

SAGE Model, v. 3.0 Review - Migration to WiNDC

On Science Advisory Board (SAB) advice, the Environmental Protection Agency (EPA) has built capacity in computable general equilibrium (CGE) modeling through the operationalization, maintenance and use of the SAGE model. SAGE is a forward-looking intertemporal model of the U.S. economy with subnational representation and multiple households per region. Each region also has 23 representative firms, with greater disaggregation in the manufacturing and energy sectors. Production technologies are represented with nested CES functions, which may include natural resource inputs. Capital for these firms is represented in a partial putty-clay framework to aid in capturing transition dynamics. A single government agent levies taxes on labor earnings, capital earnings, imports, commodities, and production. The model relies on the Armington framework governing both domestic and international trade. Version 1.2.2 of SAGE was reviewed by the SAB in 2020.¹ In response to its Tier 1 recommendations, versions 2.0 and above include improved calibration of government expenditures and deficits and the foreign trade deficit; allow for greater flexibility in the consumer demand system; and represent the United States as a large open economy.

The Science Advisory Board (SAB) also highlighted the importance of relying on open-source models for EPA analyses to increase both accessibility and transparency. In its 2020 review of SAGE, it recommended that the EPA move away from the proprietary IMPLAN database as the source for SAGE's social accounting matrix (SAM) to an open data source such as the Wisconsin National Data Consortium (WiNDC) database.² The EPA has now implemented this recommendation in SAGE version 3.0. Other updates to SAGE - new to version 3.0 - are described in the "What's New in Version 3.0.0?" section of the model documentation.

To ensure consistency with SAB guidance, we are initiating a targeted review of the migration from IMPLAN to WiNDC. We specifically ask reviewers to address the following charge questions:

- Is the migration of the model to WiNDC operationalized in a transparent and defensible way?
- Is the implementation of WiNDC data within SAGE clearly explained in the documentation?

Reviewers are also free to opine on other aspects of the SAGE model where they feel improvements can be made.

To facilitate the review, we are providing the reviewers with the SAGE model (verified to run in GAMS v. 50.4 or earlier), the build stream for producing the benchmark dataset used in the model (written using both R and GAMS), the model documentation, and a memo describing key similarities and differences between the IMPLAN and WiNDC datasets. Section 3.1 of the model documentation also contains a high-level description of the way in which WiNDC has been utilized to create the SAM in SAGE.

We request that reviewers organize their advice into three bins: (1) recommendations that should be addressed prior to using SAGE v. 3.0 in a regulatory analysis of an EPA rule; (2) recommendations that should be prioritized in the near-term (1 – 2 years) but do not prevent use of SAGE v. 3.0 in a regulatory analysis; and (3) longer-run recommendations (>2 years) for improvements to SAGE and areas where more research is needed.

¹ https://www.epa.gov/system/files/documents/2022-01/epa-sab-20-010_1.pdf

² <https://windc.wisc.edu/>

Overview of changes to WiNDC data

We make several modifications to the WiNDC data to conform to the structure and resolution of the SAGE model. These changes are summarized below with reference to the specific programs that implement the change in the build stream, if applicable. As described in the model documentation, to run the full SAGE build stream, we refer you to the R program, `"build/build_default_datasets.r"`.

- 1) The WiNDC data used as an input to the SAGE model build stream are based on the version that relies on the Current Population Survey (CPS). This is produced by running both the `"core"` and `"household"` modules of the WiNDC build stream assuming a static model (we align capital and investment accounts in a later step to satisfy conditions for a Ramsey-based growth model). However, we make one change in the household build to create the dataset. We modify the R script in the WiNDC build stream that downloads the CPS data through the Census API to allocate household observations to income bins (defined as income before taxes and transfers) based on pure income quintiles instead of predefined bins based on income thresholds. This change is not included in the distributed materials for review. We limit the size of the SAGE model package by only including the outputs of the WiNDC build stream in the `"build/data"` subdirectory. The version of the WiNDC build stream that we use is available upon request.
- 2) The default sectoral resolution in the WiNDC data is the summary sector level defined by the Bureau of Economic Analysis (BEA). While most of these sectors easily aggregate into the SAGE sectoring scheme, several must be disaggregated (e.g., the utilities sector is disaggregated into electricity transmission and distribution, natural gas distribution, and water and sewage). The process for disaggregating the sectors is broadly described in Section 3.1.1 of the model documentation. The program that implements the disaggregation of summary sectors is in `"build/disaggregate_summary_sectors.gms"`. The disaggregation is informed by national-level detailed BEA supply and use tables, Bureau of Labor Statistics data on employee compensation by NAICS code and region, and data on fuel use in electricity generation from the Energy Information Administration (EIA). The code operates by constructing national shares using the detailed BEA data, projects a calibrated set of control totals to the benchmark year (2022), and recalibrates the subnational data using national control totals, regional targets for the distribution of production, and targets for energy intensities used in electricity production.
- 3) The SAGE v3.0 build stream also contains tools for disaggregating the crude oil and natural gas extraction sector. The code is in `"build/disaggregate_oil_gas.gms"` and uses production data from EIA and trade data from Census to inform the disaggregation. This methodology is similar to what is used in prior SAGE model versions, with the exception of updates to the parameter namespace that corresponds with WiNDC-based parameters and the incorporation of margin demands into the balancing problem.
- 4) As in prior versions, the SAGE build stream contains a matrix balancing routine for recalibrating capital and investment accounts to satisfy steady state growth conditions in the intertemporal modeling framework. The version circulated with this package updates the parameter namespace to correspond with WiNDC-based parameters. The code is in `"build/balance_benchmark.gms"`.