Minutes of the Gekko steering comittee meeting, Oct. 29, 2020, virtually on Zoom[1]

Participants:

- Asger Olsen (Statistics Denmark/ADAM)
- Tony Maarsleth Kristensen (Statistics Denmark/ADAM)
- Michael Osterwald-Lenum (Statistics Denmark/ADAM)
- Rasmus Bjerre (Ministry of Finance)
- Mads Harslund (Ministry of Finance)
- Rasmus Rold Sørensen (Central Bank of Denmark)
- John Smidt (Danish Economic Councils)
- Thomas Thomsen (Gekko editor)

Note: In the following, version "2.4" covers Gekko versions 2.2, 2.4, and 2.5.x, whereas version "3.0" covers Gekko versions 3.0 and 3.1.x. Most current development is on 3.1.x.

Re 1. Welcome + choice of minute taker + approval of last year's minutes.

Asger bid the participants welcome. Thomas was chosen as minutes taker, and Michael as moderator. There were no comments regarding last year's minutes.

Re 2. Editor's report regarding 2019-20.

Thomas summed up the editor's report regarding 2019-20.

Asger asked whether Statbank Denmark (Statistikbanken.dk) ought to support the Apache Arrow format too, in addition to PC-Axis files, csv., etc? Thomas said that in his view, Arrow is a much faster, more compact and generally better format than the current text-based data files (like PC-Axis). There would be some questions regarding how to implement metadata in Arrow files though. A Statbank data table *is* a dataframe, so an Arrow implementation would make a lot of sense.

Apart from that, there were no comments or objections to the report.

Re 3. Status regarding Gekko 2.4 and 3.0

This is covered by the editor's report.

[1] Links: Gekko main webpage: www.t-t.dk/gekko, organization: www.t-t.dk/gekko/organization. Gekko on GitHub: https://github.com/thomsen67/GekkoTimeseries.

Re 4. Feedback regarding Gekko 2.4 and 3.0

Rasmus R. S. expressed interest in the Gekcel project, since the Central Bank (besides using Gekko) also still uses AREMOS a lot, and they often use an AREMOS add-in for Excel, when producing plots etc. Gekcel is developed in collaboration with the Quarterly National Accounts (Statistics Denmark). Gekcel runs under Gekko 3.0 and will not be available for Gekko 2.4.

The Central Bank still uses Gekko 2.4 for Mona model simulations proper, and a mix of Gekko 3.0 and AREMOS for other systems. They are considering migrating to Gekko 3.0 for the Gekko 2.4 and AREMOS parts, but translating the AREMOS system is a quite large undertaking.

John mentioned that the Danish Economic Councils had migrated from AREMOS and Gekko 2.4 to running all these systems in Gekko 3.0 for the last year or so. Basic estimations etc. are also performed in Gekko 3.0, and it is nice that for instance recursive parameter plots etc. are now an easy click away. Apart from quite some getting used to when migrating from AREMOS, it works out well. John mentioned Gekko error messages that – he tried to put it politely -- leave something to be desired. These cryptic error messages are a hindrance for new users, to which Thomas agreed. John would send Thomas his own list of bugs and incomprehensible error messages as seen with the eyes of a new Gekko user, and in general such feedback is very helpful.

Tony mentioned that Statistics Denmark/ADAM have migrated to using Gekko 2.5.x for simulations. This is because of the usefulness of the model module/submodel capabilities of this version, and more of such model capabilities would be most welcome.

On behalf of Daniel Freyr Gustafsson from the Quarterly National Accounts (Statistics Denmark), who could not attend the meeting, Thomas mentioned that he (Thomas) had been involved in their migration from AREMOS to Gekko 3.0. They are going to run AREMOS and Gekko in parallel in the coming months, and subsequently run their systems Gekko-only in February 2021. The migration seems to work well, apart from a few cases where it seems there are some differences between AREMOS and Gekko. This *could* be due to precision and rounding errors, and it does not seem to be a major issue. Daniel has contributed to the C# source code on GitHub (specifically the Gekcel part of Gekko).

Michael was a bit worried that moving the user forum to Stack Overflow would scare off potential new users. Gekko could have its own tag at Stack Overflow, for instance #gekkotimeseries, but the question is whether consulting Stack Overflow would still feel overwhelming?

Thomas could not answer completely on behalf of the MAKRO team (DREAM), but to his knowledge they use Gekko 3.0 a lot for interacting with data and the GAMS solver etc. The new DECOMP needs to be finished and made available also for other Gekko users. As mentioned in the editor's report, there are some videos available demonstrating how the component can turn decomposition into a kind of pivot table (of which the existing DECOMP is a special case).

Rasmus B. mentioned that they are still running Gekko 2.4 (2.2. to be exact), primarily for simulations with ADAM. This is running very well, and they do not object to developing the data parts of Gekko too, but he emphasized that in the future they need to be able to run the MAKRO

model as well. Because of this, he feels that from their point of view, the resources for new developments are best used in relation to the needs of the MAKRO model project. Simulations in 2.4 run well, and if the simulations crash, it is often because of faulty human input, rather than the software. Mads agreed with Rasmus B.

Re 5. Prioritization of main lines regarding 2020-21

Regarding 1 (higher frequencies), Thomas mentioned that this project would run its cause, since it was already funded by the Economic Council. However, within this project, it would not be possible to address all issues and questions regarding higher frequencies, and Thomas envisioned getting the most interesting parts done, and then the rest of the issues could perhaps be done more gradually on a need basis. The focus is on daily and weekly frequencies. Tony was happy to hear about the latter, since such data is prevalent in their systems. There is a blueprint available describing the proposed implementation of daily and weekly frequencies in Gekko, and if no-one objects, something like this is going to be implemented.

Item 2 and 3 (DECOMP and Gekcel) were mostly a question of supplying a package of these existing components for others to use. Michael asked whether something like Gekcel could be made avaible for Open Office (Libre Office) to avoid being tied to Excel? Thomas answered that the interface between Gekko and Excel uses a component that makes it possible for Excel to talk to .NET, and this component is tailored to that end. Doing something like this for Open Office would be a completely different project.

Item 4 (Apache Arrow): EXPORT<arroos> is up and running, and IMPORT<arrow> would be really easy to do, too. So this project is more a question of how the architecture of dataframes is to be designed. Regarding array-series etc. Tony mentioned the issue of RAM in the existing 32-bit Gekko, and the RAM issues of a 32-bit system. Therefore, they envisioned porting some systems to R or Python to deal with larger datasets, and Tony asked whether these Arrow files would not become very large? Thomas said that they do potentially, but the format is built for big data under RAM restrictions, and therefore the arrow-files can be read or written as chunks/batches. At the moment, Gekko does not compress the arrow-files on desk, but this could be done easily.

Asger asked about whether Gekko ought to transfer to 64-bit to address the memory issues. Thomas wouldn't mind transferring the whole of Gekko to 64 bit, and he mentioned that GAMS went 64-bit only, about a year ago. Cross-compiling to both 32- and 64-bit could perhaps be considered. The IT departement of Statistics Denmark could help out regarding possible hard-to understand 64 bit compilation errors.

Michael asked if it would be possible to first transfer the 2.4 version (specifically 2.5.x) to 64-bit, to see if Gekko could handle for instance the large datasets in the foreign trade statistics? Thomas agreed that this would be possible, and if this was done successfully, bumping 3.0 to 64-bit as well would be easy. It was a bit hard to estimate how long it would take to transfer to 64-bit, but it could be anything from two days to two weeks of work, but probably not longer. [Note: since the meeting, Gekko has been compiled successfully for 64-bit: expect 64-bit versions soon].

About items 5-6 (modelling improvements), Thomas mentioned that it was both a question of augmenting the .frm file syntax to accommodate loops, sums over sets/lists etc., but also a question of interacting with and composing the equations and blocks of equations from .gcm files, and perhaps even create some of the equations via the OLS command etc. Tony mentioned their up-coming ADAM versions containing long-run levels of variables, and the benefit of being able to operate on the .frm equation file as a list of equations, or a list of groups of equations (for instance pre-model, main model, after-model, but also other kinds of groupings). Long-run levels of variables could also be treated as a group of equations. Possibilities like adding, dropping, replacing equations etc. would be very useful in relation to for instance educational uses. A well thought out syntax for this would be beneficial, instead of the current rather ad hoc syntax of using "###" to separate submodels, perhaps also recursive submodels (submodels inside submodels).

Thomas deemed items 7 and 8 important (examples and better error messages). These ought to be prioritized.

Item 9 (translator from 2.4 to 3.0): The existing AREMOS translator is relatively advanced, since it was used for the translation of the Quarterly National Accounts system, and the translator was refined iteratively in that process. In contrast, the translator from Gekko 2.4 to 3.0 is a bit less advanced, but unless the 2.4 programs use "advanced" features, the syntax differences between 2.4 and 3.0 are rather modest, and the databank logic and simulation parts are identical between 2.4 and 3.0. Rasmus B. did not consider an improved translator from 2.4 to 3.0 very important, since the translation is in any case a one-time event. Still, migrating the ADAM datarevision programs from Gekko 2.4 to 3.0 could probably benefit from some translator improvements (this system is quite large).

Item 10 (more advanced plot) has been discussed with the MAKRO team and others, and among other things it is a question of getting some inspiration from how plots are handled in, say, RStudio and similar programs.

Item 11 (seasonal adjustments): It would probably be a lot less ressource-intensive to export a dataframe of Gekko data to R, and use the RJDemetra package, rather than interfacing directly with JDemetra+ (which entails talking directly to Java). Tony asked about speed, and Thomas thought it would run fast if using arrow-files for interfacing. Interfacing with RJDemetra would be a Gekko 3.0 project, but Thomas mentioned that since Gekko 3.0 reads 2.4 databanks and vice versa, existing systems written in Gekko 2.4 can relatively easily run parts of these systems in 3.0, if needed. This could also aid stepwise migration from 2.4 to 3.0.

Regarding prioritization, Thomas felt that some work could be done regarding items 5-6 within the normal funding, while at the same time leaving some resources regarding better error messages, examples, guides, etc. Thomas mentioned that Gekko has gradually become a larger and larger software package, so the maintenance and bug-fixing demands more resources than in the humble beginnings. Asger was happy that Gekko is being used more, but the growth of the capabilities of the program also means that maintenance takes more time. Statistics Denmark and the Ministry of Finance have service agreements that, besides development, cover bug-fixes, helping out with questions, and Asger asked if others had the same kind of agreement? He advised that other institutions could perhaps enter into such agreements, perhaps on a limited scale, so that bugfixes, questions, etc. do not go fundamentally unfunded. He worried that without such agreements, either

Thomas would do such maintenance free of charge or not at all, or only in relation to larger development projects. Smaller service agreements could run idle some of the time. From the users' point of view, agreements like this would also ensure that annoying bugfixes etc. are done within reasonable time. Thomas added that he has historically been rather helpful when errors were reported, including minor enhancement suggestions, etc., because the software was under construction, and also because he felt kind of responsible for any bugs that he had exposed the users to...

Thomas said that within the agreements with Statistics Denmark, Ministry of Finance and the upcoming source code documentation project, he thought that it would be possible to do some model developments alongside better error messages and better documentation. Without further funding, model improvements could not include very fancy or advanced capabilities in the coming period, but getting started and reaping the low-hanging fruit should be possible.

Michael tried to sum up.

- Regarding prioritization, item (1) is already funded, (2)-(4) are mostly a question of wrapping existing code up, so therefore items (5)-(11) are more relevant to prioritize.
- First steps of (5) and (6) could be started up and investigated. More funding or more time is needed to implement more fancy modelling capabilities.
- (7) and (8) are necessary and will be done.
- (9) not that important.
- (10) and (11) kept in drawer.
- Asger advocated that users/institutions could pool their contributions if they need something to be built/refined.

Tony asked about the possibility of back-porting the model development parts made for The Economic Council (Y-, T-, and P-type equations and more) to version 2.4 (2.5.x to be precise), so that these users could benefit from it. Thomas said that as a general rule he is a bit reluctant to back-port new development code from later versions to earlier versions, both because it is time-consuming, and also because it tends to prolong "branching" of the software.

Tony said that this was mostly a question of the model definition file (.frm file, which is a text document), but Thomas felt that new development of the syntax of the model definition file ought to be inspired by GAMS, with implicit loops over sets, summing, $-conditionals, etc., because multi-dimensional economic models in Denmark tend to be written in that way (in GAMS). Tony primarily considered model/submodel enhancements though, which Thomas admitted was easier to back-port to version 2.4. Thomas supports old Gekko versions in the sense that he keeps them running (including critical bugfixes), but he also felt that he could not continue development projects on Gekko 2.4 indefinitely, both because he felt he had a formal obligation as Gekko editor to try to direct the project in one direction, but also because he feels that branching is unhelpful regarding maximization of the general interest of the product. Since this question is not relevant for all users, it was suggested to address it separately among the interested parties.

Re 6. Commentaries regarding detailed checklist + prioritization of this

1. Regarding special symbols, Thomas would not mind them at all, if they did not interfere with or complicate interfaces to other programs, for instance Excel and others. The problem with '¤' (currency symbol, like in '¤x') is that '¤' does not exist on for instance English keyboards, whereas symbols like '~' or '|' do not look sensible as prefixes (could be used as suffixes though, like 'x~s' or 'x|s'). The GAMS way of handling this would be defining a dimension to contain the special variants of the variable, corresponding to an array-series 'x[s]'.
2. Metadata. Asger mentioned that some "annual rings" (recorded history) regarding the source metadata field could be extremely practical. The idea would that one would be able to trace back the history of timeseries (and their observations), in essence storing all the current 'source' metadata fields in a longer list. This would be verbose, and could be turned on and off (and pruned) for different purposes. Inheriting 'source' metadata when copying data could also be useful. Thomas suggesting doing a draft of this, since it should not be too difficult to do. An option to skip the update of the 'source' metadata field could also be practical.
3. Startup environment. John mentioned that at The Economic Council it takes up to 30 seconds to start up Gekko from scratch. Thomas answered that this is no doubt because a larger library of procedures are loaded at start up. The solution to this would be "lazy loading", so that a given procedure/function is not really read and compiled until it is actually used. This would be relatively easy to do. In the longer run, a kind of packaging system akin to R/CRAN could be implemented, and the Gekko website could host a library of packages (procedures/functions).
4. The option tree has been mentioned many times, but it seems people are living with it.
5. Working environment. For instance, the MAKRO team use Sublime Text as a kind of Gekko main window, from where they can run their commands. It is not the intention for Gekko to produce an advanced "editor" to compete with existing text editors.
6. Description of table cell syntax (construction of tables, cell by cell) ought to be done.
7. Better solver? The Ministry of Finance says the existing solver actually runs pretty well, so this is perhaps not the most pressing regarding ADAM-type models. Probably model enhancements (model blocks, loops, etc.) is more important right now.
8. Interface to StatBank Denmark (Statistikbanken). There is actually an old proposal/blueprint regarding the improvement of this, but it lacks funding.
9. Dataframes. There is a new Microsoft project regarding this, and a prototype already exists for C#/.NET). Implementing dataframes (as objects alongside lists, matrices and maps) would probably be beneficial at some point (but probably not this period).

Michael asked about series definitions as functions, that is, defining a series as a function, perhaps as the solution to an equation, or even as the solution to a set of equations. Thomas mention the auto-updating series concept from EViews, and using such definitions has some reminiscences of Excel cells auto-updating via formulas. There would probably be some issues to consider, if more than one databank is open, and Gekko is allowed to read from any of those other databanks.

The Economic Council would invite Thomas to come around their offices (corona permitting), to see how Gekko is used, and perhaps help them optimize their setup.

Re 7. Other potential users

Thomas mentioned some interest from non-Danish international users, for instance University of Hawaii (who are still using AREMOS), but also others.

Besides, KL (Local Government Denmark) has begun to use Gekko for some of their systems.

There is apparently also a rather large AREMOS system in existence at DI (Confederation of Danish Industry).

Re 8. Status regarding documentation, help systems, etc.

The in-built help system (F1) regarding 2.4 and 3.0 is generally kept up to date. The documentation for 3.0 could warrant a read-through though, among other things to check that all examples still work out.

Beginner guides are more lacking. According to John, newcomers seem to learn Gekko relatively quickly, but The Economic Council would be interested in helping out regarding the document "The Sun Is Always Shining In Gekko" (a beginner's guide for Gekko). Perhaps upgrading "The Sun..." for Gekko 3.0 could be a co-production between Thomas and different institutions. Thomas could take the existing "The Sun..." for Gekko 2.4 and write new 3.0 compatible code into the document, illustrating 3.0-best-practices for the same exercises. After this, the document could be "iterated" between interested institutions who can add comments and suggestions.

Staticstics Denmark has run a Gekko-data course, and the materials from that course could also be of use.

Thomas mentioned that he felt he ought to write a kind of Gekko "book", where the software is perceived from more of a birds-eye perspective ("what is the whole point of the software"), analogous to introductory books on, say R, Python, or others. In this period, however, we will try to get more practical guides up and running.

About documentation and guides John mentioned that, sure, there are still AREMOS users in existence out there, but in the documentation it would probably be unwise to focus too much on AREMOS or how stuff is done in Gekko compared to AREMOS (or how the syntax differs). The heritage should be acknowledged, but otherwise the umbilical cord should be cut.

Re 9. Organization and choice of editor for the next period

TT was elected as editor for the next period. At the next meeting, this election perhaps ought to be performed in part 1 of the meeting.

Re 10. Date of the next meeting + any other business

June 2021 suited people well, preferably not after mid-June. Potentially the meeting would have to be virtual again.