Banff and Simputation: A Comparison Using BEA's Multinational Enterprise Surveys

Larkin Terrie

FCSM Research and Policy Conference

October 25-27, 2022

Bureau of Economic Analysis
U.S. DEPARTMENT OF COMMERCE

Outline

- Overview of auto-editing at BEA
- Overview of auto-editing software options
- Approach to comparing accuracy of imputations
- Explanation of results

Auto-Editing at BEA

- Focused on annual direct investment surveys, which collect financial and operating data from:
  - U.S. multinational enterprises and their foreign affiliates
  - Foreign-owned U.S. companies
- Motivation: allow survey staff to spend more time on most complex/impactful responses, improve general efficiency of survey editing

Auto-Editing Software Options

- Banff System for Data Editing and Imputation
- R auto-editing packages (where Simputation is package for imputing missing/erroneous data)
- Both designed for production of official statistics
- Key difference: imputation methods

Differences in Imputation Methods

Key Imputation Methods in Banff and Simputation

Type	Banff	Simputation
Donor	Donor imputation with matching based on editing rules	Sequential hot deck (shd)
		Random hot deck (rhd)
		K-nearest neighbor (knn)
		Predictive mean matching (pmm)
Regression	OLS	OLS (lm)
		Robust linear regression through M-estimation (rlm)
		Elastic net/lasso/ridge regression (en)
Decision Tree	None	Classification and regression tree (cart)
		Random forest (rf)
		Multivariate imputation based on iterative random forest estimates (mf)

Approach to Analysis

- Part 1: Which Simputation methods should be used for each BEA survey form?
  - Find best donor method and best model-based method for each survey.

- Part 2: How accurate are Simputation's imputations compared to Banff's?
  - For each survey, compare Simputation-based auto-editing system developed in part 1 to Banff-based system previously developed.

Measuring Imputation Accuracy

- Problem: True values of imputed items not known

- Solution: Simulate missing/erroneous data on “clean” forms, then compare imputations to reported values

Accuracy Metrics

- Percent total error:

$$\frac{\sum_{k=1}^{m}\sum_{j=1}^{n} s_{ijk} - o_{ij}}{\sum_{k=1}^{m}\sum_{j=1}^{n} o_{ij}} \times 100$$

- Percent total absolute error:

$$\frac{\sum_{k=1}^{m}\sum_{j=1}^{n} \left| s_{ijk} - o_{ij} \right|}{\sum_{k=1}^{m}\sum_{j=1}^{n} \left| o_{ij} \right|} \times 100$$

Where s_{ijk} is the imputed value for field *i* in record *j* = 1, …, n in simulation run *k* = 1, …, *m* and o_{ij} is the reported value for the field and record in question

Data Analyzed

- Separate analyses of two BEA annual survey forms
  - BE-10D (short and simple form)
  - BE-15B (longer and more complex form)

Analysis Stage	BE-10D Survey Year Analyzed	BE-15B Survey Year Analyzed
1. Selection of Simputation procedures	2019	2015
2. Comparison of complete Simputation-based auto-editing system to Banff-based system	2014	2014

Comparing Simputation Methods

Percent of Simulated FTIs Imputed

Type	Method	BE-10D	BE-15B
Donor	K-nearest neighbor (knn)	99.08	94.06
	Predictive mean matching (pmm)	47.94	8.16
	Sequential hot deck (shd)	98.62	77.19
	Random hot deck (rhd)	97.77	77.19
Model -Based	OLS (lm)	72.87	65.21
	Robust linear models (rlm)	76.44	80.05
	Elastic net (en)	69.38	55.16
	Classification and regression tree (cart)	100.00	83.21
	Random forests (rf)	52.68	55.38
	Iterative random forests (mf)	97.01	35.94

- 100 simulation runs per method per form
- Methods vary not only in accuracy but in proportion of FTIs (fields to impute) imputed

Accuracy of Simputation Methods, 10D

Selected Pairwise Comparisons of Simputation Methods

Field	Donor				Model-Based			
	Pct. Abs. Error		Pct. Error		Pct. Abs. Error		Pct. Error	
	knn	pmm	knn	pmm	mf	rlm	mf	rlm
Assets	69.01	62.14***	-10.32	0.50***	54.77***	56.40	1.44***	-21.05
Debt Payable	116.31***	131.15	-44.82	5.51***	134.62	100.00***	7.65***	-99.97
Debt Rec.	144.05***	159.15	-28.87	6.48**	165.10	100.00***	12.15***	-100.00
Employment	115.72	101.81***	-7.14	4.66^{+}	98.99	87.22***	5.13***	-62.92
Liabilities	75.54	66.19***	-14.80	-0.36***	63.32***	69.47	2.08***	-24.03
Net Income	125.40	110.95***	22.87	3.56^{+}	98.08	94.43***	-6.61*	61.71
Sales	76.53	70.45***	-14.82	1.12***	59.03***	79.29	-1.86***	-31.11

Key to pairwise *t* test results:

*** difference significant at α = 0.0001

** significant at α = 0.001

* significant at α = 0.01

+ significant at α = 0.05

Accuracy of Simputation Methods, 15B

Selected Pairwise Comparisons of Simputation Methods

Field	Donor				Model-Based			
	Pct. Abs. Error		Pct. Error		Pct. Abs. Error		Pct. Error	
	knn	pmm	knn	pmm	rlm	lm	rlm	lm
Assets	--	--	--	--	23.41***	28.38	-8.19	-0.91
Capital Gains	198.69	343.70	199.52	203.94	105.19***	623.94	-170.38***	108,959.1
Employment	24.84	19.13^{+}	-4.52	-5.76	20.74***	22.29	-1.33**	6.02
Emp. Comp.	27.17	31.02	-8.09	-0.43	19.10***	24.07	-1.06*	4.74
Mfg. Emp.	44.88***	55.99	-2.51***	25.66	88.09	83.14***	47.15	24.91***
PP&E Exp.	59.20	51.96^{+}	-13.83	-0.15^{+}	94.32***	138.57	19.09***	87.04
Exports	62.67	40.42**	-11.34	12.79	52.34***	82.20	8.66***	54.02
Gross PP&E	50.11	21.89***	-17.80	1.18***	25.72***	37.59	-4.09***	15.92
Imports	60.81	65.04	-16.81	26.83	54.85***	109.38	23.88***	87.62
Interest Paid	67.69	34.07^{+}	-10.96	-8.85^{+}	62.64***	240.88	13.52***	201.51
Interest Rec.	64.00**	80.54	-4.40**	37.72	46.23***	380.78	-7.48***	344.92
Liabilities	38.22	1.97***	-11.68	0.28***	36.10	35.87	17.99	15.26***
Net Income	--	--	--	--	96.76***	102.60	-84.50	-39.52***
Owners' Eqty	--	--	--	--	43.09	43.33	-36.85	-25.44^{+}
R&D	82.45	66.16^{+}	-20.50	4.40	25.96***	97.26	-2.19***	82.19
Sales	27.12	18.55**	-9.16	5.97*	25.57***	27.45	5.09**	7.65
U.S. Inc. Tax	91.57*	111.10	-0.57	-4.26	98.39***	237.58	58.74***	321.40

Simputation-Based Auto-Editing Systems

Imputation Methods Selected for Simputation-Based Auto-Editing Systems

Form	Donor Method	Model-Based Method
BE-10D	Predictive mean matching (pmm)	Iterative random forests (mf)
BE-15B	*K*-nearest neighbor (knn)	Robust linear regression (rlm)

Banff vs. Simputation, 10D

Pairwise Comparison of Banff and Simputation-Based Auto-Editing Systems

Field	Pct. Abs. Error		Pct. Error	
	Simputation	Banff	Simputation	Banff
Assets	68.19***	117.76	3.62***	4.22
Debt Payable	112.65***	119.95	-0.04***	-13.41
Debt Receivable	164.05	164.52	11.28	-6.91
Employment	99.30***	103.94	5.46*	-9.42
Liabilities	66.49***	118.42	2.12***	5.66
Net Income	101.58***	123.86	21.69	22.40
Sales	61.85***	81.86	1.74***	-12.28

Banff vs. Simputation, 15B

Pairwise Comparison of Banff and Simputation-Based Auto-Editing Systems

Field	Pct. Abs. Error		Pct. Error	
	Simputation	Banff	Simputation	Banff
Assets	31.00	14.26***	-16.81	-6.67***
Capital Gains	119.35***	292.93	-119.38	15.88
Employment	93.36	38.78	67.42	-3.52^{+}
Emp. Comp.	24.78***	39.70	-7.16***	-17.11
Mfg. Emp.	52.87	42.30***	-17.52	-19.47
PP&E Exp.	79.72	74.76	-7.17	-13.79
Exports	54.10**	58.28	-12.63	-1.05**
Gross PP&E	44.50	36.30	25.96	-16.85
Imports	71.27***	89.34	13.14***	47.90
Interest Paid	61.13***	83.88	13.65***	36.37
Interest Rec.	56.08	50.07*	2.03	4.57
Liabilities	24.58	17.19	4.00	8.03
Net Income	94.98	81.12^{+}	-87.20	-50.59**
Owners’ Equity	38.14	35.86	-7.18*	-33.28
R&D	36.60	26.90^{+}	-23.09	5.97**
Sales	28.69***	40.28	7.41	-6.56
U.S. Inc. Tax	100.95	100.13	57.00	-27.47***

Summary of Results

Comparison of Banff and Simputation-Based Auto-Editing Systems

Form	Simputation More Accurate	Banff More Accurate	Results Ambiguous
BE-10D	Assets, Debt Payable, Employment, Liabilities, Net Income, Sales		Debt Receivable
BE-15B	Capital Gains, Employee Compensation, Imports, Interest Paid, Owners' Equity, Sales	Assets, Employment, Manufacturing Employment, Interest Received, Net Income, R&D, U.S. Income Tax	PP&E Expenditure, Exports, Gross PP&E, Liabilities

Conclusions

- Neither software option is universally superior
- Both are viable options for BEA auto-editing systems
- Banff may have an advantage with more complex survey forms

Contact Information

- Questions on the presentation?
  - Larkin Terrie: Larkin.Terrie@bea.gov

- Questions on BEA's direct investment statistics?
  - Internationalaccounts@bea.gov

Thank You!