

# Stata를 이용한 국제무역데이터 구축 : UN Comtrade DB 활용

발표자: 민 인 식 교수  
경희대학교 경제학과

2015년 6월 23일

## I. UN Comtrade Database

○ UN에서 공식적으로 제공하는 국제무역 데이터베이스  
[comtrade.un.org](http://comtrade.un.org)

○ Country profiles

약 250여개 국가의 무역 데이터를 제공하고 있다.

bilateral trade 데이터 : reporter and partner

reporter 국가와 전 세계(World) 무역거래

○ Commodity profiles

HS 분류 : 2 digit ~ 6 digit

SITC 분류 : 2 digit ~ 5 digit

○ Frequency profiles

연도별 데이터는 1962년부터 ~

월별 데이터는 2010년부터 ~

○ Data extraction

1) Basic Access

[comtrade.un.org](http://comtrade.un.org) 웹사이트에서 제한적으로 엑셀, CSV 파일형태로 다운로드 받을 수 있다.

2) Premium Webservice

매년 연회비를 지불하면 더 많은 데이터를 extract할 수 있다.

Option*	User Access	Records Limit	Limit per query	Annual Fee
My Comtrade	Single User	unlimited	50,000	US \$121.30
Premium Individual License	Single User	unlimited	20,000,000	US \$909.75
Premium Site License	Multiple Users	unlimited	No limit**	US \$6,065.00

### 3) Open API

연구자가 사용하는 외부 software(가령 통계패키지) 프로그램에서 Comtrade DB에 접속하여 data를 가져올 수 있도록 API(Application Programming Interface)를 공개하고 있다.

현재 API 버전에서는 접속 limit이 존재하기 때문에 Comtrade의 모든 데이터를 가져오는 일은 쉽지 않다. 이러한 접속 limit에도 불구하고 저자가 만든 코드를 이용하면 Comtrade의 모든 데이터를 가져오는 것이 가능하다.

## II. Stata and Comtrade API

○ 현재 Comtrade API 버전에서는 다음과 같은 parameter(일종의 데이터 extraction의 “조건”)를 입력하면 DB의 데이터를 불러올 수 있다.

### API의 parameters

- 1) reporting country: 수출/수입을 UN에 보고하는 국가
- 2) frequency: 연간, 월간
- 3) classification : HS, SITC
- 4) partner country: reporting country의 상대국가
- 5) trade flow: 수출/수입/재수출/재수입
- 6) commodity code: 2 digit/ 4 digit / 6 digit(5 digit)

○ Comtrade에서 공개한 API와 Stata의 기능을 활용하여 위 parameter 값을 자동으로 지정하여 Stata 데이터 파일로 만드는 코드를 작성할 수 있다

현재 저자의 코드에서는

- 1) reporting country: 250여개 국가를 지정할 수 있다. 3자리 수의 국가코드로 지정해야 한다.
- 2) frequency ; 연간 데이터만 가져올 수 있다.
- 3) classification : HS 또는 SITC
- 4) partner country: 250여개 국가를 지정할 수 있다.
- 5) trade flow: 수출/수입 무역 데이터만 가져올 수 있다.
- 6) commodity code : 2 digit/4 digit/6 digit(5 digit) 까지 모두 가져올 수 있다.

#### ○ Comtrade API의 제한

현재 Open API는 무료로 제공되는 반면, Premium Webservice를 유료로 판매하고 있다는 점을 고려하여 API에 제한을 두고 있다.

- 1) 한 번의 request에 대해 최대 50000개의 관측치(record)만 가져올 수 있다.
- 2) 1시간 동안 100번의 requests만 가능하다.
- 3) 따라서 하루 동안 최대 50000개 x 100 x 24시간 = 1억 2천만 개 records를 가져올 수 있다.

그러나 현실적으로 “1시간에 100번의 requests” 조건 때문에 API를 활용하여 연구자가 필요한 데이터로 만드는데 제약이 된다.

“One request”의 의미 : 한국(reporter) => 중국(partner)에서 2013년 발생한 모든 품목 (commodity: 가령 4-digit 품목) 수출 거래 데이터

### III. Stata 명령어와 예제 명령문

○ Comtrade API의 제약조건 특히 “1시간에 100 requests” 때문에 Stata 코드 작성에 어려움이 있다. 이 제약에 걸리지 않도록 코드를 design하는 것이 핵심이다.

○ comtrade\_use 명령어

최근 5년(2010~2014년) 무역데이터를 가져오는데 사용한다.

comtrade\_use , *start( ) end( ) reporter( ) partner( ) cc( ) regime( ) sitc [wd( )]*

*start( )*: 무역데이터를 가져오고자 하는 시작연도를 지정한다. 2010년~2014년도 중 하나를 지정할 수 있다.

*end( )*: 데이터의 마지막 연도를 지정한다. 2010년~2014년도 중 하나를 지정할 수 있다.

*reporter( )*: reporting country 코드를 지정한다. 1개 국가의 코드만 지정할 수 있다.

*partner( )*: partner country 코드를 지정한다. 최대 99개까지 지정할 수 있다.

*cc( )*: AG2, AG4, AG6 셋 중 하나를 지정한다.

AG2 : 2 digit commodity 무역데이터를 가져올 수 있다.

AG4 : 4 digit commodity 무역데이터를 가져올 수 있다.



AG6 : 6 digit commodity 무역데이터를 가져올 수 있다.

sitc 옵션을 사용하는 경우에는 AG6 대신 AG5를 입력해야 한다.

*regime( )*: 1 또는 2를 지정할 수 있다.

1 : 수입 거래 데이터를 가져온다.

2 : 수출 거래 데이터를 가져온다.

*sitc* : HS commodity 대신 SITC commodity (S1 revision) 분류를 가져오고자 할 때 사용한다. sitc 옵션이 없다면 자동으로 HS 분류를 저장한다.

*wd( )* : 필수 옵션은 아니다. 데이터가 저장되는 폴더를 지정한다.

○ comtrade\_use1 명령어

2010년 이전 무역 데이터를 가져오고자 할 때 사용한다.

comtrade\_use1 , *year( ) reporter( ) partner( ) cc( ) regime( ) sitc [wd( )]*

*year( )*: 2010년 이전 년도에 해당하는 1개 년도만 지정할 수 있다.

나머지 옵션은 comtrade\_use 명령어와 같다.

○ comtrade\_add 명령어

comtrade\_use와 comtrade\_use1 명령어에 의해 만들어진 임시데이터 파일을 모두 합쳐서 1개의 데이터 세트로 구축해준다.

comtrade\_add, *start( ) end( ) country( ) regime( ) [wd( ) save( )]*

*start( )* : 임시로 만든 데이터파일의 시작년도를 지정한다.

*end( )* : 임시로 만든 데이터파일의 마지막 년도를 지정한다.

*country( )* : reporter( ) 옵션에서 지정한 국가코드를 입력한다. 다수의 국가코드를 입력할 수 있다.

*regime( )* : “1” 또는 “2” 또는 “1 2” 셋 중 하나를 지정할 수 있다.

“1” (수출 데이터만 구축한다)

“2” (수입 데이터만 구축한다)

“1 2” (수출과 수입데이터를 모두 구축한다)

*wd( )* : 데이터가 저장되는 폴더를 지정한다.

*save( )* : 마지막에 저장되는 데이터 파일의 이름을 임의로 지정할 수 있습니다. save 옵션을 사용하지 않으면 자동으로 trade\_final\_7.dta 파일로 저장된다.

○ 예제 명령문

다음과 같은 데이터를 구축하고자 한다.

commodity code = HS

reporting 국가	partner 국가	year	regime	cc
Korea (국가코드 410)	United Kingdom(826) China (156) USA (842) Japan (392)	2009~2014	수출/수입	AG2
United Kingdom (국가코드 826)	Korea (410) China (156) USA (842) Japan (392)	2009~2014	수출/수입	AG2

```
set more off
set trace off
cd C:\₩DaumCloud₩민인식작업7₩trade
comtrade_use , start(2010) end(2014) reporter(410) partner(826 156 842 392) regime(1) cc(AG2) //410
comtrade_use , start(2010) end(2014) reporter(410) partner(826 156 842 392) regime(2) cc(AG2) //410

comtrade_use1 , year(2009) reporter(410) partner(826 156 842 392) regime(1) cc(AG2) //410
comtrade_use1 , year(2009) reporter(410) partner(826 156 842 392) regime(2) cc(AG2) //410

comtrade_use , start(2010) end(2014) reporter(826) partner(410 156 842 392) regime(1) cc(AG2) //826
comtrade_use , start(2010) end(2014) reporter(826) partner(410 156 842 392) regime(2) cc(AG2) //826

comtrade_use1 , year(2009) reporter(826) partner(410 156 842 392) regime(1) cc(AG2) //826
comtrade_use1 , year(2009) reporter(826) partner(410 156 842 392) regime(2) cc(AG2) //826

comtrade_add, start(2009) end(2014) country(410 826) regime(1 2) // 2014년 데이터가 없는 나라도 있다.
```

위 코드는 Comtrade API에 총 40번 requests를 실행하게 된다. 따라서 1시간에 100번 requests라는 제약조건을 위배하지 않는다.

○ 위 코드에 의해 만들어진 최종 데이터의 일부분

최종적으로 만들어진 데이터 파일 이름: trade\_final\_7.dta

Contains data from trade\_final\_7.dta

```
obs:      4,264
vars:      35                      11 Jun 2015 00:13
size:     1,735,448
```

variable name	storage type	display format	value label	variable label
id	str7	%9s		reporter국가+partner국가
year	int	%8.0g		Year
commoditycode	byte	%02.0f		Commodity Code
classification	str2	%9s		Classification
period	int	%8.0g		Period
perioddesc	int	%8.0g		Period Desc.
aggregatelevel	byte	%8.0g		Aggregate Level
isleafcode	byte	%8.0g		Is Leaf Code
reportercode	int	%8.0g		Reporter Code
reporter	str14	%14s		Reporter
reporteriso	str3	%9s		Reporter ISO
partnercode	int	%8.0g		Partner Code
partner	str14	%14s		Partner
partneriso	str3	%9s		Partner ISO
commodity	str275	%275s		Commodity

qtyunitcode	byte	%8.0g
qtyunit	str11	%11s
im_qty	byte	%8.0g
im_netweightkg	byte	%8.0g
im_tradevalueus	double	%20.0g
world_im_qty	byte	%8.0g
world_im_netw~g	byte	%8.0g
world_im_trad~s	double	%20.0g
total_im_qty	float	%20.0g
total_im_netw~g	float	%20.0g
total_im_trad~s	float	%20.0g
ex_qty	byte	%8.0g
ex_netweightkg	byte	%8.0g
ex_tradevalueus	double	%20.0g
world_ex_qty	byte	%8.0g
world_ex_netw~g	byte	%8.0g
world_ex_trad~s	double	%20.0g
total_ex_qty	float	%20.0g
total_ex_netw~g	float	%20.0g
total_ex_trad~s	float	%20.0g

Qty Unit Code
Qty Unit
해당 품목 수입 수량 (reporter 국가에서 partner 국가로부터 수입)
해당 품목 수입 중량 (reporter 국가에서 partner 국가로부터 수입)
해당 품목 수입 금액 (reporter 국가에서 partner 국가로부터 수입)
해당 품목 수입 수량 (reporter 국가의 전체 수입 수량)
해당 품목 수입 중량 (reporter 국가의 전체 수입 중량)
해당 품목 수입 금액 (reporter 국가의 전체 수입 금액)
전체 품목 수입 수량 (reporter 국가에서 partner 국가로부터 수입)
전체 품목 수입 중량 (reporter 국가에서 partner 국가로부터 수입)
전체 품목 수입 금액 (reporter 국가에서 partner 국가로부터 수입)
해당 품목 수출 수량 (reporter 국가에서 partner 국가로 수출)
해당 품목 수출 중량 (reporter 국가에서 partner 국가로 수출)
해당 품목 수출 금액 (reporter 국가에서 partner 국가로 수출)
해당 품목 수출 수량 (reporter 국가의 전체 수출 수량)
해당 품목 수출 중량 (reporter 국가의 전체 수출 중량)
해당 품목 수출 금액 (reporter 국가의 전체 수출 금액)
전체 품목 수출 수량 (reporter 국가에서 partner 국가로 수출)
전체 품목 수출 중량 (reporter 국가에서 partner 국가로 수출)
전체 품목 수출 금액 (reporter 국가에서 partner 국가로 수출)

id	year	commoditycode	reporteriso	im_tradevalueus	ex_tradevalueus
KOR_CHN	2009	01	KOR	5416898	30616
KOR_CHN	2009	02	KOR	85138	4409922
KOR_CHN	2009	03	KOR	743279716	120786302
KOR_CHN	2009	04	KOR	1777353	6190812
KOR_CHN	2009	05	KOR	22580511	6198419
KOR_CHN	2009	06	KOR	16988853	14436024
KOR_CHN	2009	07	KOR	262755548	10391945
KOR_CHN	2009	08	KOR	25393380	15668573
KOR_CHN	2009	09	KOR	26466597	3111984
KOR_CHN	2009	10	KOR	172195188	34617
KOR_CHN	2009	11	KOR	45872186	2100018
KOR_CHN	2009	12	KOR	301234547	12085296
KOR_CHN	2009	13	KOR	19439817	2050016
KOR_CHN	2009	14	KOR	1407370	11773
KOR_CHN	2009	15	KOR	15392705	6753542
KOR_CHN	2009	16	KOR	88272187	10796123
KOR_CHN	2009	17	KOR	65282333	68115595
KOR_CHN	2009	18	KOR	1730624	3956632
KOR_CHN	2009	19	KOR	88122719	72719152
KOR_CHN	2009	20	KOR	200174570	13492028
KOR_CHN	2009	21	KOR	99578283	67723884

## IV. Comtrade 무역데이터의 활용

### 1. 한국(reporter)-중국(partner) 무역거래에서 품목별(2-digit) 무역수지

```
use trade_final_7, clear
keep if reportercode==410
keep if year==2013
keep if id=="KOR_CHN"

gen bb=-im_tradevalueus
egen surplus=rowtotal(ex_tradevalueus bb)
capture drop bb
replace surplus=surplus/1000
gsort -surplus
format surplus %15.0fc
list id year commoditycode surplus in 1/5
list id year commoditycode surplus in -5/L
```



## ○ 무역수지 상위 5개 품목

```
. list id year commoditycode surplus in 1/5
```

	id	year	commod~e	surplus
1.	KOR_CHN	2013	85	21,812,986
2.	KOR_CHN	2013	90	18,197,822
3.	KOR_CHN	2013	29	12,433,140
4.	KOR_CHN	2013	39	8,921,823
5.	KOR_CHN	2013	27	7,568,429

## ○ 무역수지 하위 5개 품목

```
. list id year commoditycode surplus in -5/L
```

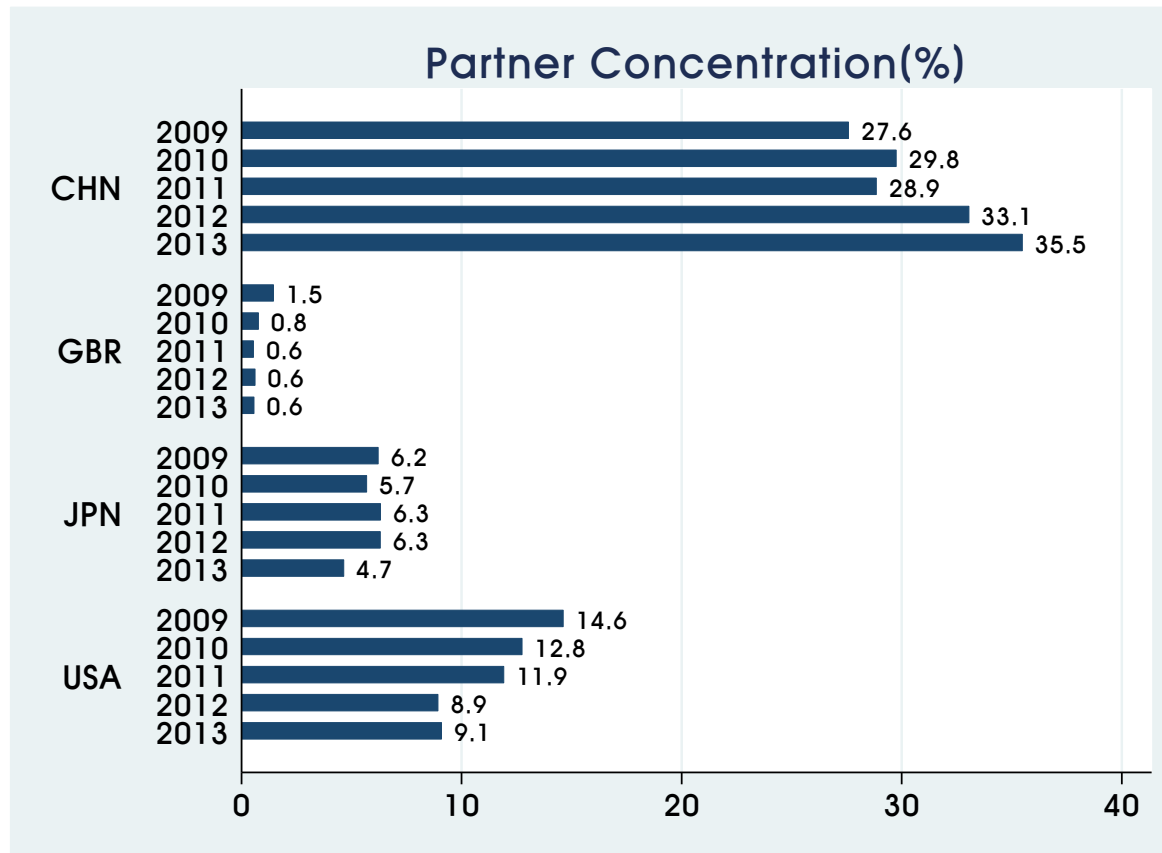
	id	year	commod~e	surplus
93.	KOR_CHN	2013	61	-1,063,337
94.	KOR_CHN	2013	70	-1,197,301
95.	KOR_CHN	2013	62	-1,936,537
96.	KOR_CHN	2013	73	-2,612,295
97.	KOR_CHN	2013	72	-3,109,739

## 2. Partner Concentration 계산

○ 특정 품목(2-digit) 수출에서 partner 국가로 수출금액이 reporter 국가의 전체 수출금액에서 차지하는 비율

○ HS Commodity code=85

```
set more off
set trace off
cd C:\WDaumCloud\W민인식작업7\trade
use trade_final_7, clear
keep if reportercode==410
keep if commoditycode==85 // electronics
gen ex_concentration=100*(ex_tradevalueus/world_ex_tradevalueus)
format ex_concen %9.1f
graph hbar (asis) ex_concen, over(year) over(partneriso) blabel(bar, format(%5.1f)) title("Partner Concentration(%)")
```



### 3. 패널데이터 만들기 (1)

○ commodity code = 81~85 (5개 품목)에 대한 패널데이터 만들기

○ 여러 품목 수출(수입)금액 간의 상관계수를 계산하거나 특정 품목수출액이 종속변수이고 다른 품목 수출액이 설명변수인 회귀모형을 추정할 때

○ 특정 품목의 수출(수입) 금액이 종속변수이고 다른 거시변수(환율) 등이 설명변수일 때

```
set more off
set trace off
cd C:\DaumCloud\민인식작업7\trade
use trade_final_7, clear
keep if commoditycode>=81 & commoditycode<=85
keep id year commoditycode ex_tradevalueus
reshape wide ex_trade, i(id year) j(commoditycode)
tsset id year
```

id	year	ex_tradevalueus81	ex_tradevalueus82	ex_tradevalueus83	ex_tradevalueus84	ex_tradevalueus85
KOR_CHN	2009	30475775	235408156	177395471	9447561188	24500144143
KOR_CHN	2010	44464776	287112961	203431960	14403998683	32971424749
KOR_CHN	2011	62764732	266736372	238271785	15840550745	34206755103
KOR_CHN	2012	74871265	321920582	261985563	12840369485	39375990335
KOR_CHN	2013	86902329	329774368	271463360	14274721035	48094867738
KOR_GBR	2009	12497832	15421040	7446732	590776066	1303316226
KOR_GBR	2010	14851104	18245221	10468518	886164654	863933094
KOR_GBR	2011	21334818	20089225	12555673	823883108	660396481
KOR_GBR	2012	4821463	26610567	11993886	754882640	751665209
KOR_GBR	2013	8972790	28932871	10095316	889178038	783702923
KOR_JPN	2009	68859979	105834707	54357519	2261984470	5527923219
KOR_JPN	2010	110798984	153287805	67970655	2732342850	6315105301
KOR_JPN	2011	193920250	195051856	78188569	3350943185	7503810858
KOR_JPN	2012	81896349	218154525	86440800	3455509642	7523120303
KOR_JPN	2013	70781580	192520440	80045518	3050780819	6310350509
KOR_USA	2009	14920025	177006655	85868628	6278093950	12994054615
KOR_USA	2010	25509210	197564008	117516949	8811107474	14136966378
KOR_USA	2011	48849802	192137854	146829182	9927338624	14143340653
KOR_USA	2012	57531329	274475861	230537796	10361213249	10644480251
KOR_USA	2013	42807090	262948800	267977016	10906099836	12331948840

#### 4. 패널데이터 만들기 (2)

○ 패널그룹 : reporter국가\_partner국가\_commoditycode

시간 : year

○ 패널데이터에서는 하나의 id 내에서 time이 반복되지 않아야 하는데, commodity 별로 수출입 금액을 데이터로 가진 경우에는 이 규칙에 어긋난다.

○ 따라서 국가+commoditycode를 하나의 그룹변수로 만든 후 tsset 명령문을 실행하면 된다.

○ 아래와 같이 실행하면 long type 패널데이터를 wide type으로 만들지 않고 tsset 명령문을 실행할 수 있다.

```
set more off
set trace off
cd C:\WDaumCloud\W민인식작업7\Wtrade
use trade_final_7, clear
gen code1=string(commoditycode) // 번호를 문자로 만든다.
replace code1="0"+code1 if commoditycode<10 // 0으로 시작하는 code가 있기 때문에 필요하다.
gen id_code=id+"_"+code1
encode id_code, gen(id_code1) // tsset을 위해서 필요하다.
order id_code id_code1 year id code1
tsset id_code1 year
```

\* *xtreg*와 같은 *panel regression* 명령어를 실행할 수 있다.

	id_code	id_code1	year	id	code1	ex_tradevalueus
2327	KOR_CHN_01	KOR_CHN_01	2009	KOR_CHN	01	30616
2328	KOR_CHN_01	KOR_CHN_01	2010	KOR_CHN	01	4640
2329	KOR_CHN_01	KOR_CHN_01	2011	KOR_CHN	01	33195
2330	KOR_CHN_01	KOR_CHN_01	2012	KOR_CHN	01	30356
2331	KOR_CHN_01	KOR_CHN_01	2013	KOR_CHN	01	8420
2332	KOR_CHN_02	KOR_CHN_02	2009	KOR_CHN	02	4409922
2333	KOR_CHN_02	KOR_CHN_02	2010	KOR_CHN	02	3756203
2334	KOR_CHN_02	KOR_CHN_02	2011	KOR_CHN	02	6553031
2335	KOR_CHN_02	KOR_CHN_02	2012	KOR_CHN	02	913196
2336	KOR_CHN_02	KOR_CHN_02	2013	KOR_CHN	02	3759817
2337	KOR_CHN_03	KOR_CHN_03	2009	KOR_CHN	03	120786302
2338	KOR_CHN_03	KOR_CHN_03	2010	KOR_CHN	03	195157503
2339	KOR_CHN_03	KOR_CHN_03	2011	KOR_CHN	03	414000247
2340	KOR_CHN_03	KOR_CHN_03	2012	KOR_CHN	03	306477170
2341	KOR_CHN_03	KOR_CHN_03	2013	KOR_CHN	03	289746739
2342	KOR_CHN_04	KOR_CHN_04	2009	KOR_CHN	04	6190812
2343	KOR_CHN_04	KOR_CHN_04	2010	KOR_CHN	04	8360320
2344	KOR_CHN_04	KOR_CHN_04	2011	KOR_CHN	04	1321554
2345	KOR_CHN_04	KOR_CHN_04	2012	KOR_CHN	04	6326756
2346	KOR_CHN_04	KOR_CHN_04	2013	KOR_CHN	04	13268888
2347	KOR_CHN_05	KOR_CHN_05	2009	KOR_CHN	05	6198419
2348	KOR_CHN_05	KOR_CHN_05	2010	KOR_CHN	05	11455826
2349	KOR_CHN_05	KOR_CHN_05	2011	KOR_CHN	05	23403523