**Assessing the accuracy of a program to perform household simulation**

Selecting a sample of 5,000 households without imposing any rejection threshold on the discrepancy between the sample and the population results in a sum of squared differences of around 0.0014 and takes around 52 seconds to run the simulation.



**age n p\_pop p\_sam diff**

1 563423 0.0792 0.103 0.024

2 588815 0.0828 0.0918 0.009

3 929086 0.131 0.115 -0.016

4 1134463 0.159 0.142 -0.018

5 1117056 0.157 0.154 -0.003

6 1234916 0.174 0.171 -0.002

7 854647 0.120 0.111 -0.009

8 401899 0.0565 0.0635 0.007

9 289831 0.0407 0.0483 0.008

Imposing a rejection threshold of 0 on the discrepancy between the sample and the population, in other words, each additional household must cause a reduction in the sum of squared differences, results in a simulation that never ends.

Imposing a rejection threshold of 0.0001 on the discrepancy between the sample and the population, results in a sum of squared differences of around 0.0013 and takes around 60 seconds to run the simulation. The rejection rate was 9.6%.



**age n p\_pop p\_sam diff**

1 563423 0.0792 0.0998 0.021

2 588815 0.0828 0.0895 0.007

3 929086 0.131 0.112 -0.018

4 1134463 0.159 0.145 -0.015

5 1117056 0.157 0.153 -0.004

6 1234916 0.174 0.168 -0.006

7 854647 0.120 0.116 -0.004

8 401899 0.0565 0.0613 0.005

9 289831 0.0407 0.0554 0.015

In another simulation of 10,000 households, the resulting sum of squared differences was 0.014 and took 2.2 minutes to run. The rejection rate was 10.3%.



In order to improve the accuracy of the simulation, a dynamic threshold was applied. An initial threshold of 0.001 was provided. At each iteration, the threshold was multiplied by a factor of *f*/*k* where *k* is the iteration number and *f* is initially given a value of 1. The value of *f* increases by 1 if, at each iteration, the household is accepted and decreases by 1 (but never falling below 1) if the household is rejected, thus keeping the rejection rate constant at around 50%. This results in a sum of squared differences of 3.2×10-8 and takes 3.7 minutes to run.



**age n p\_pop p\_sam diff**

1 563423 0.0792 0.0792 0

2 588815 0.0828 0.0828 0

3 929086 0.131 0.131 0

4 1134463 0.159 0.159 0

5 1117056 0.157 0.157 0

6 1234916 0.174 0.174 0

7 854647 0.120 0.120 0

8 401899 0.0565 0.0564 0

9 289831 0.0407 0.0408 0