Time Series Database Interface: R ODBC (TSodbc)

June 7, 2012

1 Introduction

The code from the vignette that generates this guide can be loaded into an editor with *edit(vignette("TSodbc"))*. This uses the default editor, which can be changed using *options()*. It should be possible to view the pdf version of the guide for this package with *print(vignette("TSodbc"))*.

WARNING: running these example will overwrite tables in the ODBC "test" database on the server.

Once R is started, the functions in this package are made available with

```
> library("TSodbc")
```

This will also load required packages *TSdbi*, *DBI*, *RODBC*, *methods*, and *tframe*. Some examples below also require *zoo*, and *tseries*.

The ODBC user, password, hostname, etc, should be set in ODBC client configuration file (~/.odbc.ini on Linux/Unix systems) before starting R. An example of this file is provided in the final section of this vignette. Alternatively, this information can be set with environment variables ODBC_USER and ODBC_PASSWD. The variable ODBC_HOST does not seem to work for passing the ODBC connection, so a properly setup ODBC configuration file is also needed, but the environment variables will override the user and passwd setting in that file. (An environment variable ODBC_DATABASE can also be set, but "test" is specified below.) Below, the environment variable ODBC_USER is used to determine which of these methods is being used. If this environment variable is empty then it is assumed the configuration file will be used.

```
>    user    <- Sys.getenv("ODBC_USER")
>    if ("" != user) {
        passwd  <- Sys.getenv("ODBC_PASSWD")
        if ("" == passwd)   passwd <- NULL
       }
```

The next small section of code is necessary to setup database tables that are used in the examples below. It needs to be done only once for a database and might typically be done by an administrator setting up the database, rather than by an end user.

```
> m <- dbDriver("ODBC")
> con <-  if ("" == user) odbcConnect(dsn="test") else
      odbcConnect(dsn="test", uid=user, pwd=passwd)
> if(con == -1) stop("error establishing ODBC connection.")
> source(system.file("TSsql/CreateTables.TSsql", package = "TSdbi"))
> odbcClose(channel=con)
```

More detailed description of the instructions for building the database tables is given in the vignette for the *TSdbi* package. Those instruction show how to build the database using database utilites rather than R, which might be the way a system administrator would build the database.

2 Using the Database - TSdbi Functions

This section gives several simple examples of putting series on and reading them from the database. (If a large number of series are to be loaded into a database, one would typically do this with a batch process using the database program's utilities for loading data.) The first thing to do is to establish a connection to the database:

```
>  m <- dbDriver("ODBC")
>  con <-  if ("" == user) TSconnect(m, dbname="test") else
      TSconnect(m, dbname="test", uid=user, pwd=passwd)
```

TSconnect uses *odbcConnect* from the *RODBC* package, but checks that the database has expected tables, and checks for additional features. (It cannot be used before the tables are created, as done in the previous section.)

This puts a series called *vec* on the database and then reads is back

```
>  z <- ts(rnorm(10), start=c(1990,1), frequency=1)
>  seriesNames(z) <- "vec"
>  if(TSexists("vec", con)) TSdelete("vec", con)
>  TSput( z, con)
>  z <- TSget("vec", con)
```

If the series is printed it is seen to be a "ts" time series with some extra attributes.

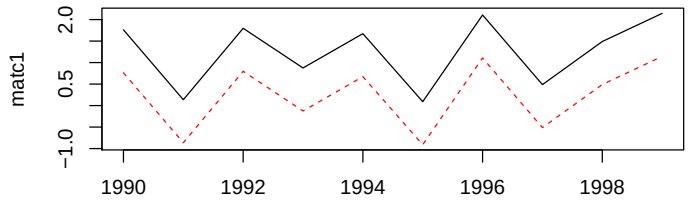
TSput fails if the series already exists on the *con*, so the above example checks and deletes the series if it already exists. *TSreplace* does not fail if the series does not yet exist, so examples below use it instead. Several plots below show original data and the data retrieved after it is written to the database. One is added to the original data so that both lines are visible.

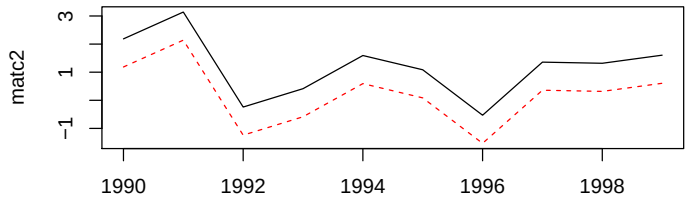
And now more examples:

```
> z <- ts(matrix(rnorm(20),10,2), start=c(1990,1), frequency=1)
> seriesNames(z) <- c("matc1", "matc2")
> TSreplace(z, con)

[1] TRUE

> TSget("matc1", con)

Time Series:
Start = 1990
End = 1999
Frequency = 1
         1          2          3          4          5          6          7
 0.7617099 -0.8635676  0.7971689 -0.1250809  0.6723554 -0.9092022  1.1066850
         8          9         10
-0.5126828  0.4897491  1.1436553
attr(,"seriesNames")
[1] matc1
attr(,"TSmeta")
serIDs:  matc1
 from dbname  test using  TSodbcConnection

> TSget("matc2", con)

Time Series:
Start = 1990
End = 1999
Frequency = 1
          1           2           3           4           5           6
 1.18574704  2.13989356 -1.24048715 -0.58658087  0.59324020  0.08595303
          7           8           9          10
-1.53079830  0.35797303  0.31710223  0.60516688
attr(,"seriesNames")
[1] matc2
attr(,"TSmeta")
serIDs:  matc2
 from dbname  test using  TSodbcConnection

> TSget(c("matc1","matc2"), con)

Time Series:
Start = 1990
End = 1999
Frequency = 1
          matc1       matc2
1990  0.7617099  1.18574704
1991 -0.8635676  2.13989356
```

```
1992  0.7971689 -1.24048715
1993 -0.1250809 -0.58658087
1994  0.6723554  0.59324020
1995 -0.9092022  0.08595303
1996  1.1066850 -1.53079830
1997 -0.5126828  0.35797303
1998  0.4897491  0.31710223
1999  1.1436553  0.60516688
attr(,"TSmeta")
serIDs:  matc1 matc2
 from dbname  test using  TSodbcConnection
```

```
> require("tfplot")
> tfplot(z+1, TSget(c("matc1","matc2"), con),
           lty=c("solid", "dashed"), col=c("black", "red"))
```





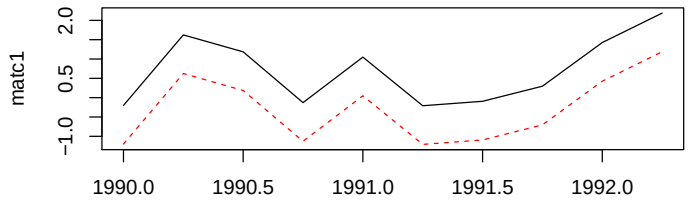

```
> z <- ts(matrix(rnorm(20),10,2), start=c(1990,1), frequency=4)
>  seriesNames(z) <- c("matc1", "matc2")
> TSreplace(z, con)
```

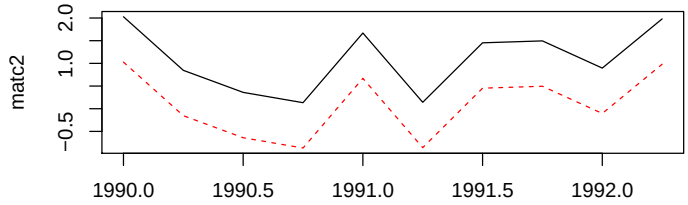
```
[1] TRUE
```

```
> TSget(c("matc1","matc2"), con)
```

```
              matc1      matc2
1990 Q1 -1.20124841  1.0275715
1990 Q2  0.62463717 -0.1533841
1990 Q3  0.18571034 -0.6395359
1990 Q4 -1.12850755 -0.8665698
1991 Q1  0.04945359  0.6692699
1991 Q2 -1.20883527 -0.8579593
1991 Q3 -1.09390847  0.4533236
1991 Q4 -0.70278790  0.4955156
1992 Q1  0.42619050 -0.1040841
1992 Q2  1.18780054  0.9806373
attr(,"TSmeta")
serIDs:  matc1 matc2
 from dbname  test using  TSodbcConnection
```

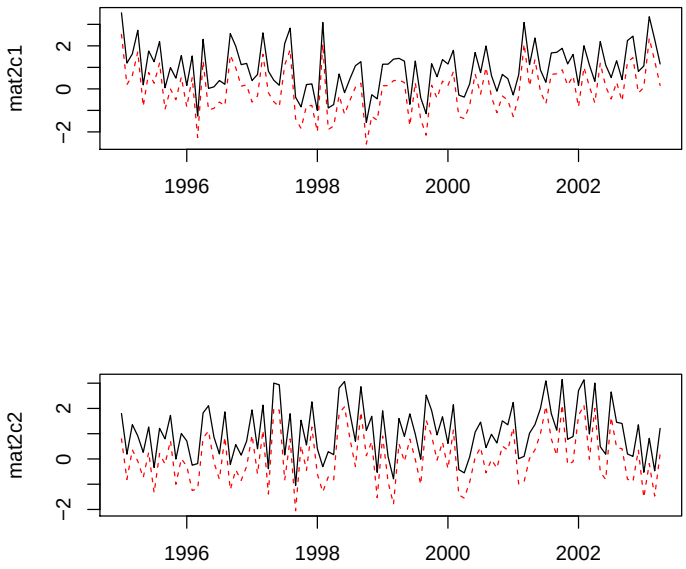
```
> tfplot(z+1, TSget(c("matc1","matc2"), con),
            lty=c("solid", "dashed"), col=c("black", "red"))
```






```
> z <- ts(matrix(rnorm(200),100,2), start=c(1995,1), frequency=12)
>  seriesNames(z) <- c("mat2c1", "mat2c2")
> TSreplace(z, con)

[1] TRUE
```

```
> tfplot(z+1, TSget(c("mat2c1","mat2c2"), con),
           lty=c("solid", "dashed"), col=c("black", "red"))
```



The following extract information about the series from the database, although not much information has been added for these examples.

```
> TSmeta("mat2c1", con)
> TSmeta("vec", con)
> TSdates("vec", con)
> TSdescription("vec", con)
> TSdoc("vec", con)
```

Below are exampoles that make more use of *TSdescription* and codeTSdoc. Often it is convenient to set the default connection:

```
> options(TSconnection=con)
```

and then the *con* specification can be omitted from the function calls unless another connection is needed. The *con* can still be specified, and some examples below do specify it, just to illustrate the alternative syntax.

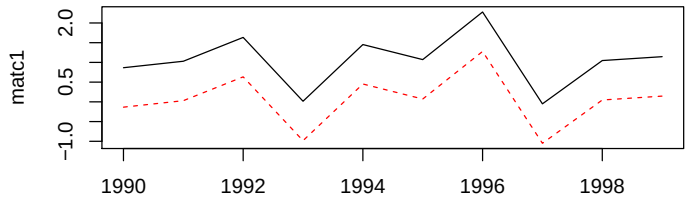
```
> z <- TSget("mat2c1")
> TSmeta("mat2c1")
```

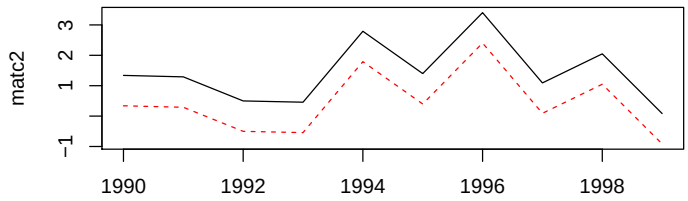
```
serIDs:  mat2c1
 from dbname  test using  TSodbcConnection
```

Data documentation can be in two forms, a description specified by *TSdescription* or longer documentation specified by *TSdoc*. These can be added to the time series object, in which case they will be written to the database when *TSput* or *TSreplace* is used to put the series on the database. Alternatively, they can be specified as arguments to *TSput* or *TSreplace*. The description or documentation will be retrieved as part of the series object with *TSget* only if this is specified with the logical arguments *TSdescription* and *TSdoc*. They can also be retrieved directly from the database with the functions *TSdescription* and *TSdoc*.

```
>  z <- ts(matrix(rnorm(10),10,1), start=c(1990,1), frequency=1)
>  TSreplace(z, serIDs="Series1", con)

[1] TRUE

>  zz <- TSget("Series1", con)
>  TSreplace(z, serIDs="Series1", con,
       TSdescription="short rnorm series",
       TSdoc="Series created as an example in the vignette.")

[1] TRUE

>  zz <- TSget("Series1", con, TSdescription=TRUE, TSdoc=TRUE)
>  start(zz)

[1] 1990    1

>  end(zz)

[1] 1999    1

>  TSdescription(zz)

[1] "short rnorm series"

>  TSdoc(zz)

[1] "Series created as an example in the vignette."

>  TSdescription("Series1", con)

[1] "short rnorm series"

>  TSdoc("Series1", con)

[1] "Series created as an example in the vignette."
```

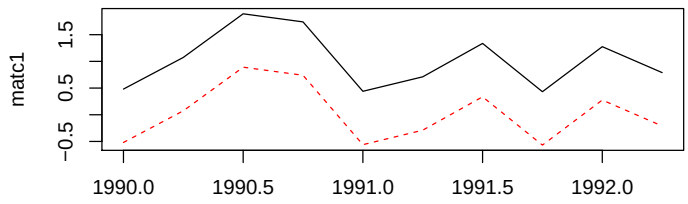
```
>  z <- ts(rnorm(10), start=c(1990,1), frequency=1)
>  seriesNames(z) <- "vec"
>  TSreplace(z, con)

[1] TRUE

>  zz <- TSget("vec", con)
>  z <- ts(matrix(rnorm(20),10,2), start=c(1990,1), frequency=1)
>  seriesNames(z) <- c("matc1", "matc2")
>  TSreplace(z, con)

[1] TRUE

>  tfplot(z+1, TSget(c("matc1","matc2"), con),
            lty=c("solid", "dashed"), col=c("black", "red"))
>
```

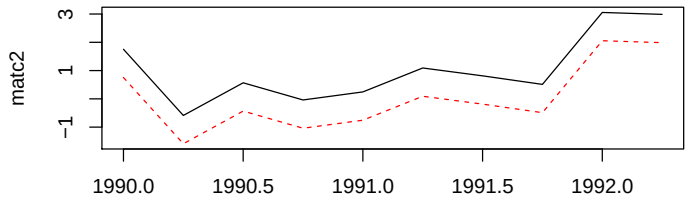




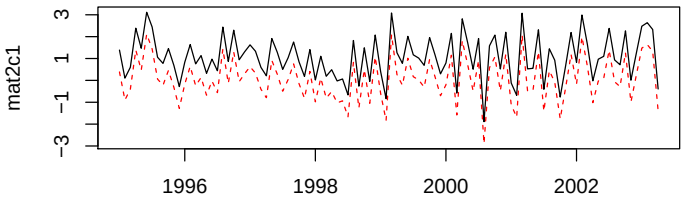

```
>  z <- ts(matrix(rnorm(20),10,2), start=c(1990,1), frequency=4)
>  seriesNames(z) <- c("matc1", "matc2")
>  TSreplace(z, con)

[1] TRUE
```

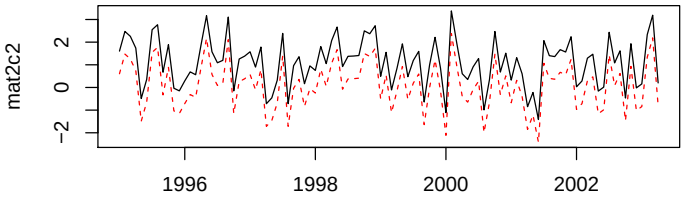
```
> tfplot(z+1, TSget(c("matc1","matc2"), con),
           lty=c("solid", "dashed"), col=c("black", "red"))
>
```





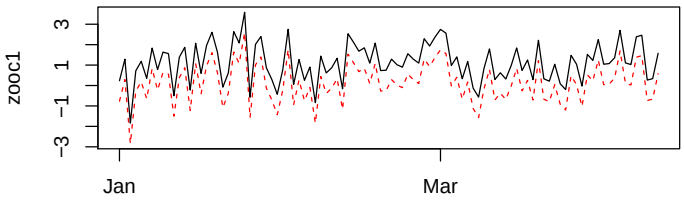

```
> z <- ts(matrix(rnorm(200),100,2), start=c(1995,1), frequency=12)
> seriesNames(z) <- c("mat2c1", "mat2c2")
> TSreplace(z, con)

[1] TRUE

> tfplot(z+1, TSget(c("mat2c1","mat2c2"), con),
           lty=c("solid", "dashed"), col=c("black", "red"))
```

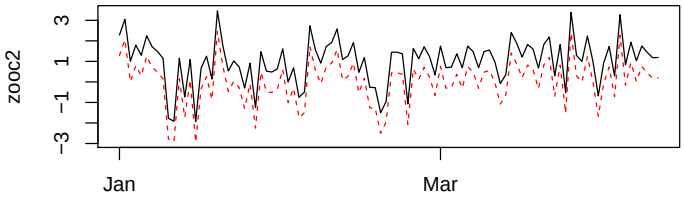




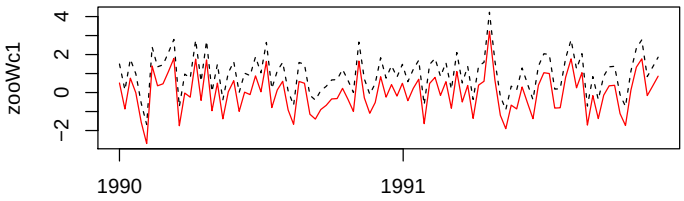
The following examples use dates and times which are not handled by *ts*, so the *zoo* time representation is used.

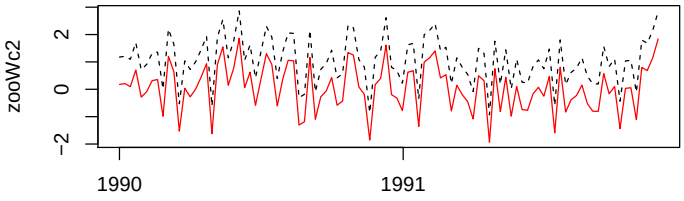
```
> require("zoo")
>  z <- zoo(matrix(rnorm(200),100,2), as.Date("1990-01-01") + 0:99)
>   seriesNames(z) <- c("zooc1", "zooc2")
>  TSreplace(z, con, Table="D")

[1] TRUE

>  tfplot(z+1, TSget(c("zooc1","zooc2"), con),
           lty=c("solid", "dashed"), col=c("black", "red"))
>
```






```
>  z <- zoo(matrix(rnorm(200),100,2), as.Date("1990-01-01") + 0:99 * 7)
>   seriesNames(z) <- c("zooWc1", "zooWc2")
>  TSreplace(z, con, Table="W")

[1] TRUE

>  tfplot(z+1, TSget(c("zooWc1","zooWc2"), con), col=c("black", "red"),
      lty=c("dashed", "solid"))
```






```
> dbDisconnect(con)
> #dbUnloadDriver(m)
```

3 Examples Using Web Data

This section illustrates fetching data from a web server and loading it into the database. This would be a very slow way to load a database, but provides examples of different kinds of time series data. The fetching is done with *TShistQuote* which provides a wrapper for *get.hist.quote* from package *tseries* to give syntax consistent with the *TSdbi*.

Fetching data may fail due to lack of an Interenet connection or delays.

First establish a connection to the database where data will be saved:

```
> con <- if ("" == user) TSconnect("ODBC", dbname="test") else
    TSconnect("ODBC", dbname="test", uid=user, pwd=passwd)
```

Now connect to the web server and fetch data:

```
> require("TShistQuote")
> Yahoo <- TSconnect("histQuote", dbname="yahoo")
> x <- TSget("^gspc", quote = "Close", con=Yahoo)
> plot(x)
```

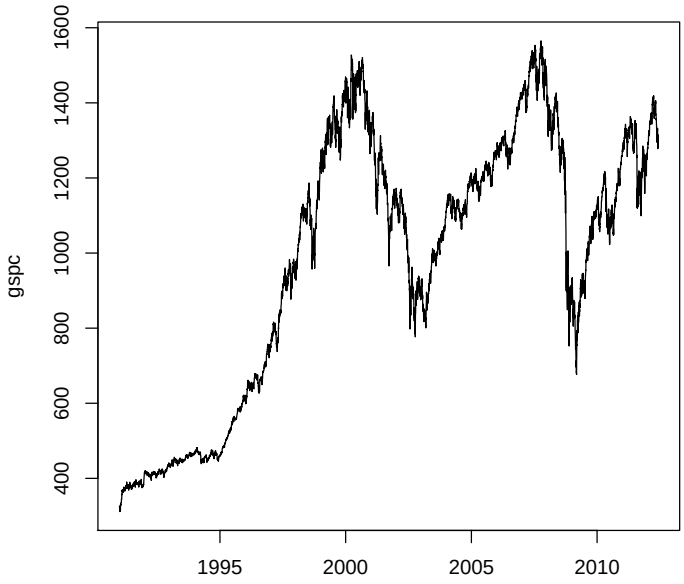
```
>   tfplot(x)
>   TSrefperiod(x)

[1] "Close"

>   TSdescription(x)

[1] "^gspc Close  from  yahoo"

>   TSdoc(x)

[1] "^gspc Close  from  yahoo retrieved  2012-06-07 17:05:10"

>   TSlabel(x)

[1] "^gspc Close"
```

Then write the data to the local server, specifying table B for business day data (using TSreplace in case the series is already there from running this example previously):

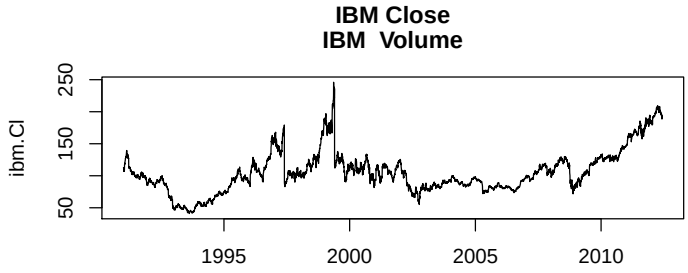
```
>    TSreplace(x,  serIDs="gspc", Table="B", con=con)

[1] TRUE
```

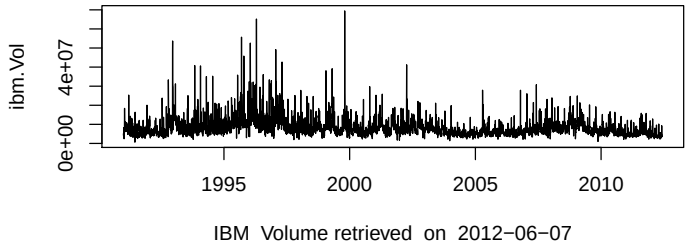
and check the saved version:

```
>   TSrefperiod(TSget(serIDs="gspc", con=con))

[1] 1

>   TSdescription("gspc", con=con)

[1] "^gspc Close  from  yahoo"

>   TSdoc("gspc", con=con)

[1] "^gspc Close  from  yahoo retrieved  2012-06-07 17:05:10"

>   TSlabel("gspc", con=con) # this is not yet supported on the db

[1] NA

>   tfplot(TSget(serIDs="gspc", con=con))
```




```
>    x <- TSget("ibm", quote = c("Close", "Vol"), con=Yahoo)
>    TSreplace(x, serIDs=c("ibm.Cl", "ibm.Vol"),  con=con, Table="B",
               TSdescription.=c("IBM Close","IBM  Volume"),
               TSdoc.= paste(c(
                  "IBM  Close   retrieved on ",
                  "IBM  Volume retrieved  on "), Sys.Date()))

[1] TRUE

>    z <- TSget(serIDs=c("ibm.Cl", "ibm.Vol"),
              TSdescription=TRUE, TSdoc=TRUE, con=con)
>    TSdescription(z)

[1] "IBM Close"   "IBM  Volume"

>    TSdoc(z)

[1] "IBM  Close   retrieved on  2012-06-07"
[2] "IBM  Volume retrieved  on  2012-06-07"

>    tfplot(z, xlab = TSdoc(z), Title = TSdescription(z))
>    tfplot(z, Title="IBM", start="2007-01-01")
```





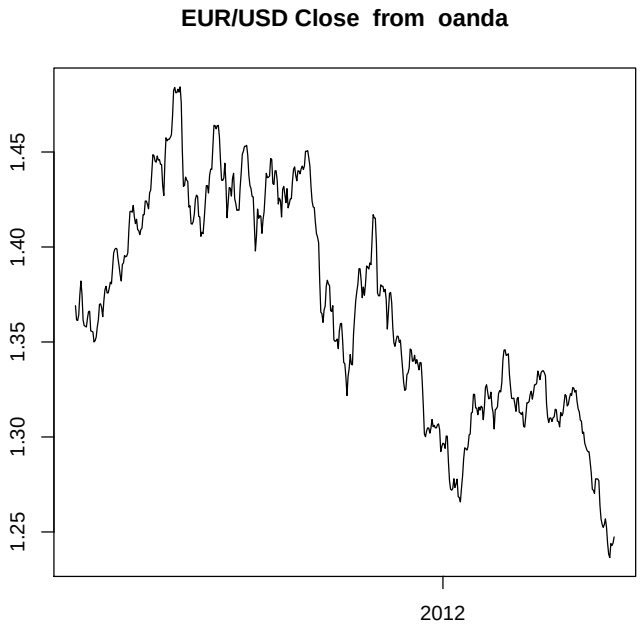
Oanda has maximum of 500 days, so the start date is specified here so as to not exceed that.

```
>    Oanda <- TSconnect("histQuote", dbname="oanda")
>    x <- TSget("EUR/USD", start=Sys.Date() - 495, con=Oanda)
>    TSreplace(x, serIDs="EUR/USD", Table="D", con=con)
```

```
[1] TRUE
```

Then check the saved version:

```
>    z <- TSget(serIDs="EUR/USD",TSlabel=TRUE, TSdescription=TRUE,  con=con)
>    tfplot(z, Title = TSdescription(z), ylab=TSlabel(z))
>    tfplot(z, Title = "EUR/USD", start=Sys.Date() - 495)
>    tfplot(z, Title = "EUR/USD", start=Sys.Date() - 100)
>    tfplot(z, Title = "EUR/USD", start=Sys.Date()  -14, end=Sys.Date(),
       xlab = format(Sys.Date(), "%Y"))
```




```
>   dbDisconnect(con)
>   dbDisconnect(Yahoo)
>   dbDisconnect(Oanda)
```

3.1 Examples Using TSdbi with ets

A more complete set of examples will be available as a vignette in a package tentatively called *TSdata*. This should include examples from a more extensive database of economic time series.

4 Examples Using DBI and direct SQL Queries

The following examples are queries using direct SQL queries. They should not often be needed to access time series, but may be useful to get at more detailed information, or formulate special queries. Some databases support special calls to access database or table information, but the following try to use generic SQL.

```
>  m <- dbDriver("ODBC")
>  con <-  if ("" == user) TSconnect(m, dbname="test") else
      TSconnect(m, dbname="test", uid=user, pwd=passwd)
>  options(TSconnection=con)
```

```
> dbListTables(con)

 [1] "a"    "b"    "d"    "i"    "m"    "meta" "q"    "s"    "t"    "u"
[11] "w"
```

If schema queries are supported then table information can be obtained in a generic SQL way, but on some systems this will fail because users do not have read priveleges on the INFORMATION_SCHEMA table, so the following are wrapped in *tryCatch()*. (SQLite does not seem to support this at all.)

```
>   tryCatch( dbGetQuery(con, paste(
         "SELECT COLUMN_NAME FROM INFORMATION_SCHEMA.Columns ",
         " WHERE TABLE_SCHEMA='test' AND table_name='A' ;"))  )

NULL

>   tryCatch( dbGetQuery(con, paste(
         "SELECT COLUMN_NAME, COLUMN_DEFAULT, COLLATION_NAME, DATA_TYPE,",
         "CHARACTER_SET_NAME, CHARACTER_MAXIMUM_LENGTH, NUMERIC_PRECISION",
   "FROM INFORMATION_SCHEMA.Columns WHERE TABLE_SCHEMA='test' AND table_name='A' ;"))  )

NULL

>   tryCatch( dbGetQuery(con, paste(
     "SELECT COLUMN_NAME, DATA_TYPE, CHARACTER_MAXIMUM_LENGTH, NUMERIC_PRECISION",
   "FROM INFORMATION_SCHEMA.Columns WHERE TABLE_SCHEMA='test' AND table_name='M';"))  )

NULL

>
```

Finally, to disconnect gracefully, one should

```
>   dbDisconnect(con)
>   #dbDisconnect(options()$TSconnection)
>   options(TSconnection=NULL)
>   odbcCloseAll()
>   dbUnloadDriver(m)
```

5 Example ODBC configuration file

Following is an example ODBC configuration file I use in Linux (so the file is in my home directory and called ".odbc.ini") to connect to a remote PostgreSQL server:

```
[test]

Description             = test DB (Postgresql)
```

```
Driver                  = Postgresql
Trace                   = No
TraceFile               = /tmp/test_odbc.log
Database                = test
Servername              = some.host
UserName                = paul
Password                = mySecret
Port                    = 5432
Protocol                = 6.4
ReadOnly                = No
RowVersioning           = No
ShowSystemTables        = No
ShowOidColumn           = No
FakeOidIndex            = No
ConnSettings            =


[ets]

Description             = ets DB (Postgresql)
Driver                  = Postgresql
Trace                   = No
TraceFile               = /tmp/test_odbc.log
Database                = ets
Servername              = some.host
UserName                = paul
Password                = mySecret
Port                    = 5432
Protocol                = 6.4
ReadOnly                = No
RowVersioning           = No
ShowSystemTables        = No
ShowOidColumn           = No
FakeOidIndex            = No
ConnSettings            =
```