

Automated Data Editing and Imputation for Surveys of Multinational Enterprises, a Banff Implementation



Mark Xu, Andy K. Kim, and Larkin Terrie
CONFERENCE OF EUROPEAN STATISTICIANS, UNECE
Work Session on Statistical Data Editing
(The Hague, Netherlands, 24–26 April 2017)

- Introduction
- The advantages of automated editing
- Banff's constraints and BEA's solutions
- Approach to implementing Banff
- Testing BEA's implementation of Banff
- Discussion and conclusion

- BEA surveys of multinational enterprises
 - For U.S. direct investment abroad
 - For foreign direct investment in the United States
- BEA's development of an auto-editing system
 - Selection of Banff
 - Two tests

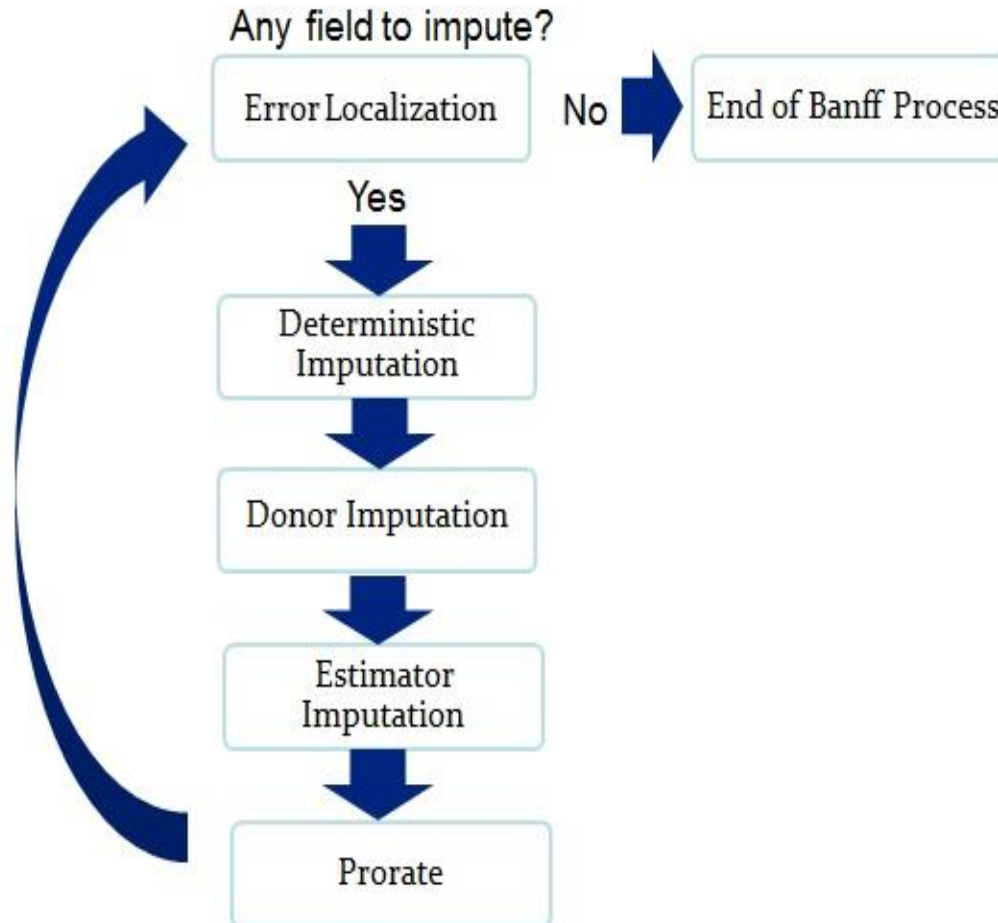
- Pros and cons of traditional data editing
- Pros and cons of auto-editing
- Mixed editing is the answer

Banff's Constraints and BEA's Solutions



Banff's constraints	BEA's solutions
Linear edits only	Convert non-linear edits to linear edits
"Hard" edits only	Convert "soft" edits to "hard" edits
Numeric data only	Pre-editing

Approach to implementing Banff



Percent of Fields Edited Manually vs. Auto-Edited

2011 BE-15 A forms: U.S. affiliates of foreign MNE

		Manual Editing		
		Edited	Not Edited	Total
Auto-Editing	Edited	4.1%	2.0%	6.1%
	Not Edited	8.7%	85.3%	93.9%
	Total	12.8%	87.2%	100%

Comparison of Auto-Edit and Manual Edit by Data Item, 2011 BE-15A



Variables	% Difference Auto vs. Manually Edited	% Share of Total Fields Auto- Edited	% Share of Total Fields Manually Edited
Assets	0.67	2.52	12.91
Inventories	0.74	1.19	4.04
Equity Investment	-0.47	7.95	14.24
Net PPE	-3.36	1.66	7.02
Other Assets	1.06	8.01	24.37
Liabilities	1.07	12.19	19.80
Owners' Equity	-1.73	1.79	14.44
Capital Stock	0.47	2.85	12.45
Retained Earnings	-58.02	2.32	18.34
Treasury Stock	25.73	0.33	0.60
Acc. Comp. Income or Loss	-35.54	0.86	13.58
Trans. Adj.	107.26	0.20	4.90
Other Acc. Comp. Income or Loss	-11.43	2.05	13.05
Other Owners' Equity	-2.51	0.66	11.66

Frequency of Different Imputation Methods, 2011 BE-15A



Imputation Method	Status Code	Description	Imputed Fields %
Deterministic	IDE	Deterministic (1 st Round)	92.2
	IDE2	Deterministic (Subsequent Rounds)	1.9
Donor	IDN	Donor Imputation	0.4
Estimator	ILR1	Historical Regression Model	1.9
	IDIF	Residual Difference Between Two Variables	0.2
	IHRT	Historical Subcomponent Share	0.3
	ICA	Carry Forward	0.4
Prorate	IPR	Prorated (Previously Imputed Fields)	2.7
All			100.0

Jackknife Estimates, 2011 BE-15A

Variables	Range of Percent Difference between Auto-Edit and Manual Edit	
Assets	-0.63%	1.80%
Inventories	-0.55%	2.03%
Equity Investment	-9.95%	8.52%
Net PPE	-8.47%	1.75%
Other Assets	-0.37%	2.38%
Liabilities	-0.16%	2.21%
Owners' Equity	-6.39%	2.95%
Capital Stock	-3.67%	1.53%
Retained Earnings	-143.58%	29.68%
Treasury Stock	-21.93%	59.81%
Accum. Comp. Inc. Loss	-89.08%	0.75%
Translation Adj.	-75.54%	289.92%
Other Acc. Comp. Inc. Loss	-43.07%	5.80%
Other Owners' Equity	-45.04%	35.50%

Discussion and Conclusion

- Substantial effort involved in setting up the system, especially in working around Banff's constraints
- The system is capable of efficiently and accurately editing data
- More work is still needed to improve the system

Questions



Percent of Fields Edited Manually vs. Auto-Edited

2009 BE-10 short A forms: U.S. parents of U.S. MNE

		Manual Editing		
		Edited	Not Edited	Total
Auto-Editing	Edited	0.6%	0.5%	1.0%
	Not Edited	2.9%	96.0%	99.0%
	Total	3.5%	96.5%	100%

Comparison of Auto-Edit and Manual Edit by Data Item, 2009 BE-10 Short A



Variables	% Difference Auto vs. Manually Edited	% Share of Total Fields Auto- Edited	% Share of Total Fields Manually Edited
Sales	0.62	3.57	8.93
Employment	8.87	5.44	17.52
Net Income	149.34	0.34	6.97
Assets	1.48	0.26	14.03
Liabilities	0.66	0.60	6.55
Exports	-3.79	6.29	5.44
Imports	-0.44	3.23	3.57

Frequency of Different Imputation Methods, 2009 BE-10 Short A

Imputation Method	Status Code	Description	Imputed Fields %
Deterministic	IDE	Deterministic Imputation	46.8
Donor	IDN	Donor Imputation	10.7
Estimator	ICOM	Imputation by Regression	15.7
	ISUM	Imputation by Sum	7.7
	IPRO	Imputation by Proportion	4.7
Prorate	IPR	Prorated	10.2
	EDT	Editors' recommendation	4.2
All			100.0

Jackknife Estimates, 2009 BE-10 Short A

Variables	Range of Percent Difference between Auto-Edit and Manual Edit	
Sales	-0.20%	1.15%
Employment	2.94%	15.18%
Net Income	-33.01%	157.33%
Assets	0.87%	3.46%
Liabilities	-2.30%	0.92%
Exports	-7.79%	-1.30%
Imports	-8.22%	3.46%