Using ODBC with Stata

Rita Sousa (rcsousa@bportugal.pt)

Bank of Portugal

Center for Mathematics and Applications
Summary

- Introduction
- ODBC Support
- Configuring ODBC
- ODBC in Stata
- Loading data
- Practical Example
- Useful Tips
- Final Remarks
Introduction

- Open DataBase Connectivity (ODBC) is a standardized set of function calls that can be used to access data stored in database management systems.
- Stata's `odbc` command allows us to load, write, and view data from ODBC sources.
- In my presentation I want to show you how to deal with large databases using the ODBC resources.
ODBC Support

- Import data from any ODBC data source, such as Oracle, SQL Server, Access, Excel, MySQL, and DB2.
- Export data to new or existing ODBC tables
- Execute custom SQL commands
Configuring ODBC for Windows

Before you start using the `odbc` command in Stata, you must first set up a data source name (DSN) in the ODBC Data Source Administrator.

1. Select **Control Panel**
2. Select **System and Security** in the Control Panel
3. Next select **Administrative Tools**
4. Double-click on **Data Sources (ODBC)** to open the ODBC Data Source Administrator.
Configuring ODBC for Windows

List of User Data Sources Names (DSN)

Create a New DSN
Configuring ODBC for Windows

- Select the appropriate driver from the list.
- For example, choose **Microsoft Access Driver (*.mdb)**.
Configuring ODBC for Windows

- Select your MS Access file to be in the Data Source test that we are creating
- After click OK
Configuring ODBC for Windows

- 64-bit Windows ships with two different ODBC Data Source Administrators, 64-bit and 32-bit
- Make sure you are using the correct version according to your Stata version
- Stata provides odbc as the Client interface

Client interface (Stata)
- `odbc list`
- `odbc query`
- `odbc describe`
- `odbc load`
- `odbc insert`
- `odbc exec`
- `odbc sqlfile`
ODBC in Stata

- **`odbc list`** produces a list of ODBC data source names to which Stata can connect

  ```
  . odbc list
  +-----------------+------------------+
  | Data Source Name | Driver           |
  +-----------------+------------------+
  | IES             | SQL Server       |
  | dBASE Files     | Microsoft Access dBASE Driver (*.dbf, *.ndx) |
  | Excel Files     | Microsoft Excel Driver (*.xls, *.xlsx, *.xl |
  | MS Access Database | Microsoft Access Driver (*.mdb, *.accdbo) |
  | SPAT            | SQL Server       |
  | test            | Driver do Microsoft Access (*.mdb)          |
  +-----------------+------------------+
  ```

- **`odbc query`** show a list of table names available from a specified data source

- **`odbc describe`** lists column names and types for each table available
Loading Data

- In Stata we can use some commands to import data stored in formats different from .dta such as `insheet`, `infix`, `import excel`, etc.

- `odbc load` reads an ODBC table into memory.

- You can load an ODBC table specified in the `table()` option or load an ODBC table generated by an SQL Command specified in the `exec()` option.
Loading Data

* odbc load id=nuemp nemp=pemp turnover=vn in 1/5,
  table("Table_Test") dsn("test")

* odbc load, exec(`"Select top 5 nuemp, pemp as pemp, vn as turnover
  From Table_Test"') dsn("test")
Practical Example

- Suppose now you want to access data stored in a SQL Server database.
- You have to register your ODBC database with the ODBC Data Source Administrator:
  1. Create a new Data Source
  2. Choose Sql Server Driver
  3. Specify the name of the server you want to connect
Practical Example

Using ODBC with Stata

@Rita Sousa

III Portuguese Stata Users Group Meeting
Practical Example

Creating the connection to the Data Source “SPAI”

Defining the Query statement

Loading and Saving Data
Practical Example

```
. describe
Contains data from C:\Users\eetu264\Documents\DST.dta
    obs:     38  
    vars:    3   
    size:   1,064 

storage  display  value
variable name  type format label

PAIDST_IDBP   long  %12.0g
PAIDST_CodDstr str4  %9s
PAIDST_DsgDis~o str20 %20s
```
Practical Example

Loading data from a master table with information of enterprises

```stata
clear
set more off
odbc list
odbc query "SFAI"

local year=2015
global path "C:\Users\ese264\Documents"
local query1 "select PAIDST_IDBP, PAIDST_CodDstr, PAIDST_DegDistrito from Distrito_BPLIM"
local query2 "SELECT * from EntEstrut_BPLIM WHERE left(PATENT_DataRefIn1,4)='year'"

*Loading the descriptions of district variable
odbc load, exec("'query1'") clear dsn("SFAI")
save "$\{path\}/DST", replace
describe

*Importing the data from the table EntEstrut_BPLIM for the year 2018
odbc load, exec("'query2'") clear dsn("SFAI")
merge m:1 PAIDST_IDBP using "$\{path\}/DST"
keep if _merge==1 | _merge==3
drop _merge
```
Practical Example

```
. save "${path}/Data`year'", replace
file C:\Users\eeu264\Documents\Data2015.dta saved
```
Writing Data

- `odbc insert` writes data from memory to an ODBC table.
- The data can change an existing table or create a new ODBC table.

```stata
odbc insert, table("Table_Name") dsn("DSN") insert_options
```

- `create` - create a simple ODBC table
- `overwrite` - clear data and write the data in memory to the ODBC table
- `insert` - default mode to append data in memory to the ODBC table
Some Useful Tips

- Stata keeps its entire dataset in memory. Usually it is faster but it can be a disadvantage when your dataset is very large.
  - `drop` any variables you don't need for your analysis
  - use the same variable names when you are combining datasets
  - Use `compress` to optimize variables format
  - `encode` strings
  - Use the cycles `foreach` or `forvalues` to do repeated commands
  - Use `preserve` and `restore` to make temporary changes to datasets
Final Remarks

- Stata command `odbc` allows Stata to load, write, or view data from ODBC sources.
- This command offers a useful set of function calls that can be used to access data stored in many different types of database management systems.
- Oracle, SQL Server, Access, Excel, MySQL and DB2 are some examples of available ODBC data sources.
- Stata’s `odbc` is a resourceful solution that allows us to query external databases and insert or update records in those databases.
Thank you for your attention

Using ODBC with Stata

*Rita Sousa* (rcsousa@bportugal.pt)

Bank of Portugal

Center for Mathematics and Applications