The Stata Journal

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	
Printed & electronic		**Printed & electronic**	
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 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
Electronic only		**Electronic only**	
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.

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.

The Stata Journal (2014)
14, Number 2, pp. 381–388

Importing Chinese historical stock market quotations from NetEase

Xuan Zhang
School of Statistics and Mathematics
Zhongnan University of Economics and Law
Wuhan, China
zhangx@znufe.edu.cn

Chuntao Li
School of Finance
Zhongnan University of Economics and Law
Wuhan, China
chtl@znufe.edu.cn

Abstract. In this article, we describe a command, `cntrade`, that imports historical stock quotations from NetEase, a key Chinese website for stocks listed on China's two exchanges.

Keywords: dm0074, cntrade, stock quotes, China, NetEase

1 Introduction

Similar to Yahoo! Finance, NetEase is a key disseminator of financial information in China.[1] This article describes `cntrade`, a command we developed to automatically download historical trading quotes. For example, users can download the daily opening and closing prices and the trading volume for China Merchants Bank and China Southern Airlines starting from their initial public-offering prices and ranging to their most recent closings.

The Chinese stock market has increasingly drawn attention worldwide. Although several databases offer Chinese financial information, many colleges and universities do not subscribe to any of them. Even the institutions that do subscribe to databases allow users to access them from only a few terminals or restrict users to a limited number of parallel downloads. These restrictions inconvenience both the institutions and their users. Furthermore, the trading quotes must be downloaded manually, which is tedious and time consuming. After downloading the quotes, users must unzip the files, transfer them from Excel to text, and clean, merge, and append the data. The Chinese language also creates a barrier for non-Chinese researchers. Our command, `cntrade`, overcomes all of these difficulties and makes accessing Chinese stock trading quotes easy, quick, and free of charge. It is especially helpful to teachers and students interested in the

1. NetEase is a NASDAQ-listed Chinese company in the form of an American depository receipt; its ticker is NTES.

Chinese stock market. It may also encourage more students in China to learn to use Stata.

2 The cntrade command

2.1 Syntax

`cntrade` *codelist* [`, path(`*foldername*`)`]

codelist is a list of stock codes to download from NetEase. The stock codes should be separated by spaces. For each stock code, a separate `.dta` file will be saved to the user's local drive. The stock code will be used as the filename, and `.dta` will be the extension name. In China, every stock code is a six-digit number theoretically ranging from 000001 to 999999; however, there are fewer than 2,500 listed firms, which means many of the numbers in this range are invalid. `cntrade` will distinguish any nonvalid code and issue an error message. In the Shenzhen Stock Exchange, all stock codes are numbered as less than 10,000, such as 000001, 000002, and 002007, but users may omit all leading zeros. For instance, 000002 stands for China Vanke Co., Ltd., but investors may simply use 2 as the stock code instead of entering the entire number. Therefore, when people input codes of fewer than six digits, `cntrade` will automatically insert the leading zeros to make the necessary six-digit code. For example, a code of 2 or 02 will be treated the same as 000002.

2.2 Option

`path(`*foldername*`)` specifies the folder where the output `.dta` files will be saved. *foldername* can be an existing folder or a new folder. If *foldername* does not exist, `cntrade` will create it automatically. If `path()` is not specified, the output will be saved to the current working directory. Specifying `path()` is strongly recommended.

2.3 Variables

When a download is finished, a typical output file contains the following 12 variables:

- `stkcd`: The stock codes stored as long integers. The `cntrade` command automatically omits the leading zeros for the firms listed on the Shenzhen Stock Exchange.
- `stknme`: The company name in Chinese.
- `clsprc`: The daily closing price for each trading day.
- `hiprc`: The daily highest price.
- `lowprc`: The daily lowest price.
- `opnprc`: The daily opening price.

- `rit`: The daily return as a percentage. We suggest the user divide this variable by 100 to convert it to the original value.
- `turnover`: The daily turnover rate as a percentage.
- `volume`: The number of shares traded each day.
- `transaction`: The transaction volume in terms of Chinese renminbi (RMB).
- `mktcap`: The market capitalization at the end of each trading day.
- `date`: The trading date.

3 Examples

3.1 Download historical transaction data for one company

Suppose we want to download the historical trading data for a company whose stock code is 000008. Here we set Stata's working directory to `d:\` and do not specify the `path()` option of `cntrade`, so `cntrade` will save the output file in Stata's current working directory (`d:\` in this case).

```
. cd d:\
d:\
. cntrade 000008
d:
(15 vars, 5311 obs)
(165 observations deleted)
(note: file d:/000008.dta not found)
file d:/000008.dta saved
. describe
Contains data from d:/000008.dta
  obs:         5,146
 vars:            12                          26 Feb 2014 17:41
 size:       272,738
-------------------------------------------------------------------------------
              storage  display     value
variable name   type   format      label      variable label
-------------------------------------------------------------------------------
stkcd           byte   %10.0g                 stock code
stknme          str8   %9s                    Stock Name
clsprc          float  %9.0g                 closing price
hiprc           float  %9.0g                 highest price
lowprc          float  %9.0g                 lowest price
opnprc          float  %9.0g                 opening price
rit             float  %9.0g                 daily return
turnover        float  %9.0g                 turnover rate
volume          long   %12.0g                trading volume
transaction     float  %9.0g                 trading amount in RMB
mktcap          double %10.0g                total market capitalization
date            float  %dCY_N_D              trading date
-------------------------------------------------------------------------------
Sorted by:  date
```

```
. list stkcd date opnprc hiprc lowprc clsprc rit in 1/5
```

	stkcd	date	opnprc	hiprc	lowprc	clsprc	rit
1.	8	1992 05 07	19.9	20.65	19.9	20.65	106.5
2.	8	1992 05 08	20.6	20.6	17.75	17.75	-14.0436
3.	8	1992 05 11	18	18	16.8	17.95	1.1268
4.	8	1992 05 12	18.1	19	17	18.55	3.3426
5.	8	1992 05 13	18.55	18.55	17.8	18.2	-1.8868

```
. cd d:\
d:\
. cntrade 8
d:
(15 vars, 5311 obs)
(165 observations deleted)
file d:/000008.dta saved
. describe
Contains data from d:/000008.dta
  obs:         5,146
 vars:            12                          26 Feb 2014 17:41
 size:       272,738
```

variable name	storage type	display format	value label	variable label
stkcd	byte	%10.0g		stock code
stknme	str8	%9s		Stock Name
clsprc	float	%9.0g		closing price
hiprc	float	%9.0g		highest price
lowprc	float	%9.0g		lowest price
opnprc	float	%9.0g		opening price
rit	float	%9.0g		daily return
turnover	float	%9.0g		turnover rate
volume	long	%12.0g		trading volume
transaction	float	%9.0g		trading amount in RMB
mktcap	double	%10.0g		total market capitalization
date	float	%dCY_N_D		trading date

```
Sorted by:  date
. list stkcd date opnprc hiprc lowprc clsprc rit in 1/5
```

	stkcd	date	opnprc	hiprc	lowprc	clsprc	rit
1.	8	1992 05 07	19.9	20.65	19.9	20.65	106.5
2.	8	1992 05 08	20.6	20.6	17.75	17.75	-14.0436
3.	8	1992 05 11	18	18	16.8	17.95	1.1268
4.	8	1992 05 12	18.1	19	17	18.55	3.3426
5.	8	1992 05 13	18.55	18.55	17.8	18.2	-1.8868

3.2 Download data for multiple companies

To download multiple firms' data, we add stock codes after `cntrade`. Suppose we want to download data for companies with the stock codes 000002, 000008, 000099, 600010, and 601898 and save the output to folder `d:\temp`. We simply invoke the following command:

```
. cntrade 8 2 99 600010 601898, path(d:\temp)
(15 vars, 5311 obs)
(165 observations deleted)
file d:\temp/000008.dta saved
(15 vars, 5640 obs)
(109 observations deleted)
file d:\temp/000002.dta saved
(15 vars, 3283 obs)
(49 observations deleted)
file d:\temp/000099.dta saved
(15 vars, 3140 obs)
(110 observations deleted)
file d:\temp/600010.dta saved
(15 vars, 1470 obs)
(11 observations deleted)
file d:\temp/601898.dta saved
. list stkcd date opnprc hiprc lowprc clsprc rit in 1/5
```

	stkcd	date	opnprc	hiprc	lowprc	clsprc	rit
1.	601898	2008 02 01	24	24.89	21.35	22.2	31.9073
2.	601898	2008 02 04	22.9	24.42	22.6	24.42	10
3.	601898	2008 02 05	23.8	24.25	22.84	22.98	-5.8968
4.	601898	2008 02 13	22.28	22.44	21.55	21.65	-5.7876
5.	601898	2008 02 14	22.1	22.76	21.81	22.24	2.7252

Here the output includes five `.dta` files, which are saved to folder `d:\temp\`.

To download the data for many companies, we can save the stock codes into a local, such as `` `stocklist´ ``.

```
. set more off
. local stocklist 2 8 9 99 600000 600010 601898 601988 601666 100 2024
. cntrade `stocklist´, path(d:\temp)
(15 vars, 5640 obs)
(109 observations deleted)
file d:\temp/000002.dta saved
(15 vars, 5311 obs)
(165 observations deleted)
file d:\temp/000008.dta saved
(15 vars, 5533 obs)
(132 observations deleted)
file d:\temp/000009.dta saved
(15 vars, 3283 obs)
(49 observations deleted)
file d:\temp/000099.dta saved
(15 vars, 3453 obs)
(85 observations deleted)
```

```
(note: file d:\temp/600000.dta not found)
file d:\temp/600000.dta saved
(15 vars, 3140 obs)
(110 observations deleted)
file d:\temp/600010.dta saved
(15 vars, 1470 obs)
(11 observations deleted)
file d:\temp/601898.dta saved
(15 vars, 1857 obs)
(16 observations deleted)
(note: file d:\temp/601988.dta not found)
file d:\temp/601988.dta saved
(15 vars, 1761 obs)
(12 observations deleted)
(note: file d:\temp/601666.dta not found)
file d:\temp/601666.dta saved
(15 vars, 2449 obs)
(166 observations deleted)
(note: file d:\temp/000100.dta not found)
file d:\temp/000100.dta saved
(15 vars, 2332 obs)
(57 observations deleted)
(note: file d:\temp/002024.dta not found)
file d:\temp/002024.dta saved
```

3.3 Data management after downloading

Combine

Suppose we define a local `` `stocklist´ ``, which contains all the listed firms' stock codes, and then run `cntrade `stocklist´, path(d:/temp)`. After several hours, we will get more than 2,500 listed firms' trading data files saved in the folder specified by the `path()` option, which is `"d:/temp"`. Financial research, such as event study, normally requires us to combine the files vertically. We can use the following code to combine them:

```
. clear
. local stocklist 2 8 9 99 600000 600010 601898 601988 601666 100 2024
. cd d:/temp
d:\temp
. cntrade `stocklist´
d:\temp
(15 vars, 5642 obs)
(109 observations deleted)
file d:\temp/000002.dta saved
(15 vars, 5314 obs)
(165 observations deleted)
file d:\temp/000008.dta saved
(15 vars, 5535 obs)
(132 observations deleted)
file d:\temp/000009.dta saved
(15 vars, 3285 obs)
(49 observations deleted)
file d:\temp/000099.dta saved
```

```
(15 vars, 3455 obs)
(85 observations deleted)
file d:\temp/600000.dta saved
(15 vars, 3142 obs)
(110 observations deleted)
file d:\temp/600010.dta saved
(15 vars, 1472 obs)
(11 observations deleted)
file d:\temp/601898.dta saved
(15 vars, 1859 obs)
(16 observations deleted)
file d:\temp/601988.dta saved
(15 vars, 1763 obs)
(12 observations deleted)
file d:\temp/601666.dta saved
(15 vars, 2451 obs)
(166 observations deleted)
file d:\temp/000100.dta saved
(15 vars, 2334 obs)
(57 observations deleted)
file d:\temp/002024.dta saved
. local datafiles: dir . files "*.dta"
. foreach file in `datafiles´ {
  2.    append using d:/temp/`file´
  3. }
(note: variable stkcd was int, now long to accommodate using data´s values)
. sort stkcd date
. save d:/trading_data, replace
(note: file d:/trading_data.dta not found)
file d:/trading_data.dta saved
. list stkcd date hiprc lowprc rit in 6118/6121
```

	stkcd	date	hiprc	lowprc	rit
6118.	8	1994 09 23	9.29	8.95	-3.9362
6119.	8	1994 09 26	9.15	8.03	-.443
6120.	8	1994 09 27	9.2	8.8	-.3337
6121.	8	1994 09 28	9.2	8.86	.7813

Transfer daily data into weekly, monthly, or annual data

After we combine the files, we can transfer the daily trading data into weekly, monthly, or even annual data. For annual data, the code is as follows:

```
. use d:\trading_data, clear
. gen year = year(date)
. sort stkcd year date
. by stkcd year: egen ann_rit = sum(ln(1+rit/100))
. replace ann_rit = (exp(ann_rit)-1)*100
(35164 real changes made)
. by stkcd year: egen ann_turnover = sum(turnover)
. by stkcd year: egen ann_volume = sum(volume)
```

```
. by stkcd year: egen ann_transaction = sum(transaction)
. by stkcd year: egen ann_hiprc = max(hiprc)
. by stkcd year: egen ann_lowprc = min(lowprc)
. by stkcd year: gen ann_opnprc = opnprc[1]
. by stkcd year: gen ann_clsprc = clsprc[_N]
. by stkcd year: keep if _n==_N
(35001 observations deleted)
. keep stkcd year ann*
. sort stkcd year
. save d:\ann_trading_data, replace
(note: file d:\ann_trading_data.dta not found)
file d:\ann_trading_data.dta saved
```

4 Comparison with fetchyahooquotes

The idea for this code comes from the command `fetchyahooquotes`, designed by Dicle and Levendis (2011). Why do we need `cntrade` when an effective command already exists? There are several reasons. First, `fetchyahooquotes` cannot download trading quotes for listed Chinese firms, because Yahoo! Finance does not cover Chinese-listed companies. On the other hand, NetEase provides comprehensive financial information for Chinese-listed firms but requires users to manually download the trading quotes. Our code automates downloading, which makes it a nice extension to `fetchyahooquotes`. Second, the data format provided by Yahoo! Finance is different from the data format provided by NetEase. Third, there is a huge demand for inexpensive and accessible financial data services in China, especially among university students. `cntrade` is a useful command that will foster these students' enthusiasm for Stata.

5 Reference

Dicle, M. F., and J. Levendis. 2011. Importing financial data. *Stata Journal* 11: 620–626.

About the authors

Xuan Zhang is a lecturer in statistics at the Zhongnan University of Economics and Law in Wuhan, China.

Chuntao Li is an associate professor of finance at the Zhongnan University of Economics and Law in Wuhan, China.