

# Contents

- Foreword** ..... vii
- Acknowledgements** ..... ix
- Acronyms** ..... xv
- List of Figures** ..... xvii
- List of Tables** ..... xix
- 1 Introduction** ..... 1
- 2 Background on Multiply Imputed Synthetic Datasets** ..... 7
  - 2.1 The history of multiply imputed synthetic datasets ..... 7
  - 2.2 Advantages of multiply imputed synthetic datasets compared with other SDC methods ..... 10
- 3 Background on Multiple Imputation** ..... 13
  - 3.1 Two general approaches to generate multiple imputations ..... 14
    - 3.1.1 Joint modeling ..... 14
    - 3.1.2 Fully conditional specification (FCS)..... 15
    - 3.1.3 Pros and cons of joint modeling and FCS ..... 18
  - 3.2 Real data problems and possible ways to handle them ..... 18
    - 3.2.1 Imputation of semi-continuous variables ..... 19
    - 3.2.2 Bracketed imputation ..... 19
    - 3.2.3 Imputation under linear constraints ..... 20
    - 3.2.4 Skip patterns ..... 20
- 4 The IAB Establishment Panel** ..... 23

<b>5</b>	<b>Multiple Imputation for Nonresponse</b>	27
5.1	Inference for datasets multiply imputed to address nonresponse	27
5.1.1	Univariate estimands	27
5.1.2	Multivariate estimands	28
5.2	Analytical validity for datasets multiply imputed to address nonresponse	30
5.3	Multiple imputation of the missing values in the IAB Establishment Panel	31
5.3.1	The imputation task	31
5.3.2	Imputation models	32
5.3.3	Evaluating the quality of the imputations	33
<b>6</b>	<b>Fully Synthetic Datasets</b>	39
6.1	Inference for fully synthetic datasets	40
6.1.1	Univariate estimands	40
6.1.2	Multivariate estimands	40
6.2	Analytical validity for fully synthetic datasets	41
6.3	Disclosure risk for fully synthetic datasets	42
6.4	Application of the fully synthetic approach to the IAB Establishment Panel	44
6.4.1	The imputation procedure	46
6.4.2	Measuring the analytical validity	47
6.4.3	Assessing the disclosure risk	48
<b>7</b>	<b>Partially Synthetic Datasets</b>	53
7.1	Inference for partially synthetic datasets	53
7.1.1	Univariate estimands	54
7.1.2	Multivariate estimands	55
7.2	Analytical validity for partially synthetic datasets	56
7.3	Disclosure risk for partially synthetic datasets	56
7.3.1	Ignoring the uncertainty from sampling	57
7.3.2	Accounting for the uncertainty from sampling	58
7.4	Application of the partially synthetic approach to the IAB Establishment Panel	59
7.4.1	Measuring the analytical validity	60
7.4.2	Assessing the disclosure risk	61
7.5	Pros and cons of fully and partially synthetic datasets	62
<b>8</b>	<b>Multiple Imputation for Nonresponse and Statistical Disclosure Control</b>	65
8.1	Inference for partially synthetic datasets when the original data are subject to nonresponse	65
8.1.1	Univariate estimands	66
8.1.2	Multivariate estimands	67
8.2	Analytical validity and disclosure risk	68

8.3	Generating synthetic datasets from the multiply imputed IAB Establishment Panel	68
8.3.1	Selecting the variables to be synthesized	68
8.3.2	The synthesis task	70
8.3.3	Measuring the analytical validity	71
8.3.4	Caveats in the use of synthetic datasets	76
8.3.5	Assessing the disclosure risk	78
<b>9</b>	<b>A Two-Stage Imputation Procedure to Balance the Risk–Utility Trade-Off</b>	<b>87</b>
9.1	Inference for synthetic datasets generated in two stages	88
9.1.1	Fully synthetic data	88
9.1.2	Partially synthetic data	90
9.2	Analytical validity and disclosure risk	90
9.3	Application of the two-stage approach to the IAB Establishment Panel	90
9.3.1	Analytical validity for the panel from one-stage synthesis	91
9.3.2	Disclosure risk for the panel from one-stage synthesis	93
9.3.3	Results for the two-stage imputation approach	96
<b>10</b>	<b>Chances and Obstacles for Multiply Imputed Synthetic Datasets</b>	<b>99</b>
<b>A</b>	<b>Bill Winkler’s Microdata Confidentiality References</b>	<b>103</b>
<b>B</b>	<b>Binned Residual Plots to Evaluate the Imputations for the Categorical Variables</b>	<b>119</b>
<b>C</b>	<b>Simulation Study for the Variance-inflated Imputation Model</b>	<b>127</b>
	<b>Bibliography</b>	<b>131</b>
	<b>References</b>	<b>131</b>
	<b>Index</b>	<b>137</b>

Synthetic Datasets for Statistical Disclosure Control  
Theory and Implementation

Drechsler, J.

2011, XX, 138 p. 19 illus., Softcover

ISBN: 978-1-4614-0325-8