

Running GTAP in R

A presentation for the 23rd Annual Conference on Global
Economic Analysis “Global Economic Analysis Beyond 2020”
on June 17–19, 2020

Maros Ivanic

Economic Research Service, United States Department of Agriculture

The findings and conclusions in this paper are those of the author and should not be construed to represent any official USDA or U.S. Government determination or policy.

GEMPACK has greatly facilitated GE analysis

GEMPACK brought much needed standards to general equilibrium analysis

- ▶ A common language (TABLO) to formulate equations needed to perform GE analysis
- ▶ Made it relatively easy to replicate models or reuse them
 - ▶ A model could be shared among researchers
- ▶ Easy to specify shocks and share them with others
- ▶ Fast to run
 - ▶ GEMPACK is created to solve linear models very efficiently
- ▶ Additional supporting tools to make the life of modelers easier (e.g., AnalyzeGE, ViewSol)

After 36 years of service GEMPACK is getting outdated

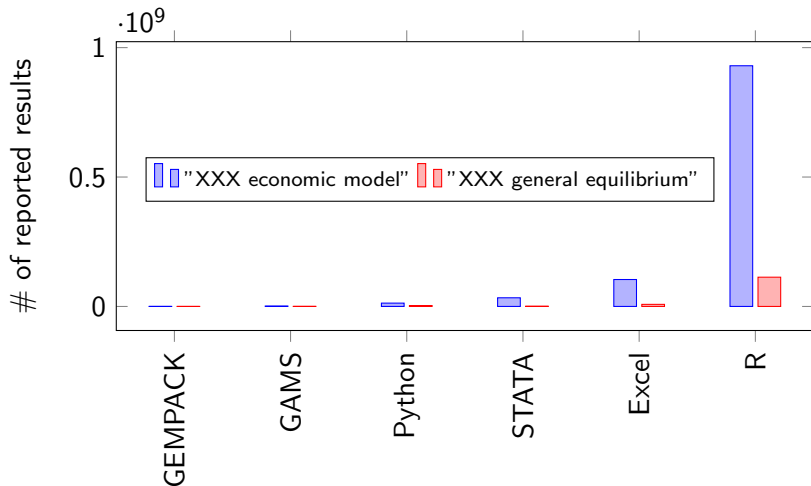
- ▶ Version control is a bit hard to implement in TABLO
 - ▶ It is easy to break the model by introducing a change because of the many connections and the use of a single file (e.g., objects have to be declared)
- ▶ Integration is extremely hard
 - ▶ GEMPACK works with custom data formats (HAR files work well with data tapes but cannot be understood by modern software)
- ▶ The proprietary nature of GEMPACK results in slow response of the system to the users' needs
 - ▶ Users are unable to create and share their own addons or improvements to GEMPACK
- ▶ The users cannot easily use their preferred work environments to conduct analysis
 - ▶ It is possible to avoid TABMATE, but much harder to avoid AnalyzeGE
 - ▶ The GEMPACK-provided tools have not kept up with the rest of the software development in terms of integration and standards

Can open-source solution replace GEMPACK?

- ▶ Open-source platforms benefit from the collective development effort by the entire community
 - ▶ Everyone who wants to contribute is welcome to do so
 - ▶ The community decides whose solution is the best by downloading it, using and contributing to it
- ▶ Open-source solutions tend to be modularized which allows improvements to be shared and generate synergies
 - ▶ If two systems share common components (e.g., an efficient method to solve a linear system, a more beautiful way of plotting a dat set) then fixing one component for one system also fixes it for all others that use it
- ▶ Open-source systems tend to be available for many more operating systems
 - ▶ There is always someone who wants or needs to make it work on Androids, iOS etc. opening technological frontiers for all
- ▶ **Free to use!**

R appears to dominate economic modeling efforts

Number of results returned by Google for a search string "XXX economic model" and "XXX general equilibrium" (e.g., "R economic model", "STATA general equilibrium", etc.) on www.google.com (June 16, 2020)



Moving GEMPACK (and GTAP) into R

Key steps

- ▶ A reader for HAR files
 - ▶ Move the data from an obscure binary format into a standard data structure (list)
- ▶ An interpreter of the TABLO language
 - ▶ Translate business logic contained in TABLO files to standard R sets, coefficients, expressions
- ▶ A solver to find the solution to the system
 - ▶ Generate a solver for the system that operates roughly like the GEMPACK solver (e.g., provide shocks to exogenous variables, obtain endogenous variable values)

Reading HAR files into R directly

Enabling the use of GTAP Center's data

- ▶ GTAP data are normally distributed as a set of HAR files
- ▶ HAR files represent a fairly dated way of storing data as they appear to be designed to be capable of rewinding and fast-forwarding (think data tapes)
- ▶ We could not find a modern system that can read HAR files directly
- ▶ The GEMPACK provided routines for data extraction (e.g., Seehar) are buggy and limited in the types of conversion they can support (e.g., GAMS output, on particular flavor of SQL script without escaping apostrophes)

Package HARr

Reading HAR files into R directly

- ▶ HARr is able to read the binary HAR files
- ▶ Can interpret real, integer and character headers
 - ▶ Also interprets sets in those headers
- ▶ Outputs an R list with the data
- ▶ Reverse engineered and therefore may need to be updated if the specifications for the HAR files change
 - ▶ Tested on various outputs generated by GTAPAgg
- ▶ Available in GitHub at
<https://github.com/USDA-ERS/MTED-HARr>
- ▶ **The community is welcome to use and improve as needed**

Package TabloToR

Interprets a TABLO model and solves it

- ▶ Understands and interprets the following statements: file, set, coefficient, formula, variable, equations
- ▶ Understands the difference between change and percent change variables
- ▶ Creates an abstract R object of the TABLO model (reference class) which can
 - ▶ Load the data (provided by HARr)
 - ▶ Specify shocks (the values of exogenous variables)
 - ▶ Solve with a specified number of subintervals
- ▶ Available in GitHub at <https://github.com/USDA-ERS/MTED-TabloToR>
- ▶ **The community is welcome to use and improve as needed**

Solver in TabloToR

Using the existing solvers for linear systems in R

- ▶ When processing the coefficients and equations, TabloToR creates a sparse $N \times M$ matrix M of variable weights in each equation where N is the number of equations and M number of variables
- ▶ The shock specification should provide a vector of $N - M$ values
- ▶ TabloToR then generates a linear system to solve: $Ex = -Ny$ where E is a square matrix of the endogenous components of M , x is a vector of unknown endogenous variables, N is the section of M representing exogenous columns, y is the set of shocks
- ▶ To speed up the calculations TabloToR identifies and backsolves all equations that determine a single variable
- ▶ Package Matrix provides a relatively efficient way of solving a linear system
- ▶ The process can be repeated for multi-pass solutions

An example of solving GTAP in R code

```
1 # Initialize a new object of class GEModel
2 GTAP = tabloToR::GEModel$new()
3
4 # Load a TABLO file into the object (and interpret it)
5 GTAP$loadTablo('gtap.tab')
6
7 # Read in the data from HAR files
8 data = list(
9     GTAPSETS = HARr::read_har('sets.har'),
10    GTAPPARM = HARr::read_har('default.prm'),
11    GTAPDATA = HARr::read_har('basedata.har')
12 )
13
14 # Load the data to the model
15 GTAP$loadData(data)
16
17 # Specify shocks (a vector of values for all exogenous variables)
18 GTAP$setShocks(shocks)
19
20 # Solve the model in a single iteration (=Johansen)
21 GTAP$solveModel(iter = 1)
```

Comparison of runs between GEMPACK and R

- ▶ To compare the results between GEMPACK and R, we perform a simple simulation in both systems and compare the results
- ▶ GTAP Database 10
 - ▶ 10 regions (Oceania, East Asia, Southeast Asia, South Asia, North America, Latin America, EU, MENA, SSA, ROW)
 - ▶ 5 factors (Skilled Labor, Unskilled Labor, Capital, Land, Natural resources)
 - ▶ 20 commodities
- ▶ Shocks
 - ▶ 20 percent reduction in UnSkLab in Latin America, North America and the EU
- ▶ Solution methods
 - ▶ Linear: one-pass in both R and GEMPACK (Johansen)
 - ▶ Multipass: six-masses in R, Gragg 1-3-5 in GEMPACK

Results—Single iteration in R vs. Johansen in GEMPACK

Comparison of the two runs

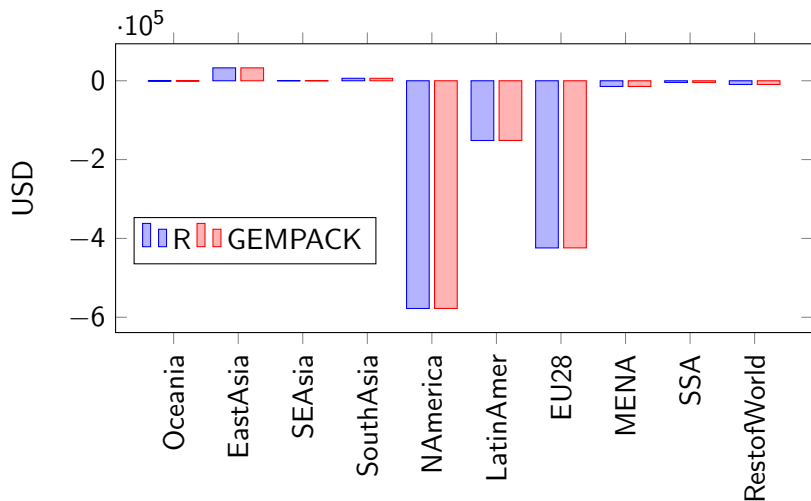


Figure: EV

Results—Single iteration vs. Johansen in GEMPACK

Comparison of the two runs

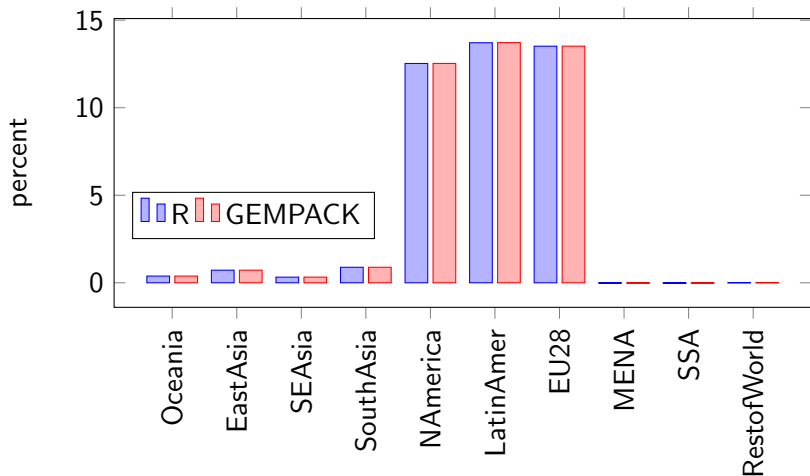


Figure: pm[UnSkLab,]

Comparison of the values—single iteration in R, Johansen in GEMPACK

Table: EV

	GEMPACK	R
Oceania	-877.46	-877.46
EastAsia	32765.12	32765.12
SEAsia	117.91	117.91
SouthAsia	6273.46	6273.47
NAmerica	-577674.06	-577674.29
LatinAmer	-151445.88	-151445.86
EU_28	-424023.56	-424023.54
MENA	-14477.49	-14477.48
SSA	-4209.67	-4209.67
RestofWorld	-9370.85	-9370.85

Table: pm["UnSkLab",]

	GEMPACK	R
Oceania	0.382161	0.382161
EastAsia	0.721118	0.721118
SEAsia	0.323598	0.323599
SouthAsia	0.885713	0.885714
NAmerica	12.528844	12.528842
LatinAmer	13.717290	13.717290
EU_28	13.516024	13.516023
MENA	-0.019788	-0.019788
SSA	-0.027744	-0.027743
RestofWorld	0.006760	0.006761

Results—6 iterations vs. Gragg

Comparison of the two runs

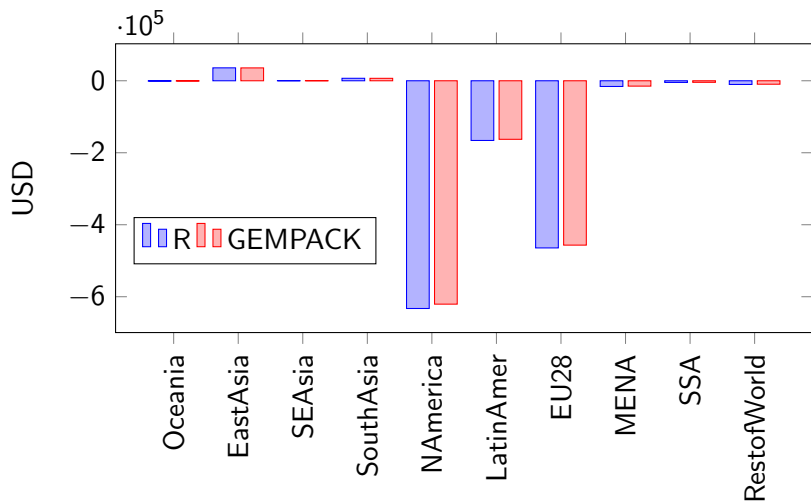


Figure: EV

Results—6 iterations vs. Gragg

Comparison of the two runs

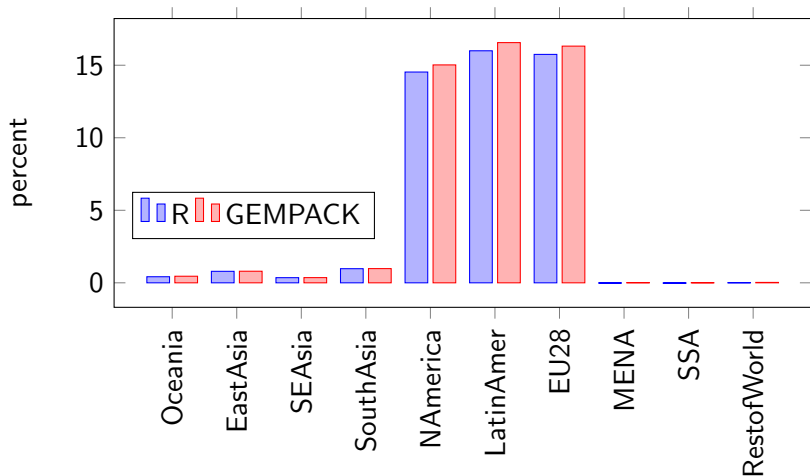


Figure: pm[UnSkLab,]

Comparison of the values—six iterations in R, Gragg in GEMPACK

Table: EV

	GEMPACK	R
Oceania	-732.90	-961.02
EastAsia	35736.14	35885.19
SEAsia	192.99	129.14
SouthAsia	6769.05	6870.86
NAmerica	-620605.19	-632683.60
LatinAmer	-162656.00	-165867.36
EU_28	-456636.84	-464401.39
MENA	-14891.53	-15856.11
SSA	-4308.89	-4610.54
RestofWorld	-9581.44	-10263.19

Table: pm["UnSkLab",]

	GEMPACK	R
Oceania	0.453670	0.419283
EastAsia	0.798568	0.792391
SEAsia	0.358550	0.354937
SouthAsia	0.979884	0.973986
NAmerica	15.022247	14.530789
LatinAmer	16.559319	15.995961
EU_28	16.316963	15.746737
MENA	0.012397	-0.021670
SSA	0.004999	-0.030381
RestofWorld	0.027524	0.007405

Benchmarks

Running GTAP in R is currently much slower

- ▶ A model of 10 regions, 20 commodities and 5 factors takes about 800 seconds to complete six iteration in R (little more than two minutes per iteration)
 - ▶ It takes just a few seconds to complete this in GEMPACK
- ▶ Parallelization could reduce this time substantially (expect less than 20 seconds with eight cores for this model)
- ▶ The time saved processing the results from GEMPACK should compensate for this time loss

Ideas for future development

In no particular order

- ▶ Implement more efficient methods for solving linear systems
 - ▶ Find or develop a better package for solving larger sparse models faster
 - ▶ Break the solving of linear models into smaller tasks to allow parallelization
- ▶ Increase the precision of the solution
 - ▶ GEMPACK-like solution extrapolation
 - ▶ Taylor series expansion
- ▶ Create a modular language to replace TABLO
 - ▶ Support mixing and matching of smaller modules, e.g., for demand, supply etc.)
 - ▶ Support version control

Conclusions

- ▶ The R framework can be used to run reasonably sized GTAP models with the original TABLO definition and original HAR files and with reasonable accuracy
- ▶ The run times in R are significantly higher (minutes instead of seconds) but it is likely that with collaborative effort the run times can be shortened
- ▶ Opening the framework to the community of GE modelers is likely to produce creative and time-saving solutions
 - ▶ Perhaps TABLO language could be simplified
- ▶ The quality of the solution in R is probably still below GEMPACK's Gragg solution
 - ▶ It would be easy to implement an existing extrapolation solution (an existing package) or develop a new one
- ▶ The model can be fully controlled by existing user-interface solutions (e.g., web-based Shiny)
- ▶ The results from the GTAP model in R are immediately available for a large number of post-simulation tools (e.g., visualization, database update etc.)