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
### 9 A Two-Stage Imputation Procedure to Balance the Risk–Utility Trade-Off

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

### 10 Chances and Obstacles for Multiply Imputed Synthetic Datasets

A Bill Winkler’s Microdata Confidentiality References 103

B Binned Residual Plots to Evaluate the Imputations for the Categorical Variables 119

C Simulation Study for the Variance-inflated Imputation Model 127

Bibliography 131

References 131

Index 137
Synthetic Datasets for Statistical Disclosure Control
Theory and Implementation
Drechsler, J.
2011, XX, 138 p. 19 illus., Softcover
ISBN: 978-1-4614-0325-8