

This Old Model

Updating and Improving Older Tax Models

Property Tax and Rent Rebate (PTRR) model

- The PA PTRR program provides rent or property tax rebates to senior citizens
- The model has two purposes:
 - Report statistics on current usage, using DOR PTRR data.
 - Estimate revenue impact of increasing income limits, using PUMS data

Previous Model

- The original model has been in use since 2001.
- Model consisted of DB2 tables pulled from the mainframe using QMF.
- There were about 91 DB2 tables on the mainframe – Research needed to link 10 of these tables.
- Using QMF these were joined as one table on the mainframe – this table was then moved into Access.

Previous Model - continued

- This process alone could take up to 10 hours.
- The joined table had approximately 500,000 returns and 15 columns.
- Writing the Access code was not overly difficult as there was expertise in this area. However with a limited number of Access users, it was difficult to easily have other analysts work on this project.

Previous Model - continued

- Running a simple request (claimants by city) could take several hours.
- The time involved to create databases and run simulations did not meet the needs of our customers.
- As we became aware that Property Tax Relief was going to be an important topic, it was clear that a faster more efficient model had to be developed.

Updated PTRR Model

- The first change was to move from the ACCESS software to SAS as the programming tool. SAS was better able to join different tables and to more quickly run through large tables.
- Analysis was done on which variables from the PTRR data were most likely to be used in analysis. SAS code joined all the tables and created one SAS database to be used for analysis.

Updated PTRR Model – cont.

- Several SAS programs were developed that could handle most analysis and simulations using both the PTRR data and the PUMS data.
- Programs that used to take several hours to run would now take minutes.
- Revisions to code were done more easily – instead of waiting 4 hours to find out that the code did not give the desired results, problems could be fixed almost immediately.

Be Careful What You Wish For!

- Once our customers realized how quickly simulations could be run, the requests for estimates seemed to increase exponentially.
- The quicker response time also led to more creative ideas on increased property tax relief. The PTRR data and PUMS data still had limitations as to how much could be explained.
- In addition, more complicated proposals still required time to write code and verify results. There was no “magic button” to push and get instant results.

Summary

- Overall, the newer model allowed Research to respond more quickly to requests.
- The PTRR program's income limits were increased and property tax relief was increased for older and disabled Pennsylvanians. The model directly helped with that conclusion.

PIT Tax Forgiveness (SP) Model

- To simulate tax forgiveness programs, Research used a PIT tax model, built by an outside consultant .
- The model used a sample of PIT returns and federal returns, which were averaged and blurred to limit disclosure of specific taxpayer information, certain fields (income, number of dependents) were capped at a certain level.

PIT SP Model - continued

- While the income limits were understood, the ramifications of limiting the number of dependents to three per return were not apparent at first.
- As the administration in the late 1990's continued to expand PA's tax forgiveness program, the model was used exclusively to estimate the fiscal impact. Most of the increases in tax forgiveness expanded the amount of income forgiveness for dependents.

PIT SP Model - continued

- Research came to realize that when a tax forgiveness simulation was run that expanded income levels for dependents, the outside model consistently understated the fiscal impact.
- The model was not able to be easily updated with newer data that would not limit the number of dependents.
- It was apparent that a new model would have to be built.

Issues with PIT SP Model

- PIT databases are large – over 5.5 million records.
- PIT database does not contain all the variables (certain income items, dependents) needed to calculate SP.
- Model must be able to run quickly and precisely in order to meet the legislative requests.

PIT SP Model

- Both federal data and state data were brought into SAS.
- A merge was done between federal and state data. It was particular difficult to correctly merge returns that selected married filing separate for the IRS and not the state (or vice-versa).
- The file created has 6.5 million rows and 410 columns. Uncompressed, it is 4.5 GB.

PIT SP Model – continued

- Before even writing SAS code to simulate SP, storage issues arose due to the size of the file.
- A short-term solution was that PCs were equipped with a second hard drive, containing about 40 GB. Compressed files were copied onto each hard drive and SAS was run against that hard drive.

PIT SP Model – continued

- A better solution was the issuance of 150 GB of space on the Department's Sybase server. Research has created hundreds of SAS datafiles for analysis.
- Analysts are all now using the same data and not creating mini datamarts on their PCs.
- The downside is that there is only about 22 GB of space left on the server.

PIT SP Model – continued

- Now that the database was created, code needed developed to simulate SP analysis.
- The code was written so that limits and parameters could be easily changed.
- A nice feature is that if results are counter-intuitive, the micro-data can be examined to verify the results. The old model was like a “black box” that gave results, but it was sometimes difficult to analyze the validity of the results.
- The code can be used against the most recent year of PIT/federal data.

Summary – PIT SP Model

- It took time to develop code to merge the federal and state data correctly and to deal with the size of the file.
- The resulting database is used for analysis, especially proposals that involve data not typically found in the PA PIT database.
- Though the SAS code is long and layered to run SP analysis, the advantages of being able to easily update the data and verify results is invaluable.