ">") + strlen(ticker) + 20
    if (bas_pos>100) {
        output = substr (raw, bas_pos, .)
        son_pos = strpos(strlower(output), "</span>")
        output = substr (output, 1, son_pos-1)
    }
    if (bas_pos<100) output = cat("http://download.finance.yahoo.com/d/quotes.csv?s="
+ ticker + "&f=l1")
    return (output)
}

```

A check is necessary to see whether the options table exists.

```

real check_table (string scalar raw, string scalar aranan)
{
    kalan = strpos(strlower(raw), strlower(aranan))
    return (kalan)
}

```

Options data are within the HTML table. At this point, the options table needs to be parsed.

```

string get_table (string scalar raw, string scalar tag, string scalar aranan)
{
    kalan = strlower(raw)
    while (strpos(kalan, "<" + strlower(tag)) < strpos(kalan, strlower(aranan))) {
        kalan = substr (kalan, strpos(kalan, "<" + strlower(tag)) + 2
+ strlen(tag), .)
    }
    kalan = "<" + strlower(tag) + " " + kalan
    son_pos = strpos(kalan, "</" + strlower(tag))
    kalan=substr(kalan, 1, son_pos + 2 + strlen(tag))
    return (kalan)
}

```

The above functions are called in the following order.

```
void get_options (string scalar symbol, string scalar month)
{
  icerik = file_get_contents("http://finance.yahoo.com/q/op?s=" + symbol
+ "&m=" + month)
  price = get_price(icerik,symbol)
  isthere = check_table (icerik,"Strike")
  if (isthere) {
    tablo = get_table (icerik,"table","Strike")
    tablo2 = get_table2 (icerik,"table","Strike")
    bos=table_data (tablo + tablo2)
  }
  stata("gen Price = " + price)
  icerik_rf = file_get_contents("http://finance.yahoo.com/q?s=~IRX")
  price_rf = get_price(icerik_rf,"~IRX")
  stata("gen IRX = " + price_rf)
}
```

The remaining string is parsed for individual HTML tags such as <td>, <tr>, <br>, etc. Parsing is lengthy and can be accessed through the `fetchyahoooptions.ado` file. Parsing is not shown here to conserve space.

The entire HTML source code downloaded from Yahoo! Finance is parsed into a single string using Mata. The string table that contains the options data has HTML tag <td>, which can be used as a line break that can then be converted into a Stata observation. The following Stata code is used for this string split.

```
split myvar, parse("</td>") gen(mfd)
```

The resulting dataset contains eight variables: **Strike**, **Symbol**, **Last**, **Change**, **Bid**, **Ask**, **Volume**, and **Open\_Interest**.

## 6.4 Using fetchyahoooptions to fetch equity options data

### ▷ Example

Here options data are downloaded for IBM and GOOG using `fetchyahoooptions` for the closest maturity (April 2013) and the next closest maturity (May 2013). The program will also calculate the implied volatility.

```
. fetchyahoooptions GOOG IBM, m(2013-04 2013-05 2013-06) iv(0.0001)
Options data for GOOG (2013-04) are downloaded.
Options data for GOOG (2013-05) are downloaded.
Options data for GOOG (2013-06) are downloaded.
Options data for IBM (2013-04) are downloaded.
Options data for IBM (2013-05) are downloaded.
Options data for IBM (2013-06) are downloaded.
```



```
. twoway (line IV Strike) if (Type=="Call") & (Underlying=="GOOG") &
> (Maturity==date("22jun2013","DMY"))
```

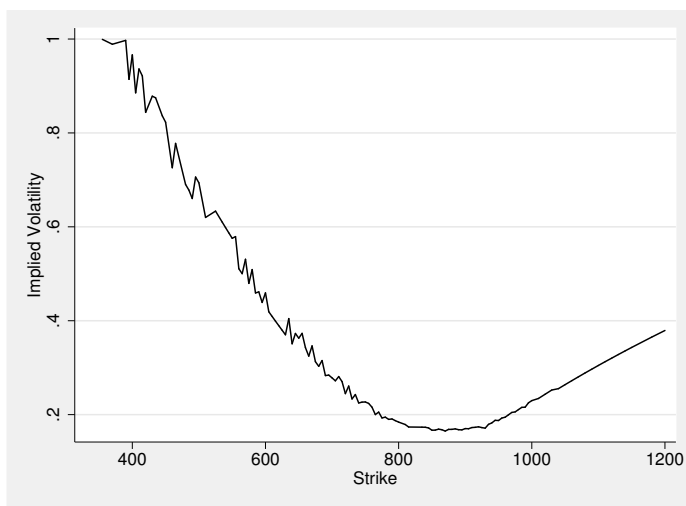


Figure 1. Implied volatility for each strike price

## 7 Acknowledgments

I thank Jiad Alqotob, College of Business, Loyola University, New Orleans, for his valuable comments during the creating of the `fetchyahooptions` command. I also thank Ashton Verdery, Department of Sociology, University of North Carolina at Chapel Hill, for suggesting an improvement (implemented as suggested) in the `fetchyahooptions` command to make it more reliable.

## 8 References

- Black, F., and M. Scholes. 1973. The pricing of options and corporate liabilities. *Journal of Political Economy* 81: 637–654.
- Dicle, M. F., and J. Levendis. 2011. Importing financial data. *Stata Journal* 11: 620–626.
- Fama, E. F., and K. R. French. 1992. The cross-section of expected stock returns. *Journal of Finance* 47: 427–465.
- . 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33: 3–56.

Lintner, J. 1965. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *Review of Economics and Statistics* 47: 13–37.

Sharpe, W. F. 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance* 19: 425–442.

———. 1970. *Portfolio Theory and Capital Markets*. New York: McGraw–Hill.

Treynor, J. L. 1965. How to rate management of investment funds. *Harvard Business Review* 43: 63–75.

#### **About the author**

Mehmet F. Dicle is an assistant professor of finance at Loyola University New Orleans.