Assessing Development Strategies to Achieve the MDGs in the Arab Region

Plenary sessions XI and XII: Implementation of the Microsimulation Methodology in STATA

Martín Cicowiez

Casablanca, December 2-5, 2008

Outline

- STATA Basics
- The Microsimulations code
- Using the Microsimulations code
- Linking MAMS and the Microsimulations

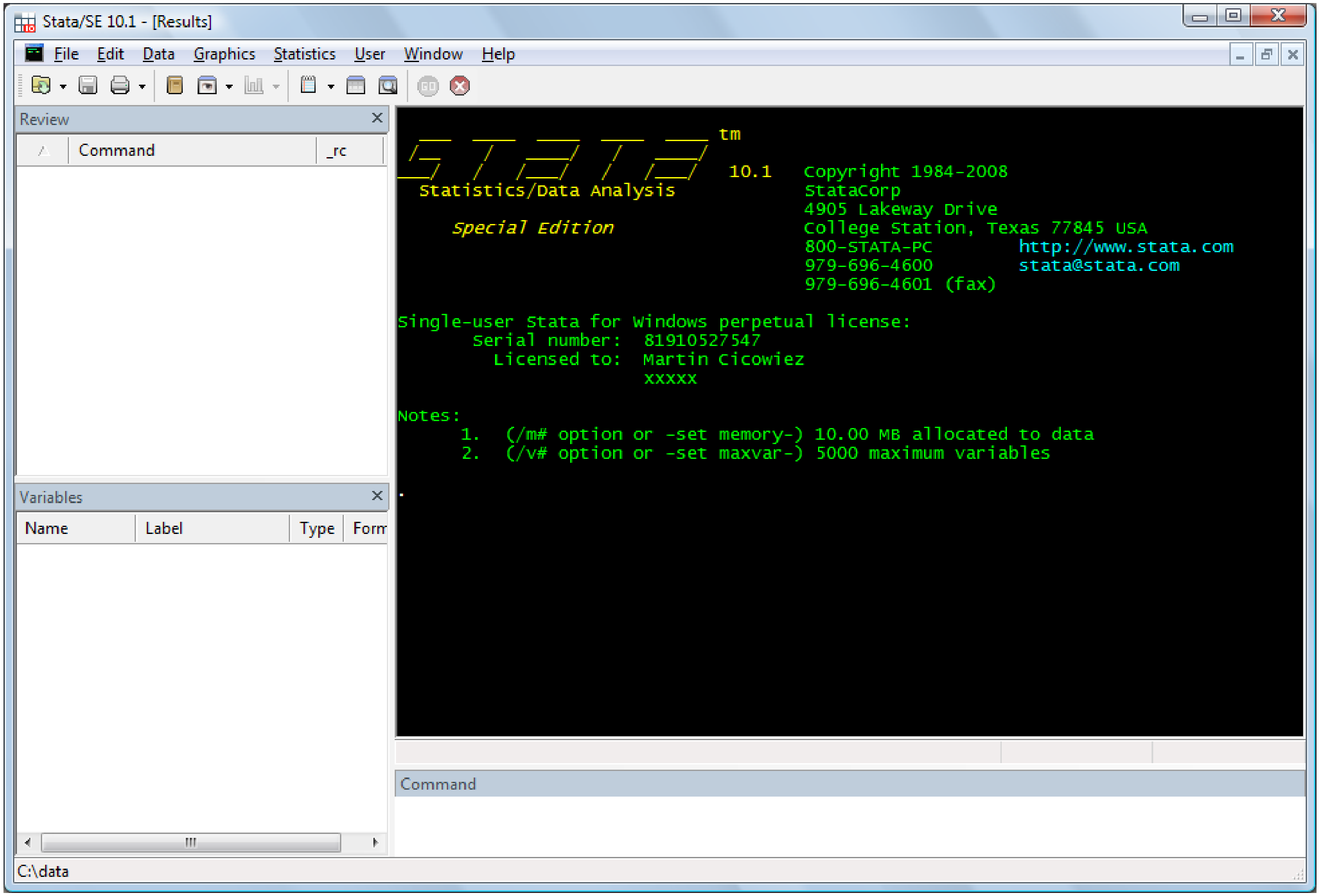
STATA Basics

- STATA is a statistical package that can be used to analyze and manipulate data, econometric analysis, and programming.
- All the information regarding STATA can be found at <*www.stata.com*>. Other useful internet resources are
  - <*www.ats.ucla.edu/stat/stata*>
  - <*www.cpc.unc.edu/services/computer/presentations/statatutorial*>
  - <*www.princeton.edu/~erp/stata*>

Installing STATA

- Run the setup.exe file.
- Run STATA for the first time and enter the license information.
- Suggestion. Read “Getting Started with STATA” and the “User’s Guide”. Read the rest of the documentation selectively, according to your needs.
- In case of a problem, use the help command.

The STATA Interface



```
  ___  ____  ____  ____  ____  tm
 /__    /   ____/   /   ____/
___/   /   /___/   /   /___/   10.1   Copyright 1984-2008
  Statistics/Data Analysis            StataCorp
                                      4905 Lakeway Drive
     Special Edition                  College Station, Texas 77845 USA
                                      800-STATA-PC        http://www.stata.com
                                      979-696-4600        stata@stata.com
                                      979-696-4601 (fax)

Single-user Stata for Windows perpetual license:
       Serial number:  81910527547
         Licensed to:  Martin Cicowiez
                       xxxxx

Notes:
      1.  (/m# option or -set memory-) 10.00 MB allocated to data
      2.  (/v# option or -set maxvar-) 5000 maximum variables

.
```

The STATA Interface

- STATA can be used by means of menus. However, it is more flexible to use the command line.
- The STATA interface is comprised of four windows: review, variables, STATA results, and STATA command.
- The commands can be introduced “immediately” trough the command line or trough text files that STATA executes line by line (i.e., do files).

The Language Syntax

The basic language syntax for STATA commands is

```
[by varlist:] command [varlist] [=exp] [if exp]
                      [in range] [weight]
                      [using filename] [, options]
```

where the elements between brackets are optional.

Some Basic Commands

- clear everything
  - `clear`
- setting the memory size for the database
  - `set mem 100m`
- setting the path
  - `cd “C:\Data”`
- importing an EXCEL database
  - `insheet using “my-data.csv”, comma`

Some Basic Commands – cont.

- computing basic statistics
  - `summarize ypc`
  - `summarize ypcf [w=popwt]`
  - `summarize ylab [w=popwt] if age >=25 & and age <=55`
- generate new variables
  - `generate ypc2 = ypc^2`
- tabulate data
  - `table skill [w=popwt], c(mean ylab)`

Some Basic Commands – cont.

- renaming variables
  - `rename ypc2 ypcf22`
- eliminating variables
  - `drop ypc22`
- replacing values
  - `replace male=0 if male==1`

Do Files and Log Files

- A do file is a text file with STATA code that STATA runs line by line, as if the sentences where written in the STATA command window.
- A log file is a text file with all the results that appear in the STATA results window.
  - the user selects when to start and when to stop logging to the log file

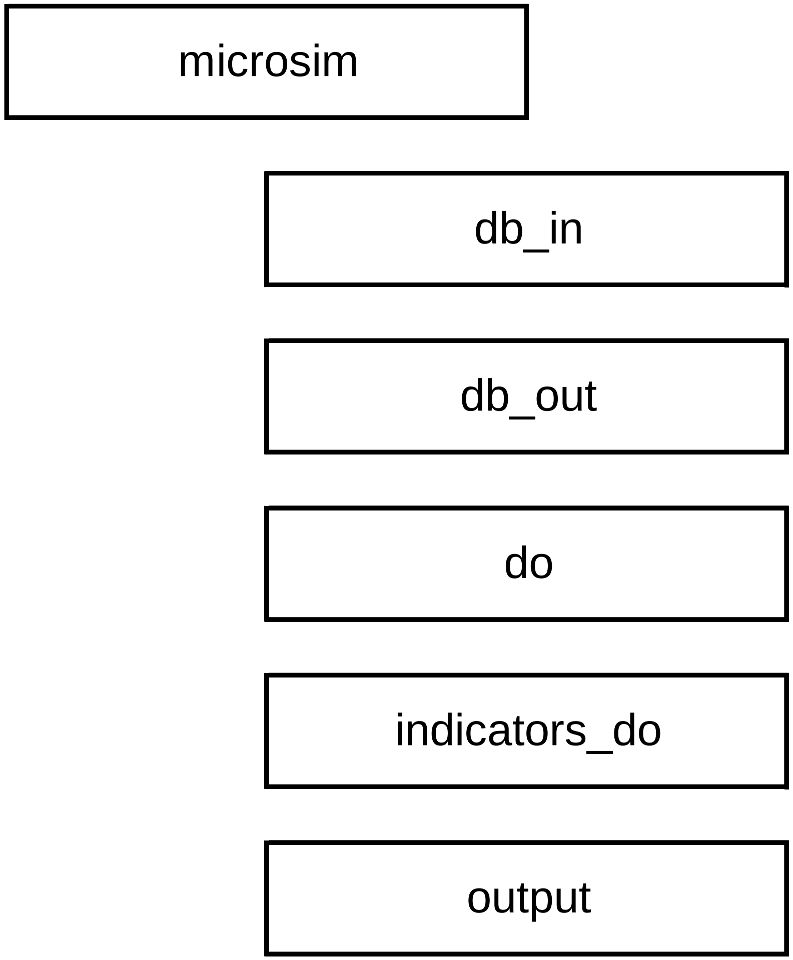
Microsimulations Files

- files to add to the MAMS folder
  - run-db-mams-ms.xls
  - test-db-mams-ms.xls
  - yem-db-mams-ms.xls
  - db-mams-ms.gms
- microsimulation package (microsim-example-2008-12-04.zip) containing two main folders – use a dedicated folder (see below)
  - link-mams-ms
  - microsim

The Microsimulations Code: Two Flavors

- stand-alone
- linked with MAMS results

The Microsimulations Code: Folder Structure



microsim

db_in

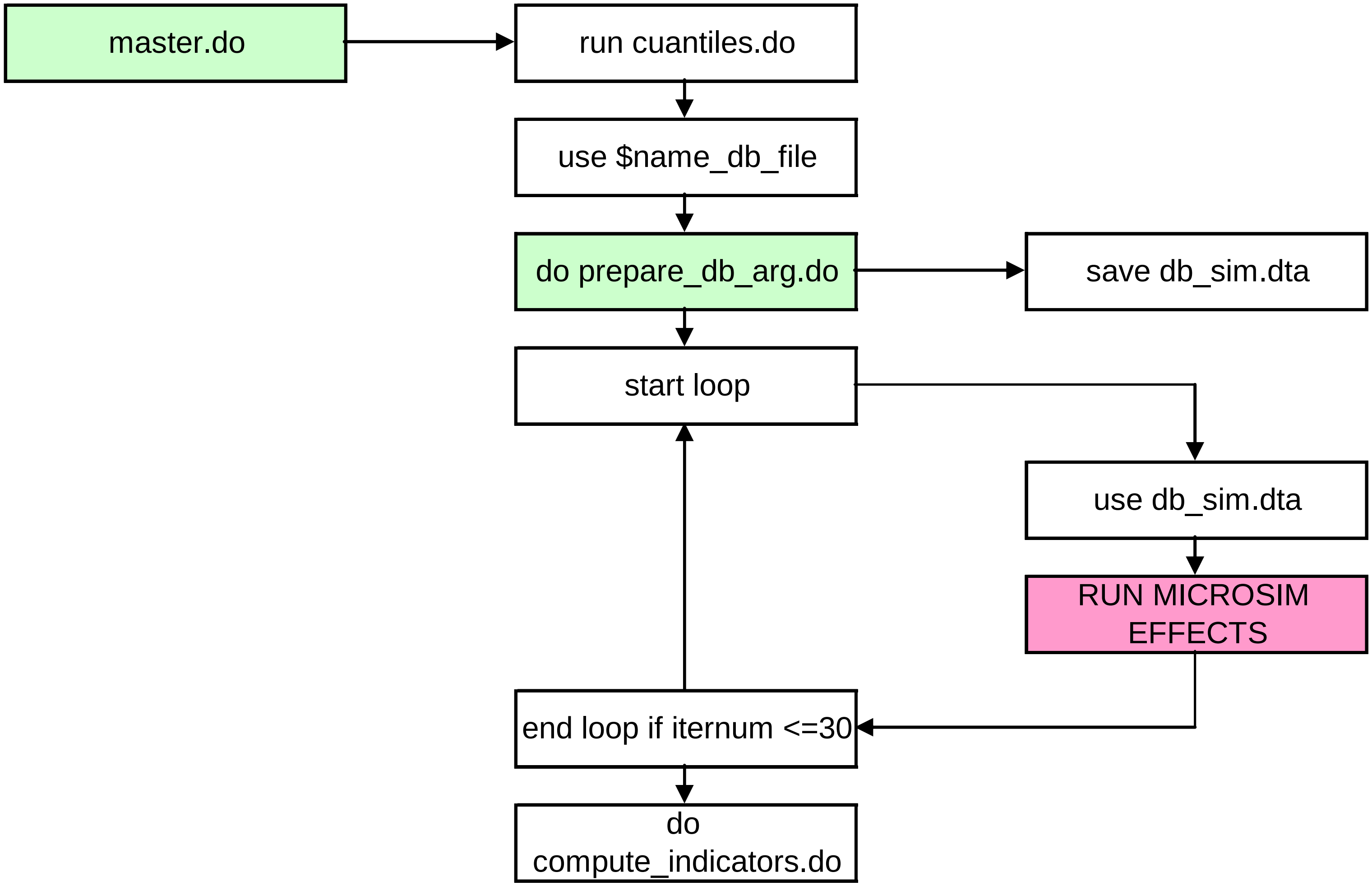
db_out

do

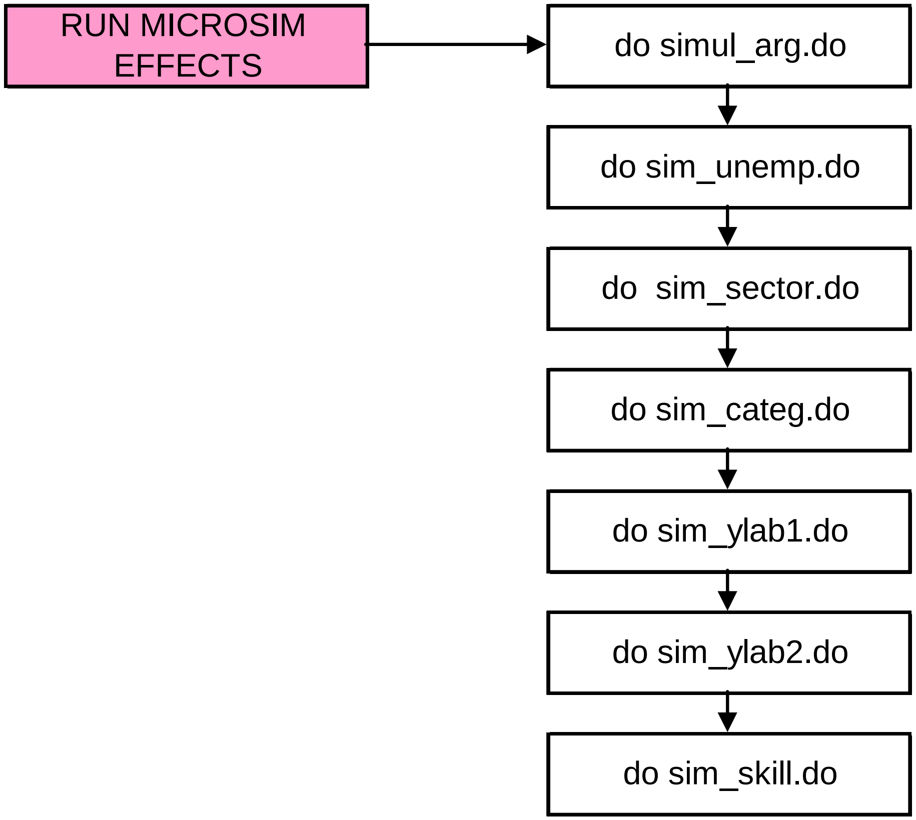
indicators_do

output

The Microsimulations Code: File Structure



The Microsimulations Code: File Structure



RUN MICROSIM EFFECTS

do simul_arg.do

do sim_unemp.do

do sim_sector.do

do sim_categ.do

do sim_ylab1.do

do sim_ylab2.do

do sim_skill.do

Installing and Preparing the MS code

- Create the folder C:\microsim-example.
- Unzip the file microsim-example-2008-12-04.zip in the folder C:\microsim-example.
  - keep the folder structure!
- Make a copy of the file prepare_db_arg.do to prepare_db_<app>.do.
- Make a copy of the file simul_arg.do to simul_<app>.do.

Installing and Preparing the MS code – cont.

- Save your raw household survey data in the folder “C:\microsim-example\microsim\db_in”.
- Adjust the prepare_db_<app> file as required based on your household or labor force survey – generate the contry-specific variables (see below).
- Adjust the simul_<app> file as required base on the shocks you want to simulate.
- Adjust the master.do file – change the paths stored in global macros (see below).

Database: Household Variables

VARIABLE	DESCRIPTION	UNITS / VALUES
id	household identifier	number / string
size	household size (number of members)	number
pondwt	population weights	number
yh	household total income	number
ypc	household per capita income	number
pl_moderate	moderate poverty line	number
pl_extrema	extreme poverty line	number
pl_1usd	1 US$ poverty line	number
pl_2usd	2 US$ poverty line	number

Database: Individual Variables

VARIABLE	DESCRIPTION	UNITS / VALUES
age	age	number
male	gender (*)	1 = male 0 = female
male_rep	gender (**)	1 = male 0 = female
skill	skill level	number
status_lab	labor status	1 = employed 2 = unemployed 3 = inactive
member	membership to the (labor market) microsim	1 = included 0 = no included

(*) in case this disaggregation is not present in CGE, assign value 1 to all -- our case.

(**) it is used for reporting results by gender.

Database: Employed Individuals Variables

VARIABLE	DESCRIPTION	UNITS / VALUES
ylab	labor income	number
sector	sector of employment	number ex: 1=agr, 2=mnf, 3=svc
categ	occupational category (*)	number ex: 1=wage-earner, 2=self-employed

(*) in case this disaggregation is not present in CGE, assign value 1 to all -- our case.

(**) it is used for reporting results by gender.

Preparing the Database

- In preparing the database for the MS it is important to follow the instructions in the prepare_db_arg.do (prepare_db_<app>) file.

Adapting the Master File

- Adjust the different access paths -- stored in global macros
  - *path_do*
  - *path_db_in*
  - *path_db_out*
  - *path_log*
  - *path_indicators_do*
- Adjust the name of the database – see global macro *name_db_file*.
- Adjust the name of the file that prepares the database – from prepare_db_arg.do to prepare_db_<app>.do.
- Select the number of iterations to perform – used to estimate confidence intervals for the results.

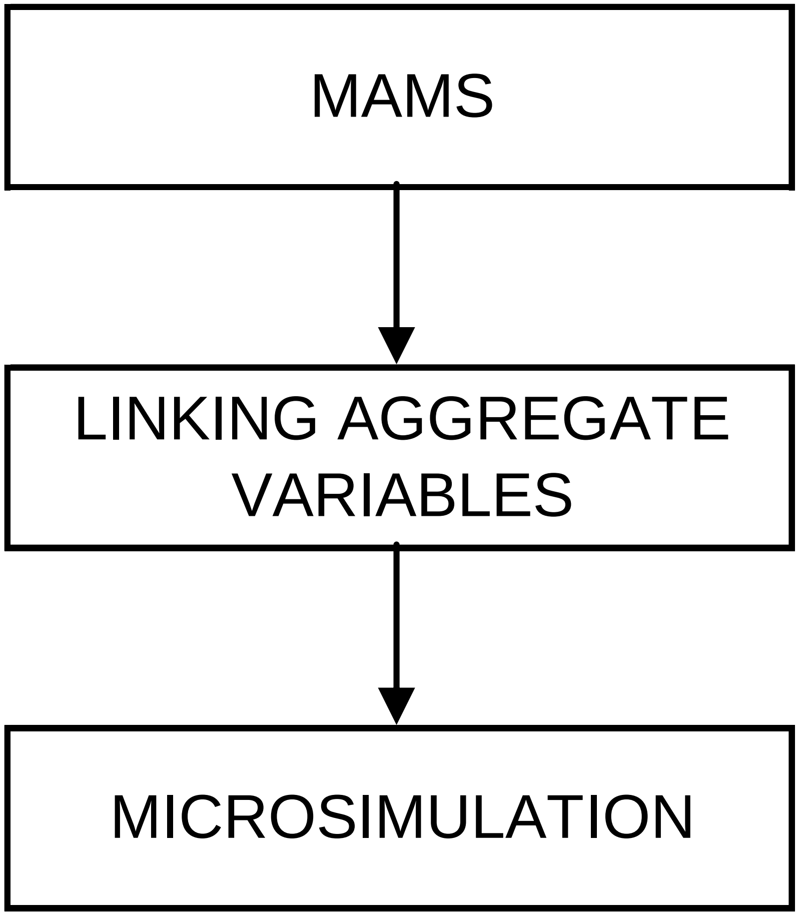
An Example

- Simulate
  - 50% decrease in the unemployment rate for f-labn
  - 50% decrease in the unemployment rate for f-labs
  - 50% decrease in the unemployment rate for f-labt
  - keep the other changes in simul_arg.do

An Example – cont.

- The results for each iteration are stored in the file microsim__.csv in C:\microsim-example\microsim\output\.
- The confidence intervals are stored in the file intervalos__.log in C:\microsim-example\microsim\output\.

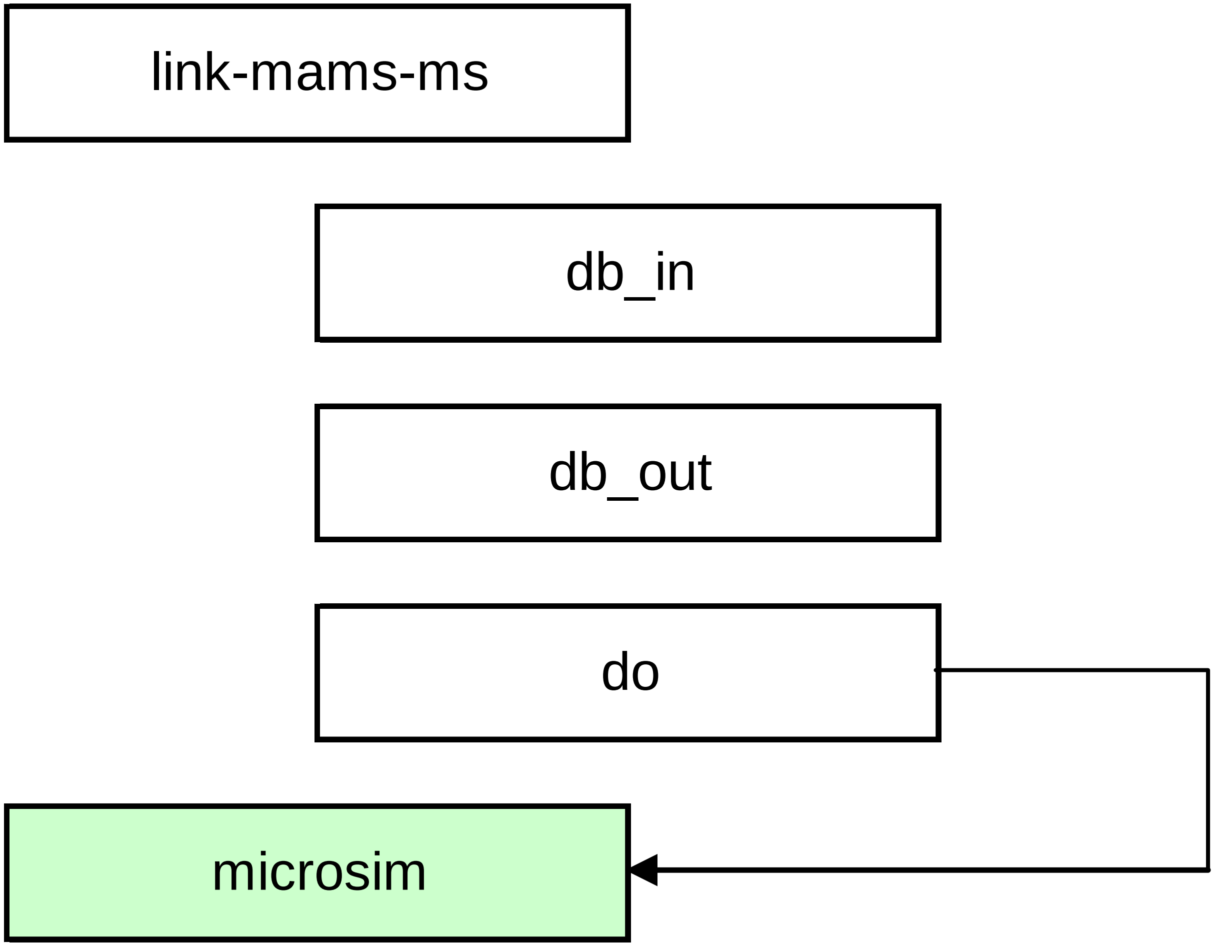
Top-Down Approach to CGE-Microsimulations


MAMS

LINKING AGGREGATE VARIABLES

MICROSIMULATION

Linking MAMS and the MS: Folder Structure


link-mams-ms

db_in

db_out

do

microsim

Steps in Linking MAMS and MS

- Copy the files test-db-mams-ms.xls and run-db-mams-ms.gms to your MAMS folder (e.g., MAMS-example).
- Make a copy of the file test-db-mams-ms.xls to <app>-db-mams-ms.xls – not needed now!
- Make changes to the sets aagg (activities in the microsimulations) and tt in <app>-db-mams-ms.xls.
  - usually, the set aagg is an aggregation of the set a (activities in MAMS)
  - tt refers to the years for which the microsimulations are run

Steps in Linking MAMS and MS -- cont

- Open the file db-mams-ms.gms and do a search and replace, replacing *test-* by <*app*>*-*.
- Run the GAMS file db-mams-ms.gms using the EXCEL file run-db-mams-ms.xls.
  - creates the CSV file mams-results.csv, in the MAMS folder
    - contains MAMS results that can be read by STATA

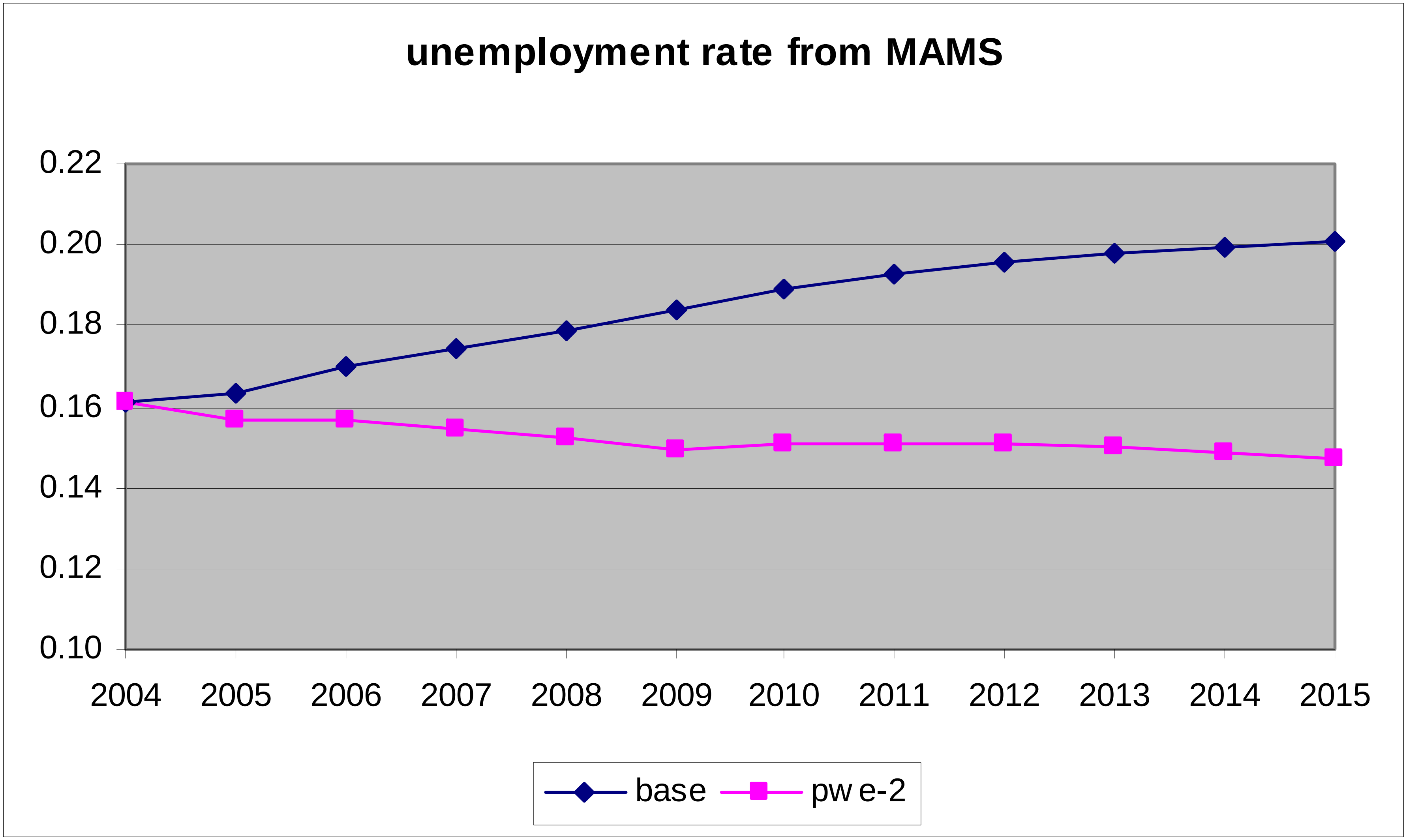
Steps in Linking MAMS and MS -- cont

- Run the master2.do file located in the folder link-mams-ms\do.
  - uses the results from MAMS for the aggregate linking variables to run the microsimulations
  - runs the microsimulations for the scenarios in simcur
- Once finished, the results can be found in the file microsim_all, it will be located in the folder link-mams-ms\output.
- Confidence intervals for each year in each simulation can be found in the folder microsim\output.

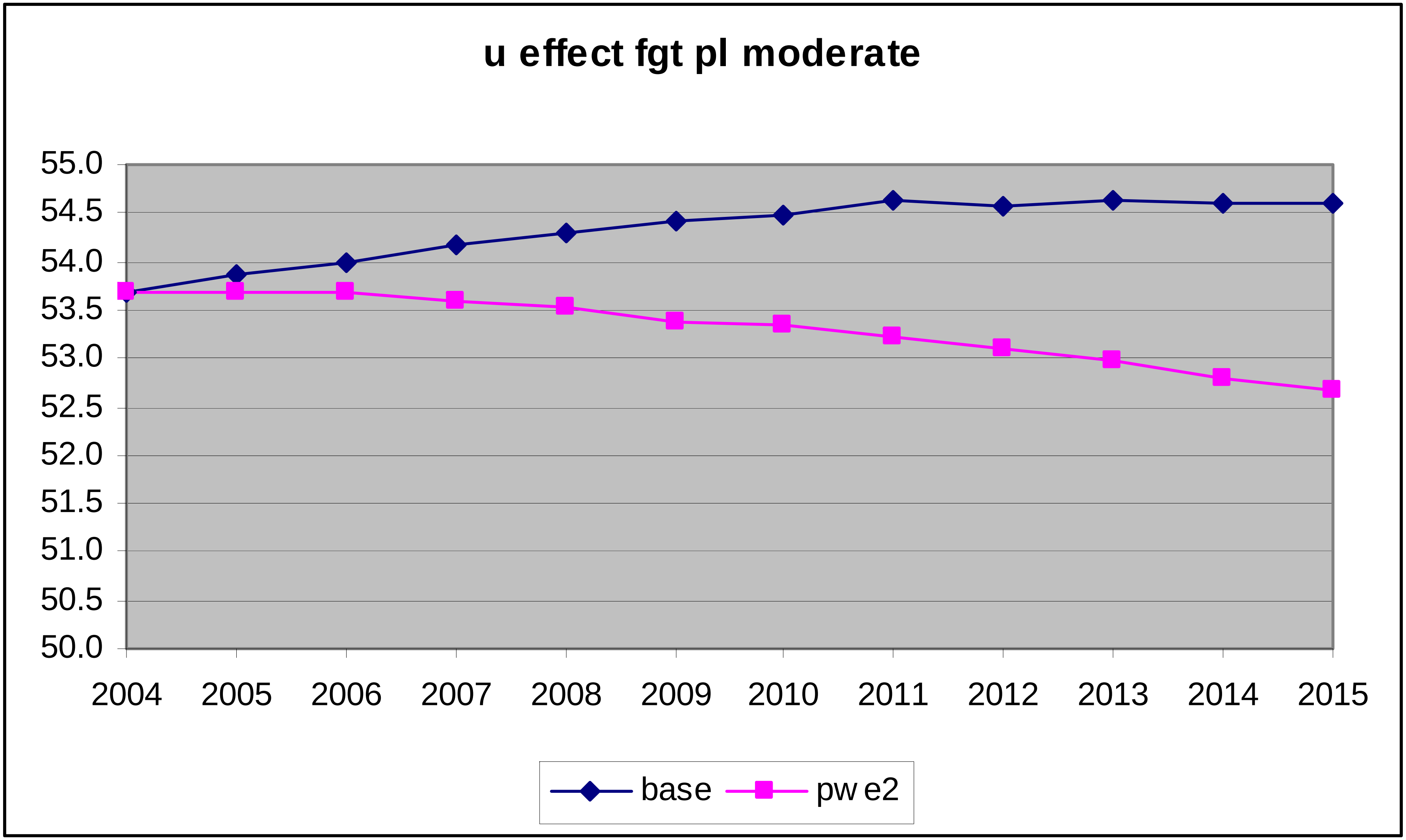
An Example

- Run the base and pwe-2 scenarios using the model in MAMS-example
  - yem-data-general.xls
  - yem-sim-mdg.xls
- Run the db-mams-ms.gms file by using the run-db-mams-ms.
- Run the microsimulations using master2.do, located in C:\microsim-example\link-mams-ms\do.

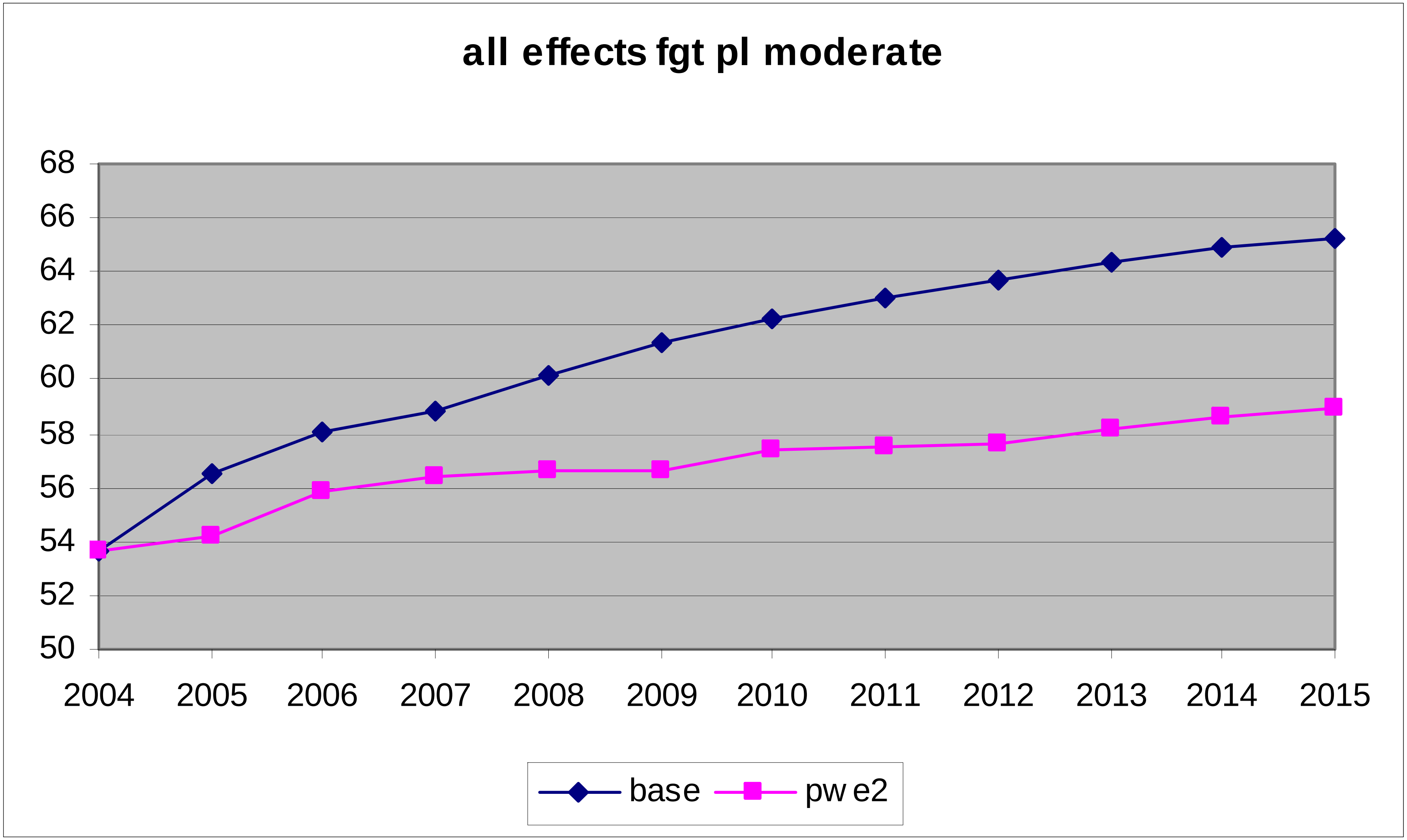
An Example – cont.



An Example – cont.



An Example – cont.


all effects fgt pl moderate
base
pw e2