2010

A Small Area Procedure for Estimating Population Counts

Emily J. Berg

Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/etd

Part of the Statistics and Probability Commons

Recommended Citation

https://lib.dr.iastate.edu/etd/11215

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
A small area procedure for estimating population counts

by

Emily Berg

A dissertation submitted to the graduate faculty
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Statistics

Program of Study Committee:
Sarah Nusser, Major Professor
Wayne A. Fuller
Mark Kaiser
Jae Kwang Kim
Dan Nettleton
Dan Nordman

Iowa State University
Ames, Iowa
2010
ACKNOWLEDGEMENTS

I would like to acknowledge and thank my major professor, Professor Fuller.
Abstract

Many large scale surveys are designed to achieve acceptable reliability for large domains. Direct estimators for more detailed levels of aggregation are often judged to be unreliable due to small sample sizes. Estimation for small domains, often defined by geographic and demographic characteristics, is known as small area estimation. A common approach to small area estimation is to derive predictors under a specified mixed model for the direct estimators. A procedure of this type is developed for small areas defined by the cells of a two-way table.

Construction of small domain estimators using the Canadian Labour Force Survey (LFS) motivates the proposed model and estimation procedures. The LFS is designed to produce estimates of employment characteristics for certain pre-specified geographic and demographic domains. Direct estimators for specific occupations in small provinces are not published due to large estimated coefficients of variation. A preliminary study conducted in cooperation with Statistics Canada investigated estimation procedures for small areas defined by the cross-classification of occupations and provinces using data from a previous Census as auxiliary information. For consistency with published estimates, predictors are desired that preserve the direct estimators of the margins of the two-way table.

One method in the Statistics Canada study is based on a nonlinear mixed model for the direct estimators of the proportions. An initial predictor is defined to be a convex combination of the direct estimator and an estimator obtained by raking the Census totals to the direct estimators of the marginal totals. The estimators resulting from the raking operation are called the SPREE estimators and are expected to have smaller variances than the direct estimators. The weight assigned to the direct estimator depends on the relative magnitudes of an estimator of a random model component and an estimator of the sampling variance. The final predictors are defined by raking the initial predictors to the direct estimators of the marginal totals. Estimation of the mean squared error (MSE) of the predictors was not fully developed.

This dissertation addresses several issues raised by the procedure discussed above. First, the method above uses SPREE to estimate a fixed expected value. SPREE is unbiased if the Census interactions persist unchanged through time and is efficient if the direct estimators of
the cell totals are realizations of independent Poisson random variables. A generalization of SPREE that is more efficient under a specified covariance structure is explored. A simulation study shows that predictors constructed under the specified covariance structure can have smaller MSE’s than predictors calculated with the direct estimators of the variances. An estimator of the MSE of the initial convex combination of the direct estimator and the estimator of the fixed expected value is derived using Taylor linearizations. The LFS procedure uses a final raking operation to benchmark the predictors. A bootstrap procedure is investigated as a way to account for the effects of raking on the MSE’s of the predictors. The procedures are applied to the Canadian Labour Force Survey, but the issues discussed are of general interest because they arise in many small area applications.
TABLE OF CONTENTS

ACKNOWLEDGEMENTS ......................................................... ii
LIST OF TABLES ................................................................. ix
LIST OF FIGURES ................................................................. xix

CHAPTER 1. Introduction ....................................................... 1
  1.1 Occupations in the Canadian Labour Force Survey .................. 2
  1.2 Challenges ................................................................. 3
  1.3 Outline ................................................................. 4

CHAPTER 2. Literature on Small Area Estimation of Counts and Proportions ....................................................... 5
  2.1 SPREE and Generalizations ........................................... 5
  2.2 Generalized Linear Mixed Models ................................... 7
  2.3 Exponential Quadratic Variance Function Models .................. 8
  2.4 Semi-parametric Models ............................................. 8
  2.5 Extensions and Reviews ............................................. 9
  2.6 Benchmarking ......................................................... 10
  2.7 MSE Estimation ..................................................... 10
  2.8 Estimation of Unknown Sampling Variances ....................... 13

CHAPTER 3. A Model for Vectors of Small Area Proportions ............. 18
  3.1 Fixed Expected Value ................................................. 19
  3.2 Small Area Effects .................................................. 20
  3.3 Sampling Errors ................................................... 21
3.4 Minimum Mean Squared Error Linear Predictors .................................... 23
3.5 Totals ........................................................................................................ 24
3.6 Multiple Two-way Tables ........................................................................ 26
3.7 Summary ..................................................................................................... 26

CHAPTER 4. Procedure ...................................................................................... 28
4.1 Estimators of Model Parameters ................................................................. 28
   4.1.1 Estimator of $c_k$ and Sampling Variances ............................................ 30
   4.1.2 Estimator of $\lambda_o$ .......................................................................... 33
   4.1.3 Estimator of $\psi$ .............................................................................. 34
4.2 Predictors of True Proportions and Totals .................................................. 40
   4.2.1 Initial Predictors of Proportions ......................................................... 40
   4.2.2 Beale Predictors of Proportions ......................................................... 41
   4.2.3 Benchmarked Predictors .................................................................... 44

CHAPTER 5. MSE Estimators ............................................................................. 50
5.1 MSE Estimators based on Taylor Linearization ......................................... 50
   5.1.1 MSE Estimators for Proportions ......................................................... 50
   5.1.2 MSE Estimators for Totals ................................................................. 56
   5.1.3 Estimating the MSE After Raking ....................................................... 58
5.2 Bootstrap Estimator of MSE ...................................................................... 59
   5.2.1 Bootstrap Distributions ....................................................................... 59
   5.2.2 Definitions of Bootstrap MSE Estimators ........................................... 62
   5.2.3 Computation of Bootstrap MSE Estimators ......................................... 64
   5.2.4 Moments of Bootstrap Distributions of Proportions and Totals ........ 65
   5.2.5 Bootstrap Distributions of Direct Estimators of Sampling Variances ... 67
   5.2.6 Details on Implementation ................................................................... 69
   5.2.7 Simplifications to Bootstrap MSE Estimator when Working Model for
       Sampling Variances Holds ....................................................................... 72
5.3 Benefits and Drawbacks of Taylor and Bootstrap MSE Estimators .......... 73
CHAPTER 6. Simulation .................................................. 76

6.1 Simulation Models .................................................. 76
  6.1.1 Parameters for the Simulation ................................. 81

6.2 Details of Computing and Estimation ............................. 85

6.3 Efficiencies of Predictors ........................................... 88
  6.3.1 Effect of Beale Ratio Estimator of $\gamma_{ik}$ on MSE of Predictor . 94
  6.3.2 Effect of Raking on MSE of Predictor ....................... 96

6.4 Comparison of Predictors Calculated with Direct Estimators of Sampling Variances to Predictors Calculated with Model Estimators of Sampling Variances . 98

6.5 Comparison of Bootstrap and Taylor MSE Estimators .......... 108
  6.5.1 SRS, 0.003 ...................................................... 112
  6.5.2 SRS, $\psi = 0.02$ ........................................... 118
  6.5.3 Two Stage, 0.003 ........................................... 124
  6.5.4 Two Stage, 0.02 ............................................ 129
  6.5.5 Estimators of Leading Terms ................................ 133
  6.5.6 Effect of Estimating Parameters ............................. 138
  6.5.7 Raking Effect ................................................ 140
  6.5.8 Summary .................................................... 144

6.6 Empirical Properties of Normal Theory Prediction Intervals .......... 147
  6.6.1 Empirical Coverages of Normal Theory 95% Confidence Intervals .... 147
  6.6.2 Interval Widths ............................................. 151
  6.6.3 Summary .................................................... 153

6.7 Augmented Model Predictors ....................................... 154

CHAPTER 7. Canadian Labour Force Survey ............................... 158

7.1 National Occupational Classification ............................ 158

7.2 Census and LFS Data ............................................... 159

7.3 Estimation and Prediction for Canadian LFS ..................... 161

7.4 Sampling Variances ................................................ 165
7.5 Estimates and Predictions for A1 ........................................ 185
  7.5.1 Model Estimates for A1 ............................................. 188
  7.5.2 Predictions for A1 .................................................. 191
  7.5.3 Comparison of MSE’s and CV’s of Predictors to MSE’s and CV’s of Direct Estimators for A1 ........................................ 196
  7.5.4 Model Assessment for A1 ........................................... 200
7.6 Estimates and Predictions for E0 ...................................... 203
  7.6.1 Model Estimates for E0 ............................................. 206
  7.6.2 Predictions for E0 .................................................. 208
  7.6.3 Comparison of MSE’s and CV’s of Predictors to MSE’s and CV’s of Direct Estimators for E0 ........................................ 212
  7.6.4 Model Assessment for E0 ........................................... 214
7.7 Summary ................................................................. 216

CHAPTER 8. Summary and Future Study ............................... 219
  8.1 Summary ............................................................... 219
  8.2 Future Study .......................................................... 225

APPENDIX 1 Linear approximation to the operation that converts totals to proportions ...................................................... 227

APPENDIX 2 Justification of the Linear Approximation for \( \hat{\lambda} \) ......................................................... 228

APPENDIX 3 Initial Estimator of \( \lambda_0 \) ........................................ 231

APPENDIX 4
  Distortion of Covariances in Bootstrap Data Generating Procedure ... 233

APPENDIX 5
  Distortion of Covariances in Bootstrap Data Generating Procedure ... 243

BIBLIOGRAPHY ................................................................. 256
### LIST OF TABLES

| Table 6.1 | Census proportions and 2-digit totals | 83 |
| Table 6.2 | Expected values of proportions and province two digit totals, CV’s of direct estimators of province two digit totals, expected province sample sizes \((n_k)\). \(n_k\) is the inverse of the squared CV for province \(k\) | 83 |
| Table 6.3 | Parameters for 2-stage simulation model | 84 |
| Table 6.4 | Ratio of \(c_k p_{T,ik}(1 - p_{T,ik})n_k^{-1}\) to MC variance of \(e_{ik}\). The \(c_k\), in the last row, are MC approximations. 2-stage sampling error model, \(\psi = 0.02\) | 84 |
| Table 6.5 | Optimal weights \(\gamma_k\) under SRS model | 85 |
| Table 6.6 | MC approximations for optimal \(\gamma_{ik}\) and working \(\gamma_k\) under 2-stage model, \(\psi = 0.02\) | 85 |
| Table 6.7 | Number of MC samples in which the modification for small \(\alpha_{ik}^*\) is used in each of 1-7 bootstrap samples | 87 |
| Table 6.8 | Number of MC samples (out of 2000) in which the multiplicative bias correction is used. SRS simulation, \(\psi = 0.02\) | 88 |
| Table 6.9 | Averages of ratios of MC MSE’s of predictors to MC MSE’s of direct estimators for SRS sampling error model | 90 |
| Table 6.10 | Averages of ratios of MC MSE’s of predictors to MC MSE’s of direct estimators for 2-stage sampling error model | 90 |
| Table 6.11 | Ratios of MC MSE’s of predictors of totals calculated with true parameters to sampling variances of totals, SRS sampling error model, \(\psi = 0.003\) | 94 |
Table 6.12  Averages of ratios of MC MSE’s of Beale predictors to MC MSE’s of standard predictors, SRS sampling error model.  

Table 6.13  Averages of ratios of MC MSE’s of Beale predictors to MC MSE’s of standard predictors, 2-stage sampling error model. MC SE’s for $\psi = 0.003 < 0.001$. MS SE’s for $\psi = 0.02 > 0.0035$. 

Table 6.14  Averages of ratios of MC MSE’s of raked predictors to MC MSE’s of initial predictors, SRS sampling error model. 

Table 6.15  Averages of ratios of MC MSE’s of raked predictors to MC MSE’s of initial predictors, 2-stage sampling error model. 

Table 6.16  Ratios of MC MSE’s of predictors calculated with direct estimator (no modification for zeros) to MC MSE’s of predictors calculated with model estimator of sampling covariance matrix. 2-stage simulation model, $\psi = 0.003$. 

Table 6.17  Averages of ratios of MC MSE’s of predictors calculated with direct estimator (no modification for zeros) to MC MSE’s of predictors calculated with model estimator of sampling covariance matrix. 

Table 6.18  Averages of ratios of MC variances of direct estimators of sampling variances (no modification for zeros) to MC variances of model based estimators of sampling variances, $\psi = 0.003$. 

Table 6.19  Averages of ratios of MC MSE’s of predictors calculated with modified direct estimator (modified for zeros) to MC MSE’s of predictors calculated with model estimator of sampling covariance matrix. 

Table 6.20  Standardized biases of model estimators of sampling variances, 2-stage sampling error model, $\psi = 0.003$. 

Table 6.21  Ratios of MC MSE’s of model estimators of sampling variances to MC MSE’s of direct estimators of sampling variances, 2-stage simulation, $\psi = 0.003$.  

Table 6.22 Empirical correlation between $\hat{\gamma}_{ik,B,dir}$ of (6.15) and $\hat{p}_{ik} - \hat{p}_{T,ik}$, 2-stage, $\psi = 0.003$ ......................................................... 106

Table 6.23 Empirical correlation between $\hat{\gamma}_{ik,B}$ of (4.32) and $\hat{p}_{ik} - \hat{p}_{T,ik}$, 2-stage, $\psi = 0.02$ ......................................................... 106

Table 6.24 Standardized biases of benchmarked predictors calculated with direct estimators of sampling variances, 2-stage, $\psi = 0.003$ ........................................... 107

Table 6.25 Standardized biases of benchmarked predictors calculated with modified direct estimators of sampling variances, 2-stage, $\psi = 0.003$ .................. 108

Table 6.26 Standardized biases of benchmarked predictors calculated with model estimators of sampling variances, 2-stage, $\psi = 0.003$ .................. 108

Table 6.27 Percent of MC MSE of $\hat{p}_{ik,B}$ explained by leading term (column 1), effect of parameter estimation (column 3), and the raking effect (column 4). Column 2 has ratios of empirical biases (row 1) and MC means of the Taylor (row 2) and bootstrap (row 3) estimators of biases of the estimator of the leading term to the MC MSE of $\hat{p}_{ik,B}$ (multiplied by 100). SRS, $\psi = 0.003$ ........................................... 115

Table 6.28 Percent of MC MSE of $\hat{M}_{ik,B}$ explained by leading term (column 1), effect of parameter estimation (column 3), and the raking effect (column 4). Column 2 has ratios of empirical biases (row 1) and MC means of the Taylor (row 2) and bootstrap (row 3) estimators of biases of the estimator of the leading term to the MC MSE of $\hat{M}_{ik,B}$ (multiplied by 100). SRS, $\psi = 0.003$ ........................................... 118

Table 6.29 Percent of MC MSE of $\hat{p}_{ik,B}$ explained by leading term (column 1), effect of parameter estimation (column 3), and the raking effect (column 4). Column 2 has ratios of empirical biases (row 1) and MC means of the Taylor (row 2) and bootstrap (row 3) estimators of biases of the estimator of the leading term to the MC MSE of $\hat{p}_{ik,B}$ (multiplied by 100). SRS, $\psi = 0.02$ ........................................... 121
Table 6.30 Percent of MC MSE of $\tilde{M}_{ik,B}$ explained by leading term (column 1), effect of parameter estimation (column 3), and the raking effect (column 4). Column 2 has ratios of empirical biases (row 1) and MC means of the Taylor (row 2) and bootstrap (row 3) estimators of biases of the estimator of the leading term to the MC MSE of $\tilde{M}_{ik,B}$ (multiplied by 100). SRS, $\psi = 0.02$ ............................... 123

Table 6.31 Percent of MC MSE of $\tilde{p}_{ik,B}$ explained by leading term (column 1), effect of parameter estimation (column 3), and the raking effect (column 4). Column 2 has ratios of empirical biases (row 1) and MC means of the Taylor (row 2) and bootstrap (row 3) estimators of biases of the estimator of the leading term to the MC MSE of $\tilde{p}_{ik,B}$ (multiplied by 100). Two stage, $\psi = 0.003$ ............................... 126

Table 6.32 Percent of MC MSE of $\tilde{M}_{ik,B}$ explained by leading term (column 1), effect of parameter estimation (column 3), and the raking effect (column 4). Column 2 has ratios of empirical biases (row 1) and MC means of the Taylor (row 2) and bootstrap (row 3) estimators of biases of the estimator of the leading term to the MC MSE of $\tilde{M}_{ik,B}$ (multiplied by 100). Two stage, $\psi = 0.003$ ............................... 129

Table 6.33 Percent of MC MSE of $\tilde{p}_{ik,B}$ explained by leading term (column 1), effect of parameter estimation (column 3), and the raking effect (column 4). Column 2 has ratios of empirical biases (row 1) and MC means of the Taylor (row 2) and bootstrap (row 3) estimators of biases of the estimator of the leading term to the MC MSE of $\tilde{p}_{ik,B}$ (multiplied by 100). Two stage, $\psi = 0.02$ ............................... 131
Table 6.34 Percent of MC MSE of $\tilde{M}_{ik,B}$ explained by leading term (column 1),
effect of parameter estimation (column 3), and the raking effect (column 4). Column 2 has ratios of empirical biases (row 1) and MC means of
the Taylor (row 2) and bootstrap (row 3) estimators of biases of the estimator of the leading term to the MC MSE of $\tilde{M}_{ik,B}$ (multiplied by
100). Two stage, $\psi = 0.02$ ................................................................. 133

Table 6.35 Expected province two digit totals and province sample sizes ........ 142

Table 6.36 Ratios of MC MSE’s of benchmarked predictors of proportions to MC
MSE’s of initial predictors of proportions in the simulation with sample
sizes given in Table 6.35 ................................................................. 142

Table 6.37 Averages of empirical coverages of nominal 95% confidence intervals for
proportions and totals. T=Taylor, B=bootstrap. SRS, $\psi = 0.003$. . . . 148

Table 6.38 Averages of empirical coverages of nominal 95% confidence intervals for
proportions and totals. T=Taylor, B=bootstrap. SRS, $\psi = 0.02$. . . . 149

Table 6.39 Averages of empirical coverages of nominal 95% confidence intervals for
proportions and totals. T=Taylor, B=bootstrap. Two Stage, $\psi = 0.003$. 150

Table 6.40 Averages of empirical coverages of nominal 95% confidence intervals for
proportions and totals. T=Taylor, B=bootstrap. Two stage, $\psi = 0.02$. 151

Table 6.41 Average widths of nominal 95% prediction intervals for proportions
constructed with Taylor MSE estimator. 2-stage sampling error model,
$\psi = 0.003$ ......................................................................................... 152

Table 6.42 Average widths of nominal 95% prediction intervals for proportions con-
structed with bootstrap MSE estimator. 2-stage sampling error model,
$\psi = 0.003$ ......................................................................................... 152

Table 6.43 Average widths of nominal 95% prediction intervals for proportions and
totals constructed with bootstrap MSE estimator. . . . . . . . . . . . . . 153

Table 6.44 Average widths of nominal 95% prediction intervals for proportions and
totals constructed with Taylor MSE estimator. . . . . . . . . . . . . . . 153
Table 6.45 \textit{aug2/rak: Average of ratios of MC MSE’s of }\hat{p}_{\text{pred},ik,aug2}\textit{ to MC MSE’s of }\tilde{p}_{ik}; \textit{aug1/rak: Average of ratios of MC MSE’s of }\hat{p}_{\text{pred},ik,aug1}\textit{ to MC MSE’s of }\tilde{p}_{ik}; \textit{aug1/aug2: Average of ratios of MC MSE’s of }\hat{p}_{\text{pred},ik,aug1}\textit{ to MC MSE’s of }\hat{p}_{\text{pred},ik,aug2}. \hspace{1.5cm} 156

Table 6.46 \textit{aug2/rak: Average of ratios of MC MSE’s of }\hat{p}_{\text{pred},ik,aug2}\textit{ to MC MSE’s of }\tilde{p}_{ik}; \textit{aug1/rak: Average of ratios of MC MSE’s of }\hat{p}_{\text{pred},ik,aug1}\textit{ to MC MSE’s of }\tilde{p}_{ik}; \textit{aug1/aug2: Average of ratios of MC MSE’s of }\hat{p}_{\text{pred},ik,aug1}\textit{ to MC MSE’s of }\hat{p}_{\text{pred},ik,aug2}. \hspace{1.5cm} 157

Table 6.47 \textit{aug2/rak: Average of ratios of MC MSE’s of }\hat{p}_{\text{pred},ik,aug2}\textit{ to MC MSE’s of }\tilde{p}_{ik}; \textit{aug1/rak: Average of ratios of MC MSE’s of }\hat{p}_{\text{pred},ik,aug1}\textit{ to MC MSE’s of }\tilde{p}_{ik}; \textit{aug1/aug2: Average of ratios of MC MSE’s of }\hat{p}_{\text{pred},ik,aug1}\textit{ to MC MSE’s of }\hat{p}_{\text{pred},ik,aug2}. \hspace{1.5cm} 157

Table 7.1 \textit{Estimated CV’s for A1, province sample sizes, and direct estimates of province 2-digit totals. \hspace{1.5cm} 160}

Table 7.2 \textit{Likelihood ratio test statistics defined in (7.1) for A1 and E0. \hspace{1.5cm} 174}

Table 7.3 \textit{Differences between AIC’s for multinomial model and AIC’s for full model, as defined in (7.2). \hspace{1.5cm} 174}

Table 7.4 \textit{Differences between BIC’s for multinomial model and BIC’s for full model, as defined in (7.3). \hspace{1.5cm} 175}

Table 7.5 \textit{Realized province sample sizes }\{n_k : k = 1, \ldots, 10\}\textit{ for A1 and E0. \hspace{1.5cm} 175}

Table 7.6 \textit{Estimates of }c_k\textit{ : }k = 1, \ldots, 10\textit{ and Taylor standard errors defined in (4.30). \hspace{1.5cm} 179}

Table 7.7 \textit{Census 2001 proportions and province two digit totals for A1 (specialist managers). \hspace{1.5cm} 188}

Table 7.8 \textit{May 2005 LFS proportions, province two digit totals }\hat{M}_k\textit{, and estimated coefficients of variation of the estimators of the province two digit totals for A1. \hspace{1.5cm} 188}
<p>| Table 7.9 | Estimates of the parameters of the reduced model for A1 and Taylor standard errors. Coefficient on Census interactions, $\theta_o = 1$ | 189 |
| Table 7.10 | Estimates of $\theta_o$ and $\psi$ for A1, A2, and A3 with Taylor standard errors. The last column has the lower bounds for the estimators of $\psi$. The second column has the number of three digit codes in each two digit code | 191 |
| Table 7.11 | Direct estimates of proportions and sampling standard deviations, A1 | 193 |
| Table 7.12 | Predictors of proportions and square roots of bootstrap MSE estimates, A1 | 193 |
| Table 7.13 | Estimates of ${p_{T,ik} : i = 1,\ldots,4; k = 1,\ldots,10}$ and Taylor standard errors, A1 | 194 |
| Table 7.14 | Ratios of bootstrap estimates of MSE’s to sampling variances for proportions in A1 | 198 |
| Table 7.15 | Ratios of bootstrap estimates of MSE’s to sampling variances for totals in A1 | 199 |
| Table 7.16 | Estimated Coefficients of Variation | 199 |
| Table 7.17 | Standardized residuals for A1 | 200 |
| Table 7.18 | Census 2001 proportions and province two digit totals for E0 (judges, lawyers, psychologists, social workers, ministers of religion, and policy and program officers) | 206 |
| Table 7.19 | May 2005 LFS proportions, province two digit totals ($\hat{M}_k$), and estimated coefficients of variation of the estimators of the province two digit totals for E0 | 206 |
| Table 7.20 | Estimates of the parameters of the reduced model for E0 and Taylor standard errors. Coefficient on Census interactions, $\theta_o = 1$ | 207 |
| Table 7.21 | Estimates of $\theta_o$ and $\psi$ for E0 and E1 with Taylor standard errors. The last column has the lower bounds for the estimators of $\psi$. The second column has the number of three digit codes in each two digit code | 208 |
| Table 7.22 | Direct estimates of proportions and sampling standard deviations, E0 | 210 |</p>
<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 7.23</td>
<td>Predictors of proportions and square roots of bootstrap MSE estimates, ( E_0 )</td>
</tr>
<tr>
<td>Table 7.24</td>
<td>Estimates of ( { \hat{p}_{T,ik} : i = 1, 2, 3; k = 1, \ldots, 10 } ) and standard errors, ( E_0 )</td>
</tr>
<tr>
<td>Table 7.25</td>
<td>Ratios of bootstrap estimates of MSE’s to sampling variances for proportions in ( E_0 ).</td>
</tr>
<tr>
<td>Table 7.26</td>
<td>Ratios of bootstrap estimates of MSE’s to sampling variances for totals in ( E_0 ).</td>
</tr>
<tr>
<td>Table 7.27</td>
<td>Estimated coefficients of variation for direct estimators of proportions in ( E_0 ).</td>
</tr>
<tr>
<td>Table 7.28</td>
<td>Estimated coefficients of variation for predictors of proportions in ( E_0 ).</td>
</tr>
<tr>
<td>Table 7.29</td>
<td>Standardized residuals defined in (7.9) for ( E_0 ).</td>
</tr>
<tr>
<td>Table A.1</td>
<td>Estimates ( \hat{p}_{T,ik} ) and ( \hat{M}_k ) and realized sample sizes from a single simulation run from the two-stage simulation with ( \psi = 0.003 ).</td>
</tr>
<tr>
<td>Table A.2</td>
<td>Standard deviations and correlations of the original estimator of the covariance matrix of the sampling errors in the totals (starting matrix) and the covariance matrix of the simulated totals (simulated matrix).</td>
</tr>
<tr>
<td>Table A.3</td>
<td>Ratios of diagonal elements of (A.13) to the diagonal elements of the generating matrix for totals.</td>
</tr>
<tr>
<td>Table A.4</td>
<td>Standard deviations and correlations of the original estimator of the covariance matrix of the sampling errors in the proportions (starting matrix) and the covariance matrix of the simulated proportions (simulated matrix).</td>
</tr>
<tr>
<td>Table A.5</td>
<td>Ratios of variances of ( \hat{p}<em>{ik}^{(b)} - p</em>{ik}^{(b)} ) to diagonal elements of ( \hat{\Sigma}_{ee,k,md} ).</td>
</tr>
<tr>
<td>Table A.6</td>
<td>Ratios of variances of ( e_k^{(b)} ) to diagonal elements of ( \hat{\Sigma}_{ee,k} ).</td>
</tr>
<tr>
<td>Table A.7</td>
<td>Ratios of variances of ( \check{e}<em>k^{(b)} ) to diagonal elements of ( \hat{\Sigma}</em>{ee,k} ).</td>
</tr>
<tr>
<td>Table A.8</td>
<td>Ratios of variances of ( \check{e}_k^{*} ) to direct estimators of sampling variances.</td>
</tr>
<tr>
<td>Table A.9</td>
<td>Ratios of variances of ( \check{e}_k^{(b)} ) to direct estimators of sampling variances.</td>
</tr>
</tbody>
</table>
Table A.10  Estimates of \( c_k : k = 1, \ldots, 10 \) with Taylor and bootstrap standard error estimates .................................................. 244

Table A.11  Estimates of \( \psi \) for A1 and E0 with bootstrap and Taylor standard error estimates ................................................................. 245

Table A.12  Estimates of \( \lambda_o \) for A1 and E0 with bootstrap and Taylor standard error estimates. E0 has 3 categories, so \( \alpha_{o,4} \) is not in the model for E0. 245

Table A.13  Percent differences between bootstrap MSE’s and Taylor MSE’s relative to Taylor MSE’s for A1 .......................................................... 247

Table A.14  Percent differences between bootstrap MSE’s and Taylor MSE’s relative to Taylor MSE’s for E0 .......................................................... 248

Table A.15  Percent differences between bootstrap and Taylor MSE estimates with Taylor MSE estimates evaluated at the bootstrap estimate of the variance of \( \hat{\psi} \). .................................................. 248

Table A.16  Ratios of bootstrap estimates of biases of estimates of leading terms to Taylor estimates of leading terms for A1 ................................. 252

Table A.17  Ratios of Taylor estimates of biases of estimates of leading terms to Taylor estimates of leading terms for A1 ................................. 252

Table A.18  Ratios of bootstrap estimates of biases of estimates of leading terms to Taylor estimates of leading terms for E0 ................................. 253

Table A.19  Ratios of Taylor estimates of biases of estimates of leading terms to Taylor estimates of leading terms for E0 ................................. 253

Table A.20  Relative differences between bootstrap and Taylor MSE estimators for A1 after eliminating effects of raking and estimating the biases of the estimators of the leading terms. ................................. 254

Table A.21  Relative differences between bootstrap and Taylor MSE estimators for E0 after eliminating effects of raking and estimating the biases of the estimators of the leading terms. ................................. 254
Table A.22  Percent differences between bootstrap and Taylor MSE estimates for
totals in A1 .............................................................. 255

Table A.23  Percent differences between bootstrap and Taylor MSE estimates for
totals in E0 .............................................................. 255
LIST OF FIGURES

Figure 6.1 Ratios of MC MSE’s of predictors of proportions and totals to MC sampling variances. Proportion $\hat{p}_{ik,B}$: ●, Total $\hat{M}_{ik,B}$: Δ. SRS sampling error model, $\psi = 0.003$. .................................................. 92

Figure 6.2 Ratios of MC MSE’s of predictors of proportions and totals to MC sampling variances. Proportion $\hat{p}_{ik,B}$: ●, Total $\hat{M}_{ik,B}$: Δ. SRS sampling error model, $\psi = 0.02$. .................................................. 93

Figure 6.3 MC relative biases of MSE estimators for proportions. B=Bootstrap, T=Taylor. SRS sampling error model, $\psi = 0.003$. ................................. 113

Figure 6.4 MC relative biases of MSE estimators for totals. B=Bootstrap, T=Taylor. SRS sampling error model, $\psi = 0.003$. ................................. 116

Figure 6.5 MC relative biases of MSE estimators for proportions. B=Bootstrap, T=Taylor. SRS sampling error model, $\psi = 0.02$. ................................. 119

Figure 6.6 MC relative biases of MSE estimators for totals. B=Bootstrap, T=Taylor. SRS sampling error model, $\psi = 0.02$. ................................. 122

Figure 6.7 MC relative biases of MSE estimators for proportions. B=Bootstrap, T=Taylor. Two stage sampling error model, $\psi = 0.003$. ................................. 124

Figure 6.8 MC relative biases of MSE estimators for totals. B=Bootstrap, T=Taylor. Two stage sampling error model, $\psi = 0.003$. ................................. 127

Figure 6.9 MC relative biases of MSE estimators for proportions. B=Bootstrap, T=Taylor. Two stage sampling error model, $\psi = 0.02$. ................................. 130

Figure 6.10 MC relative biases of MSE estimators for totals. B=Bootstrap, T=Taylor. Two stage sampling error model, $\psi = 0.02$. ................................. 132
Figure 6.11  Ratios of MC biases of bootstrap (B) and Taylor (T) estimators of leading terms to empirical leading terms, and ratios of MC means of bootstrap (b) and Taylor (t) estimators of biases of estimators of leading terms to empirical leading terms (multiplied by 100). SRS, $\psi = 0.003$, proportions. .................................................. 134

Figure 6.12  Ratios of MC biases of bootstrap (B) and Taylor (T) estimators of leading terms to empirical leading terms, and ratios of MC means of bootstrap (b) and Taylor (t) estimators of biases of estimators of leading terms to empirical leading terms (multiplied by 100). Two stage, $\psi = 0.003$, proportions. .................................................. 135

Figure 6.13  Ratios of MC biases of bootstrap (B) and Taylor (T) estimators of leading terms to empirical leading terms, and ratios of MC means of bootstrap (b) and Taylor (t) estimators of biases of estimators of leading terms to empirical leading terms (multiplied by 100). Two stage, $\psi = 0.02$, proportions. .................................................. 137

Figure 6.14  Ratios of MC biases of bootstrap (B) and Taylor (T) estimators of leading terms to empirical leading terms, and ratios of MC means of bootstrap (b) and Taylor (t) estimators of biases of estimators of leading terms to empirical leading terms (multiplied by 100). Two stage, $\psi = 0.02$, totals. .................................................. 138

Figure 6.15  MC relative biases of bootstrap (B) and Taylor (T) estimators of the contribution of estimation of the unknown parameters to the MSE. Two stage, $\psi = 0.003$, totals. .................................................. 139

Figure 6.16  Ratios of MC means of bootstrap estimates of the raking effects to the MC MSE’s of the predictors of the proportions (B). Ratios of empirical raking effects to MC MSE’s of the predictors of the proportions (E). Two stage, $\psi = 0.003$. .................................................. 141
Figure 6.17 y-axis: Ratios of differences between MC means of bootstrap estimators of benchmarked predictors and bootstrap estimators of initial predictors to MC MSE’s of benchmarked predictors. x-axis: Ratios of differences between MC MSE’s of benchmarked predictors and MC MSE’s of initial predictors. 144

Figure 7.1 Direct estimates of sampling variances and covariances (y-axis); corresponding elements of $[\text{diag}(\hat{p}_{k,ts}) - \hat{p}_{k,ts}^{\prime} \hat{p}_{k,ts}] n_{k,ts}^{-1} (t = 1, s = 1, 5)$ for A1 and E0 (x-axis). A1: closed circles; E0: open triangles 168

Figure 7.2 Direct estimates of sampling variances and covariances (y-axis); corresponding elements of $[\text{diag}(\hat{p}_{T,k,ts}) - \hat{p}_{T,k,ts}^{\prime} \hat{p}_{T,k,ts}] n_{k,ts}^{-1} (t = 1, s = 1, 5)$; for A1 and E0 (x-axis). A1: closed circles; E0: open triangles 169

Figure 7.3 Estimates of $\{c_{k,ts}: k = 1, \ldots, 10; t = 1, \ldots, D_s; s = 1, \ldots, 10\}$ calculated separately for the 31 two digit codes in the one digit codes A-J 177

Figure 7.4 Standardized residuals defined in (7.4) on the vertical axis and province sample sizes on the horizontal axis 179

Figure 7.5 Standardized residuals for evaluating goodness of fit of working model for sampling variances for A1. Residuals defined in (7.6) on y-axis; working model variance $\hat{c}_n^{-1} \hat{p}_{T,ik} (1 - \hat{p}_{T,ik})$ on x-axis. 181

Figure 7.6 Standardized residuals for evaluating goodness of fit of working model for sampling variances for E0. Residuals defined in (7.6) on y-axis; working model variance $\hat{c}_n^{-1} \hat{p}_{T,ik} (1 - \hat{p}_{T,ik})$ on x-axis. 182

Figure 7.7 Direct estimators of variances of direct estimators of province 2-digit totals (y-axis), and squares of direct estimators of province 2-digit totals multiplied by $\hat{c}_n^{-1} n_{k,ts}$ (x-axis). 184

Figure 7.8 Residuals from weighted regression of direct estimators of variances on $\hat{c}_n^{-1} M_{k,ts}^2 n_{k,ts}$. A1: open circle, A2: triangle, A3: plus, E0: ×, E1: diamond. 185
Figure 7.9  Direct estimates and Census proportions for A1. Left plot: logits of direct estimators of proportions (y-axis), Census interactions (x-axis). Right plot: LFS proportions (y-axis), Census proportions (x-axis). Numbers correspond to three digit codes, $i = 1, \ldots, 4$. 187

Figure 7.10  Direct estimates and predictors of proportions in A1 with corresponding prediction intervals. Direct estimates: + and solid lines. Predictors: · and dashed lines. 195

Figure 7.11  Standardized residuals for A1 (y-axis); $\hat{p}_{T,ik} (x$-axis) 201

Figure 7.12  Absolute values of standardized residuals for A1 (y-axis); $\hat{p}_{T,ik} (x$-axis) 201

Figure 7.13  Direct estimates and Census proportions for E0. Left plot: logits of direct estimators of proportions (y-axis), Census interactions (x-axis). Right plot: LFS proportions (y-axis), Census proportions (x-axis). Numbers correspond to three digit codes, $i = 1, 2, 3$. 205

Figure 7.14  Direct estimates and predictors of proportions in E0 with corresponding prediction intervals. Direct estimates: + and solid lines. Predictors: · and dashed lines. 212

Figure 7.15  Standardized residuals for E0 (y-axis), $\hat{p}_{T,ik} (x$-axis) 215

Figure 7.16  Absolute values of standardized residuals for E0 (y-axis), $\hat{p}_{T,ik} (x$-axis). 216

Figure A.1  Taylor MSE estimates (x-axis). Bootstrap MSE estimates (y-axis). Solid line is $y = x$. Two digit code A1. 246

Figure A.2  Taylor MSE estimates (x-axis). Bootstrap MSE estimates (y-axis). Solid line is $y = x$. Two digit code E0. 247

Figure A.3  Percent difference between bootstrap estimates of MSE’s of benchmarked predictors and Taylor MSE estimates (y-axis). Percent difference between bootstrap estimates of MSE’s of benchmarked predictors and bootstrap estimates of MSE’s of initial predictors (x-axis) for A1. 250
Figure A.4  Percent difference between bootstrap estimates of MSE’s of benchmarked predictors and Taylor MSE estimates (y-axis). Percent difference between bootstrap estimates of MSE’s of benchmarked predictors and bootstrap estimates of MSE’s of initial predictors (x-axis) for E0. 251
CHAPTER 1. Introduction

Survey agencies often desire estimates for small areas, domains in which realized sample sizes are too small to produce stable direct estimators. A widely adopted approach to small area estimation combines information across areas in a way that is optimal under a specified model. Fay and Herriot (1979) use the linear mixed model with normally distributed random effects and an assumption of a known sampling variance. Numerous applications extend the Fay and Herriot (1979) procedure to models with nonlinear expectation functions and non-normal distributions. Rao (2003) and Jiang and Lahiri (2006) review methods for small area estimation.

We consider a situation in which a current survey provides direct estimators for the cells in a two-way table. The direct estimators of the margins of the table are judged to have good design properties, but the direct estimators of the interior cells are unreliable due to small realized sample sizes. The objective is to obtain stable predictors of the interior of the two-way table that preserve the direct estimators of the margins.

Structure Preserving Estimation (SPREE) is a small area procedure that combines auxiliary information, often data from a previous census, with current survey data to improve the precision of estimators of the cell totals in a multi-way contingency table (Purcell and Kish, 1980). The idea underlying SPREE is that the dependence structure in the previous census holds in the current time period, while the census marginal levels may be out-dated. SPREE adjusts the interior of the table from the previous census in a way that preserves the interactions from the census and the margins from the current survey. We discuss SPREE and generalizations of SPREE in more detail Section 2. We also discuss alternative models that have been used to obtain small area predictors when the quantities of interest are counts and
Our investigation was motivated by the Canadian Labour Force Survey (LFS). The quantities of interest are the cells in the two-way table defined by occupations and provinces. The two-way table from the previous Census provides auxiliary information.

1.1 Occupations in the Canadian Labour Force Survey

Canada’s National Occupational Classification (NOC) organizes employment into occupations using a hierarchical system. Categories labeled with three digit codes are nested in categories labeled with two digit codes (Hidiroglou and Patak, 2009). For example, the two digit code A1 is the category for specialist managers. The four three digit codes A11-A14 subdivide specialist managers into more specific occupations. Each two-way table of interest is defined by a cross-classification of three digit codes and provinces.

In the LFS, direct estimates of occupational totals at various levels of detail are weighted sums, where the weights account for the LFS sample design, poststratification, and nonresponse. The direct estimators of two digit occupational totals at the province level and the direct estimators of national three digit totals (the margins of the two-way table) are judged to have adequate design coefficients of variation. Monthly estimates of the marginal totals are publicly available. Because occupations are unplanned domains in the LFS sample design, realized sample sizes in occupations are random. Small realized sample sizes cause the direct estimators of three digit totals and proportions in small provinces to have unacceptably large estimated coefficients of variation. Stable monthly estimators of three digit totals and proportions at the province level (the interior of the two-way table) are desired.

The Canadian Census of Population, conducted every five years, publishes occupational counts through the three digit level of detail for each province. The Census is a convenient source of auxiliary information because the Census tables are readily available across two digit codes and provinces.

In many of the two digit codes, the proportions in each three digit code calculated with the Census data by province are linearly related to the corresponding proportions calculated with
the direct LFS estimates. Although differences between the data collection protocols used in
the LFS and the Census lead to some differences in the resulting occupational data, the Census
provides a good source of auxiliary information for estimating three digit occupational totals
and proportions at the province level.

1.2 Challenges

The objectives of the LFS application present several challenges. Many are characteristic
of the difficulties encountered in small area estimation. Others are specific to estimation
of a table of counts. The model and method of this dissertation address several challenges
associated with estimation for the LFS.

First, because the quantities of interest are proportions and totals, predictors should
fall in the natural parameter space. Predictors of proportions should be between zero and one,
and predictors of totals should be nonnegative. The restriction on the parameter space makes
the usual predictors based on linear mixed models unacceptable. Predictors that respect the
parameter spaces are reviewed in Section 2.

Second, the LFS uses a complex design, and LFS estimation procedures account for
the selection probabilities, use auxiliary data, and adjust for nonresponse. As a consequence,
predictors and MSE estimators derived under an assumption of simple random sampling are
not appropriate for the LFS. Because the distribution of the direct estimators is unknown, it is
very difficult to specify a full likelihood and take advantage of the benefits of likelihood based
inference.

Third, predictors for the cells that preserve the direct estimators of the margins are
desired to give consistency with published records. Imposing that restriction is called bench-
marking in many small area applications. When the parameter space is unrestricted, linear
benchmarking procedures are often used. An advantage of linear benchmarking is that one can
obtain a closed form approximation for the effect of benchmarking on the MSE. A problem
with linear benchmarking for proportions or totals is that the benchmarked predictors can fall
outside the parameter space. Raking is a commonly used alternative when the quantities of
interest are totals and proportions because the predictors are easy to calculate and remain in the parameter space.

The unit level data are not available and little information is available about the distribution of the direct estimators of the sampling variances. Traditionally, estimated sampling obtained from a generalized variance function are treated as the true sampling variances. More recent studies address the problem of unknown sampling variances in both model specification and estimation.

The need for benchmarking and the uncertainty about the variances of estimated variances make estimation of the MSE more difficult. The effect of raking on the MSE is not thoroughly understood (Pfeffermann and Tiller, 2006). Much small area estimation literature focuses on obtaining estimates of MSE’s with biases of smaller order than $K^{-1}$, where $K$ denotes the number of independent areas. (See Section 2.) Obtaining MSE estimators with that order of accuracy requires estimating the variances of variance estimators.

1.3 Outline

In Section 2, we review previous applications of small area estimation that have dealt with some of the challenges discussed above. Sections 3-5 describe a procedure for obtaining predictors of three digit occupational totals and proportions for the LFS. We specify a model for the direct estimators in Section 3. In Section 4, we develop a model based predictor and define alternative estimators of the MSE in Section 5. In Section 6, we evaluate the properties of the procedure through simulation. We apply the procedure to two-way tables from the May 2005 LFS in Section 7.
CHAPTER 2. Literature on Small Area Estimation of Counts and Proportions

In the motivating application, predictors of the cell counts and proportions in a contingency table that preserve the direct estimators of the marginal totals and utilize the two-way table from a previous census are desired. Purcell and Kish (1980) introduced Structure Preserving Estimation (SPREE) as a way to update census data to timely estimators of the margins of a multi-way contingency table. We discuss the SPREE procedure and generalizations of SPREE. We then review alternative models used to obtain predictors of counts and proportions in small areas.

2.1 SPREE and Generalizations

Structure Preserving Estimation (SPREE) combines an auxiliary table, often from a previous census, to improve the estimators of the cell totals in a multi-way contingency table (Purcell and Kish, 1980). SPREE adjusts the interior of the auxiliary table in a way that preserves the interactions from the auxiliary table and preserves the margins from the current survey. Purcell and Kish (1980) implement SPREE by applying iterative proportional fitting.

Noble et al. (2002) explain that the model underpinning SPREE is a special case of a generalized linear model. The estimators of the cell totals obtained from SPREE are the maximum likelihood estimators of the expected counts under a generalized linear model with a Poisson random component and a log link. Main effects for rows and columns are estimated with the direct estimators. Interactions are set equal to the interactions in a saturated loglinear model fit to the census two-way table. The representation of SPREE as maximum likelihood estimation suggests Newton-Raphson as an alternative to iterative proportional fitting as a
way to implement SPREE (Noble et al., 2002).

Noble et al. (2002) extend the Poisson model underlying SPREE to the larger family of generalized linear models. In the general setting, the parameters of the linear predictor are partitioned into two sets: one set (e.g., the main effects in the case of SPREE) is estimated from the direct estimators and the second set (e.g., the interactions) from the auxiliary data. The response and explanatory variables do not need to be categorical, as in SPREE. Noble et al. (2002) illustrate the generalization of SPREE through an application to estimation of unemployment rates from the Household Labour Force Survey conducted by Statistics New Zealand. In the application, main effects for age and sex are estimated from the Labour Force Survey, while fixed effects associated with nine regions and interactions between age and sex are estimated from a previous census.

Griffiths (1996) considers two composite estimators of the cell totals in contingency tables defined by employment characteristics and income brackets in congressional districts in Iowa. One of the composite estimators is a convex combination of the SPREE estimator and the corresponding direct estimator. The weights used to form the convex combination depend on the estimated design MSE's of the SPREE estimators and the direct estimators. The second composite estimator is the EBLUP derived under a linear mixed model for the direct estimators of the totals.

Zhang and Chambers (2004) develop two extensions of the loglinear model underlying SPREE. Both are models for the true proportions of interest. The first one, called the generalized linear structural model (GLSM), is a loglinear model in which the interactions are assumed to be proportional to the interactions in a census. If the coefficient on the census interactions is assumed to equal one, then the GLSM simplifies to the loglinear model underlying SPREE. Predictors based on the GLSM are synthetic estimators. To obtain predictors with a smaller bias than predictors based on the GLSM, Zhang and Chambers (2004) extend of the GLSM to a random effects model called the generalized linear structural mixed model (GLSMM). In the GLSMM, the vector of interactions for a single area are assumed to have a multivariate normal distribution with a singular covariance matrix.
2.2 Generalized Linear Mixed Models

An alternative approach assumes that the conditional distributions of the direct estimators are members of an exponential family and models the natural parameters with normal distributions. In a Bayesian framework, Ghosh et al. (1998) demonstrate that the multinomial distribution for a multi-category response is in the exponential family and give a condition sufficient to guarantee that the posterior distribution is proper. In an application to exposure to hazards in the workplace, Ghosh et al. (1998) model observed counts as conditionally independent Poisson random variables, given the small area totals and normally distributed random effects. Lu and Larsen (2007) conduct a Bayesian analysis of a hierarchical Poisson loglinear model for enrollment in employment preparation courses in Iowa schools stratified by district size and area education agencies. Molina et al. (2007) predict employment, unemployment, and inactivity rates in the UK labour force, under the assumption that the observed totals in small areas follow a multinomial logit model with a random area effect. Molina et al. (2007) compare a bootstrap estimator to an analytic estimator of the frequentist MSE.

You et al. (2002) suggest an un-matched model as an alternative to a standard approach that applies nonlinear transformations to the direct estimators to justify use of the normal linear mixed model. The un-matched model specifies a normal linear model for the conditional distribution of the direct estimator given the true value and a nonlinear model with normally distributed random effects for the true value. You et al. (2002) argue that un-matched models have several appealing features. For one, the normal linear model captures design properties of the direct estimators, such as design unbiasedness or known features of the design variance. Simultaneously, the nonlinear link function restricts the parameter space for the true value appropriately. Use of an un-matched model can avoid problems associated with nonlinear transformations of direct estimates. For instance, in small areas with small sample sizes, a nonlinear function of an unbiased direct estimator of a total may have a non-negligible bias for the corresponding function of the expected value of the direct estimator of the total (You et al., 2002). Also, some nonlinear functions such as the log and the logit are undefined when the direct estimate is zero. You et al. (2002) apply un-matched models to estimate the net
undercoverage in small domains from the Canadian Census and conduct a Bayesian analysis. Torelli and Trevisani (2008) discuss potential applications of un-matched models in the Italian Labour Force Survey. Mohadjer et al. (2007) use unmatched models to obtain Bayes estimators of the proportions of individuals with a low literacy level in states and counties in the U.S. based on the National and State Assessments of Adult Literacy. Fabrizi et al. (2008) employ un-matched models to estimate proportions of households in three ordered poverty classes (severe poverty, poverty, and at risk of poverty) in domains defined by cross-classifications of twenty administrative regions in Italy and nine household types.

2.3 Exponential Quadratic Variance Function Models

Ghosh and Maiti (2004) discuss use of exponential quadratic variance function families for small area estimation. The normal linear mixed model is a special case. Other examples include the beta-binomial, gamma-Poisson, and multinomial-Dirichlet distributions, which may be more appropriate than the normal distribution when the true quantity to be predicted for each small area is a single proportion, a total, or a vector of proportions, respectively. Ghosh and Maiti use the beta-Binomial model to estimate poverty rates in counties in the U.S. and compare their method to a procedure similar to the procedure that the U.S. Census Bureau uses. The log normal model that the Census Bureau uses is described in Ghosh and Maiti (2004).

2.4 Semi-parametric Models

Semi-parametric models for small area estimation avoid full distributional assumptions by specifying moments of the conditional distributions of the direct estimators given the true values and moments of the distributions of the true values. One way to obtain a predictor in a semi-parametric context is to use the linear estimator that minimizes the mean squared error of the predictor (Rao, 2003, pg. 214). An alternative approach adds the assumption of posterior linearity (Ghosh and Lahiri, 1987), which assumes that the conditional distribution of the true value given the direct estimator is linear. Under posterior linearity, the predictor
that minimizes the mean squared error is linear. In a simulation experiment, Ghosh and Lahiri (1987), find that estimators derived under posterior linearity sometimes have mean squared errors smaller than the mean squared errors of estimators based on full distributional assumptions. Raghunathan (1993) incorporates covariates into semi-parametric models for small area estimation and derives predictors using quasi-likelihoods to specify quasi-posterior predictive distributions. Longford (1999) uses a linear link function to specify a multivariate mixed linear model for vectors of small area proportions. The model leads to estimators that borrow strength not only from across small areas but also from across sub-populations within a single small area. The SPREE procedure and its extensions are also examples of methods that do not rely on full distributional assumptions.

2.5 Extensions and Reviews

Several studies compare and contrast the types of models described above and also propose extensions. Torelli and Trevisani (2008) review models for discrete data, arguing that the assumptions of linear mixed models for continuous responses are often unrealistic. Torelli and Trevisani (2008) propose non-normal hierarchical models to describe over-dispersion in the sampling distributions of the direct estimators, conditional on the true underlying quantities of interest. In particular, Torelli and Trevisani (2008) suggest Poisson-lognormal and gamma-Poisson mixture models for the direct survey estimators, conditional on the true values to be predicted. Also, specific issues from the Italian Labour Force Survey motivate Torelli and Trevisani (2008) to develop adaptations for situations in which the geographic locations associated with auxiliary data differ from the small areas of interest. Liu et al. (2007) evaluate the design properties of Bayes credible intervals for proportions under several hierarchical models. The models include a traditional normal linear mixed model and an un-matched model with normally distributed sampling errors and a logit link function. Liu et al. (2007) also propose a model in which the direct survey estimators are conditionally independent beta random variables, and the logits of the true proportions have normal distributions. The reviews by Rao (2003) and Jiang and Lahiri (2006) include discussions of applications in which the
quantities to predict are counts and proportions.

2.6 Benchmarking

The model-optimal small area predictors may not satisfy the “benchmarking” requirement that the total of the predicted values in a region equals the direct estimate for the region. Survey agencies often impose a benchmarking restriction when an estimator for the larger region has desirable sampling properties such as design unbiasedness and a small design coefficient of variation. Benchmarking ensures that the survey agency publishes internally consistent estimates and may reduce a bias due to model misspecification.

Wang et al. (2008) review several benchmarking methods in the context of linear mixed models and propose an augmented model under which the standard EBLUP satisfies the benchmarking restriction. Ugarte et al. (2008) benchmark the small area predictors to a synthetic estimator for the larger domain by minimizing a weighted sum of squared residuals subject to the benchmarking restriction. For example, You et al. (2002a) use a ratio adjustment to benchmark hierarchical Bayes predictors of the net Census undercoverage in small domains. Lu and Larsen (2007) use a discrepancy measure based on the difference between the posterior mean squared error of a hierarchical Bayes estimator and a benchmarked estimator for guidance in model selection. Torelli and Trevisani (2008) achieve the benchmarking requirement by including a model for the direct estimators of the totals in large domains in the hierarchical framework.

While benchmarking is desirable in many applications, consistency between the totals of the small area estimates and the estimators for larger domains is not always necessary. Mohadjer et al. (2007) give an example in which benchmarking is inappropriate due to differences between the definitions used at different levels of aggregation.

2.7 MSE Estimation

Quantifying the uncertainty associated with small area predictors is a challenge in small area estimation. Traditionally, small area estimation emphasizes (approximately) unbiased

In the context of the Fay-Herriot (1979) model, Prasad and Rao (1990) derive an analytic approximation to the MSE of the EBLUP constructed with a particular moment based estimator of the small area random effects variance and prove that an estimator of the MSE has a bias of smaller order than $n^{-1}$, where $n$ is the number of areas. The derivation of Prasad and Rao (1990) assumes that the small area random effects and sampling errors are normally distributed and that the sampling variances are known. Lahiri and Rao (1995) demonstrate that the MSE estimator due to Prasad and Rao (1990) is robust to nonnormality of the small area random effects in the sense that the MSE estimator of Prasad and Rao (1990) retains the desired order of accuracy under certain moment assumptions for the small area random effects. Datta et al. (2000) derive MSE estimators for predictors constructed with maximum likelihood, and REML estimators of the variance of the small area random effects. Datta, Rao, and Smith (2005) derive MSE estimators for predictors based on the Fay-Herriot estimator of the variance of the small area random effects. Datta, Rao and Smith (2005) compare several estimators of the variance of the small area random effects and corresponding MSE estimators in a simulation study.

The analytic approximations discussed above are derived under the linear mixed effects model. Adaptations of the Prasad and Rao (1990) approach to nonlinear models with nonnormal errors have also been studied. For example, Gonzalez-Manteiga et al. (2007) extend the Prasad-Rao MSE approximation to logistic mixed linear models. Slud and Maiti (2006) obtain an approximation for the MSE when the final predictor involves exponentiating an EBLUP of a random effect.

Jackknife estimators in the context of small area estimation involve replicating the estimation procedure after deletion of a single small area. Jiang, Lahiri, and Wan (2002) develop a jackknife estimator of the MSE applicable to the EBLUP of the linear mixed model (not assuming normality) and to the EBP (empirical best predictor) for nonlinear models. A disad-
vantage of the Jiang, Lahiri, and Wan (2002) estimator is that the estimator can be negative. Chen and Lahiri (2002) propose an alternative that combines an analytic approximation for the variance due to estimation of regression coefficients with a jackknife estimator of the variance of due to estimation the unknown small area effects variance. The form of Chen and Lahiri (2002) ignores a cross product term, which is zero under normality but can be nonzero in general. Chen and Lahiri (2002) also propose a weighted version of the jackknife estimator and an expression that is guaranteed to be nonnegative. Chen and Lahiri (2008) “obtain the Taylor series approximation to the Jackknife [MSE for] a general linear mixed normal model.” In the special case of the Fay-Herriot model, the resulting estimator uses the Taylor approximation of Datta et al. (2000) in conjunction with jackknife estimators of small area effects variances. Lohr and Rao (2009) propose a jackknife estimator of the MSE conditional on the observed data. The Lohr and Rao (2009) MSE estimator is applicable to empirical best predictors in the context of natural exponential quadratic variance function models.

**Bootstrap Estimators**

Several parametric and nonparametric bootstrap estimators of variances have also been proposed. Pfeffermann and Glickman (2004) propose three parametric bootstrap estimators of the MSE and one nonparametric bootstrap estimator. The nonparametric bootstrap estimator of Pfeffermann and Glickman (2004) involves sampling from standardized residuals with replacement and ignores a cross-product term, which is zero under normality but nonzero in general. Hall and Maiti (2006) develop a nonparametric double bootstrap estimator of the MSE in the context of a unit-level model with uncorrelated sampling errors and uncorrelated small area random effects. The Hall and Maiti (2006) procedure first obtains nonparametric estimators of the fourth moments of the small area effects and sampling errors from the residuals and then simulates from parametric distributions uniquely specified by the second and fourth moments.

Lahiri et al. (2007) and Lahiri and Maiti (2006) use a method similar to that of Tin (1965) to estimate the bias of an estimator of the leading term in the MSE. They define a modification to parameter estimator so that the bias of the leading term in the MSE estimator
evaluated at the modified parameter estimator is of smaller order than \( n^{-1} \). The model of Lahiri and Maiti (2006) assumes a linear expectation function with additive errors and known sampling variances but allows for nonnormality. Lahiri et al. (2007) consider a more general hierarchical model allowing a nonlinear expectation function and unknown sampling variances. The more general model of Lahiri et al. (2007) results in a greater computational burden, as Lahiri and Maiti (2006) simulate from several values of the parameter to approximate both the MSE and also the bias of an estimator of the MSE numerically.

Wang and Fuller (2003) point out that, although small area estimation literature has focused on unbiased estimation of the MSE, prediction intervals provide a potentially more useful measure of uncertainty for an analyst. Chatterjee and Lahiri (2008) use the parametric bootstrap to approximate the distribution of standardized prediction errors. In the model of Chatterjee and Lahiri (2008), the variances of normally distributed small area random effects and sampling errors are functions of a set of \( q \) parameters. In the asymptotic framework of Chatterjee and Lahiri (2008), the total number of parameters, \( q \), is allowed to increase as the sample size increases. Hall and Maiti (2006 a) obtain parametric bootstrap intervals that are not necessarily centered around the small area estimator.

### 2.8 Estimation of Unknown Sampling Variances

The best predictors or best linear unbiased predictors of small area means are functions of the sampling variances. The sampling variances reflect properties of the sample design and estimation procedures, including unknown selection probabilities, adjustments for nonresponse, and calibration to population controls. In practice, the true sampling variances are unknown and need to be estimated. Several ways to handle the problem of unknown sampling variances are discussed below.

Traditionally, estimates of the sampling variances are treated as the true sampling variances for the purposes of constructing predictors and estimates of the prediction MSE (Rao, 2003, pg. 76). One way to justify ignoring the variance of the estimator of the sampling variance is to smooth a direct estimator of the sampling variance. For example, Fay and
Herriot (1979) specify a model for the log of the direct estimator of the per capita income in each area. They assume that the coefficient of variation of the direct estimator of the per capita income in area \( k \) is \( 3\hat{N}_k^{-0.5} \), where \( \hat{N}_k \) is a weighted estimate of the total in area \( k \). The model for the coefficients of variation justifies an assumption that the variance on the log scale is \( 9\hat{N}_k^{-1} \). Fay and Herriot (1979) treat the estimates \( 9\hat{N}_k^{-1} \) as the true sampling variances. More recently, the EBLUP has been used to obtain estimates of the number of school aged children in poverty using data from the Current Population Survey. The model for the direct logs of the direct estimates of the poverty counts assumes that the sampling variance is of the form \( \sigma^2e_n^{-1} \), where \( n_i \) is the sample size in area \( k \). The constant \( \sigma^2 \) is estimated from the area level data and then treated as fixed. (Rao, 2003, pg. 124). Generalized variance functions (Valliant, 1987) provide one way to smooth a direct estimate of a sampling variance obtained from unit level data. Several applications treat the smoothed estimate of the sampling variance as the true value. For example, Dick (1995) uses generalized variance functions to model direct estimates of variances of estimators of the undercount from the 1990 Canadian Census in small domains. Singh et al. (2005) use generalized design effects to smooth direct estimates of covariance matrices. O’Malley and Zaslavsky (2005) propose alternative generalized variance functions for covariance matrices of transformed ordinal responses. The estimators of the covariance matrices are obtained through iteratively reweighted least squares.

If it is difficult to find a suitable model for the sampling variances, an alternative is to use the direct estimates of the sampling variances. Battese, Harter, and Fuller (1988) treat direct estimates of the sampling variances obtained from unit level data as fixed constants for MSE estimation. In areas with small sample sizes, the direct estimator of the sampling variance may
be unreliable. Bell (2008) discusses potential drawbacks associated with treating the direct estimators of the sampling variances as known. First, use of an unstable variance estimator in weighted least squares may lead to an inefficient estimator of the regression coefficients. Second, if a direct estimator of a variance is too small, then the corresponding direct estimator of the mean is assigned an unduly large weight in the EBLUP predictor. Third, failure to account for variability in the estimators of the sampling variances may lead to underestimation of the MSE of the predictors.

Rivest and Vandal (2003) and Wang and Fuller (2003) study the effects of estimated sampling variances on the frequentist MSE of small area predictors. Rivest and Vandal (2003) derive an approximation for the MSE of a predictor calculated with an estimator of a sampling variance under an assumption that the rest of the model parameters are known and that the estimator of the sampling variance has an approximately normal distribution. In the simulation study of Rivest and Vandal (2003), the estimator of the sampling variance is the sample variance of the units in a particular small area. In their application, the estimator of the sampling variance is obtained from a generalized variance function fit to several small areas. Because a large number of small areas are used to estimate the parameters of the generalized variance function, the contribution from the variance of the estimator of the sampling variance to the MSE of the predictor is negligible. Wang and Fuller derive an analytic approximation for the MSE of a predictor calculated with an estimator of the sampling variance. The theory assumes that the estimator of the regression coefficient is a weighted least squares estimator with a known, diagonal weight matrix and that the estimators of the sampling variances are independent of the model errors. In the asymptotic framework, the degrees of freedom associated with the direct estimators of the sampling variances and the number of small areas increase. In a simulation with normally distributed data, the MSE’s of predictors based on an estimated generalized least squares estimator are uniformly smaller than the MSE’s of predictors based on an OLS estimator of the regression coefficients. The theoretical (large sample) MSE closely approximates the true MSE when the sampling variances are large or when the variance of the random effect is large. When the variance of the random effects is small
relative to the sampling variance and the sampling variance is small, the analytic approximation for the MSE is too large, and an initial estimator of the MSE severely overestimates the true MSE. Wang and Fuller (2003) give two modifications to the initial estimator of the MSE that reduce the overestimation. Wang and Fuller (2003) study the coverage properties of confidence intervals constructed with the improved MSE estimator. In the simulations, the coverages tend to exceed the nominal level when the variance of the random effects is small and approach the nominal level for larger values of the random effects variance. Wang and Fuller (2003) also assess the robustness of the MSE estimators through simulation. In the presence of skewed distributions, they suggest using an ordinary least squares estimator of the regression coefficients and using a $\chi^2$ distribution with an estimated degrees of freedom to estimate the variance of the estimator of the sampling variance.

In a Bayesian context, You (2008) and You and Chapman (2006) incorporate models for sampling variances into the modeling and estimation procedures. You (2008) contrasts two models for the sampling variances with an approach that treats a smoothed estimate of the covariance matrix as fixed. The two models assume that the covariance matrix is a function of the true parameter to be predicted. Uncertainty in the sampling variances is reflected through the posterior predictive distribution. You and Chapman (2006) conduct Bayesian analyses of two models: in one, the sampling variance is assumed to have a distribution proportional to a $\chi^2$ distribution, and in the other, the sampling variance is treated as known. The two models are applied to the county crop acres data set of Battese, Harter, and Fuller (1988) and to the milk production data set of Datta et al. (1999). The milk production data set consists of 43 areas, each with between 95 and 633 units, and the posterior MSE’s of the predictors are similar, regardless of whether or not the sampling variances are treated as known. In contrast, the county crop acres data set consists of eight areas, each with between three and five units. You and Chapman (2006) note that when the small area sample sizes are small and the number of areas are small, treating the sampling variances as known leads to significantly smaller estimates of the posterior mean squared error of the small area predictors.

An alternative approach specifies a hierarchical model for the sampling variances and
uses optimal predictors of the sampling variances under the model. For example, Maples and Bell (2009) assume that the direct estimates of the sampling variances have conditional gamma distributions given the true sampling variances, and that the true sampling variances have inverse-gamma distributions. The resulting estimators of the sampling variances are convex combinations of the direct estimators and model based estimators. O’Malley and Zaslavsky (2005) specify hierarchical models for estimates of sampling variances of transformed ordinal responses. Cho et al. (2002) use hierarchical models to obtain empirical Bayes estimators of sampling variances using data from the Current Population Survey.
CHAPTER 3. A Model for Vectors of Small Area Proportions

Using terminology appropriate for the Canadian Labour Force Survey, suppose each of the provinces is partitioned into \( K \) different categories. Let \( k \) index the provinces \((k = 1, \ldots, K)\), and let \( i \) index the categories \((i = 1, \ldots, C)\). The quantities of interest are the cell totals and column proportions in the two-way table with categories as rows and provinces as columns.

Let \( M_{ik} \) be the true total in category \( i \) and province \( k \), and let the true proportion be \( p_{ik} \), where \( p_{ik} = M_{ik}^{-1}M_k \) and \( M_k = \sum_{i=1}^{C} M_{ik} \). Assume that the two-way table from a previous census is available. Let \( N_{ik} \) denote the total in three digit code \( i \) and province \( k \) in the census. The census totals are treated as fixed constants.

The current survey provides direct estimators of the true totals and proportions. Let \( \hat{M}_{ik} \) and \( \hat{p}_{ik} \) be the direct estimators of a total and proportion, respectively, where \( \hat{p}_{ik} = \hat{M}_{ik}^{-1}\hat{M}_k \) and \( \hat{M}_k = \sum_{i=1}^{C} \hat{M}_{ik} \). Assume that the direct estimator of the proportion satisfies

\[
\hat{p}_{ik} = p_{T,ik} + u_{ik} + e_{ik},
\]  
(3.1)

where \( u_{ik} \) is a mean zero random small area effect, \( e_{ik} \) is a mean zero sampling error, and \( p_{T,ik} \) is a function of the census totals and a vector of parameters, \( \lambda_o \). Assume \( u_{ik} \) and \( e_{jt} \) are uncorrelated for all \( i, k, j, t \). Because we treat the census totals as fixed, \( p_{T,ik} \) is a fixed parameter. The true proportion to be predicted is

\[
p_{ik} = p_{T,ik} + u_{ik}.
\]  
(3.2)

The model (3.1) expresses the direct estimator as a sum of three parts: the fixed expected value, \( p_{T,ik} \), the random small area effect, \( u_{ik} \), and the random sampling error, \( e_{ik} \).

We discuss specific assumptions about the function defining the fixed parameter \( p_{T,ik} \) and the variances of the random components in the following subsections.
3.1 Fixed Expected Value

The census proportion is \( p_{c,i,k} = N_k^{-1} N_{ik} \), where \( N_{ik} \) is the total in three digit code \( i \) and province \( k \) in the census. Assuming the census totals are positive, the census proportion has the loglinear representation,

\[
\log \left( \frac{p_{c,i,k}}{p_{c,1,k}} \right) = \alpha_{i}^{cen} + (\alpha \beta)_{i,k}^{cen},
\]

where \( \alpha_{i}^{cen} \) and \( (\alpha \beta)_{i,k}^{cen} \) are maximum likelihood estimators constructed under an assumption that the census totals are independent Poisson random variables. Under the restrictions that \( \alpha_{1}^{cen} = (\alpha \beta)_{11}^{cen} = (\alpha \beta)_{1k}^{cen} = 0 \) for \( i = 1, \ldots, C \) and \( k = 1, \ldots, K \), \( \alpha_{i}^{cen} = \log(N_{i1}) - \log(N_{11}) \), and \( (\alpha \beta)_{i,k}^{cen} = \log(N_{ik}) - \log(N_{1k}) - \log(N_{i1}) + \log(N_{11}) \).

To specify a model for the \( p_{T,i,k} \) of (3.1), let

\[
p_{T,i,k}(\lambda) = \frac{\exp(\eta_{ik}(\lambda))}{\sum_{j=1}^{C} \exp(\eta_{jk}(\lambda))}, \tag{3.3}
\]

where \( \lambda = (\alpha_2, \ldots, \alpha_C, \theta)' \),

\[
\eta_{ik}(\lambda) = \log \left( \frac{p_{T,i,k}(\lambda)}{p_{T,1,k}(\lambda)} \right) = \alpha_i + \theta(\alpha \beta)_{i,k}^{cen}, \tag{3.4}
\]

and \( \alpha_1 = 0 \). The fixed expected value, \( p_{T,i,k} \), is the value of the function in (3.3) evaluated at the true parameter \( \lambda_o \);

\[
p_{T,i,k} = p_{T,i,k}(\lambda_o) = \frac{\exp(\eta_{ik})}{\sum_{j=1}^{C} \exp(\eta_{jk})}, \tag{3.5}
\]

where \( \eta_{ik} = \eta_{ik}(\lambda_o) = \alpha_{o,i} + \theta_o(\alpha \beta)_{i,k}^{cen} \), \( \lambda_o = (\alpha_{o,2}, \ldots, \alpha_{o,C}, \theta_o) \), and \( \alpha_{o,1} = 0 \).

Zhang and Chambers (2004) use the representation (3.5), called the generalized linear structural model (GLSM) by them, as a model for the true proportion, \( p_{ik} \). Our (3.5) is a model for the expected value of \( p_{ik} \). The parameter \( \theta_o \) permits the interactions in the two-way table of expected values for the current time point to be proportional to the interactions in the two-way table of census totals. The assumption that \( \theta_o = 1 \) produces the loglinear model underlying SPREE. Allowing \( \theta_o \) to differ from 1 relaxes the assumption that the interactions in the census persist unchanged through time.
If rows and columns are independent in the census (the interactions are zero), then (3.5) specifies rows and columns in the table of expected values for the current time point to be independent as well. In this case, \( p_{T,ik} = p_{T,i1} \) for \( k = 1, \ldots, K \), and \( \theta_o \) is not identifiable.

The census interactions are covariates in the model (3.5) for the expected values of the proportions. Other choices of covariates are also possible. For example, one could replace the linear function of the census interactions in (3.4) with a different function of the census totals. Noble et al. (2002) generalize SPREE by replacing the census interactions with a different set of covariates estimated from the census. For example, they suggest estimating regional effects and parameters linked to continuous variables from census data. They then include these covariates in the expectation function for the current time point. In the model of Noble et al. (2002), the coefficient assigned to the covariate obtained from the census is set equal to 1 in the model for the current time point. Allowing the coefficient to differ from 1 generalizes the model of Noble et al. (2002).

### 3.2 Small Area Effects

Let \( \mathbf{u}_k = (u_{1k}, \ldots, u_{Ck})' \) denote the vector of random effects for a province. Assume that \( E[u_{ik}] = 0 \) and that the population covariance matrix for \( \mathbf{u}_k \) is

\[
\Sigma_{uu,k} = \psi[\text{diag}(\mathbf{p}_{T,k}) - \mathbf{p}_{T,k}\mathbf{p}_{T,k}' ] := \psi \Gamma_{uu,k}, \tag{3.6}
\]

where \( \mathbf{p}_{T,k} = (p_{T,1k}, \ldots, p_{T,Ck})' \), \( \psi \) is a constant to be estimated, and \( \text{diag}(\mathbf{v}) \) denotes a diagonal matrix with the vector \( \mathbf{v} \) on the diagonal. The covariance matrix (3.6) holds under a variety of models for the random effects including multinomial and Dirichlet distributions with parameter \( \mathbf{p}_{T,k} \). Assuming the covariance between \( \mathbf{u}_k \) and \( \mathbf{u}_t \) is zero for \( t \neq k \), the covariance matrix of the vector of small area effects, \( \mathbf{u} = (\mathbf{u}_1', \ldots, \mathbf{u}_K')' \), is block-diagonal with \( \psi \Gamma_{uu,k} \) as the \( k^{th} \) block. Let \( \Sigma_{uu} \) denote the block diagonal covariance matrix of \( \mathbf{u} \), where \( \Sigma_{uu} = \psi \Gamma_{uu} \). Because the columns of the covariance matrix in (3.6) sum to zero, \( \sum_{i=1}^C u_{ik} = 0 \) for all \( k \).

The model (3.1) differs from the GLSMM of Zhang and Chambers (2004) in two ways. One, the random effects, \( \{ u_{ik} : i = 1, \ldots, C; k = 1, \ldots, K \} \) in (3.1) are specified in the
scale of $p_{ik}$, while the random effects in the GLSMM are specified in the scale of $\log(p_{ik}) - C^{-1} \sum_{j=1}^{C} \log(p_{jk})$. Also, the covariance matrix in (3.6) is proportional to a specified function of the mean. The covariance matrix of the random effects in Zhang and Chambers (2004) is a $C \times C$ matrix with the property that rows and columns sum to zero. The form of (3.6) allows us to use multiple two-way tables to estimate a common $\psi$ even if the dimensions of the tables differ.

3.3 Sampling Errors

Like many large scale surveys, the LFS uses a complex design, and direct estimators are constructed to reflect the selection probabilities as well as adjustments for nonresponse and control totals. The errors $e_k = (e_{1k}, \ldots, e_{Ck})'$ ($k = 1, \ldots, K$) of (3.1) represent the effects of sampling and estimation on the direct estimators of the proportions.

Assume that $E[e_{ik}] = 0$ so that the direct estimators are unbiased for the true proportions defined in (3.2). Let $\Sigma_{ee,k}$ denote the variance of $e_k$. Assume sampling is independent across provinces, so $\Sigma_{ee}$, the covariance matrix of the vector of design errors $e = (e'_1, \ldots, e'_K)'$, is block-diagonal. Because the direct estimators of the proportions in a province sum to 1, the covariance matrix for $e_k$ is such that $\sum_{i=1}^{C} e_{ik} = 0$ for all $k$.

A distinction is to be made between the conditional variance of $e_k$ given the finite population and the unconditional variance. Let $V(e_k | u_k)$ denote the variance of $e_k$ given $u_k$. The survey often provides a direct estimator of $V(e_k | u_k)$. For example, the LFS uses the jackknife to estimate the sampling covariance matrix in the scale of totals, and a Taylor approximation (given in Appendix 1) is used to convert the jackknife covariance matrix for the totals to a direct estimate of the sampling covariance matrix in the scale of proportions. The direct estimator is design consistent and approximately design unbiased for $V(e_k | u_k)$ given the finite population. The unconditional variance, defined previously as $\Sigma_{ee,k}$, is $E[V(e_k | u_k)]$. If the survey’s direct estimator of the variance is conditionally unbiased for the conditional variance, then the survey’s direct estimator of the variance is unbiased for $\Sigma_{ee,k}$. In contrast, even if the direct estimator of the covariance matrix is design consistent for $V(e_k | u_k)$, the
direct estimator of the covariance matrix may not be consistent for $\Sigma_{ee,k}$ because the random effects $u_k$ do not converge in probability to zero as the sample size increases.

To clarify the difference between the conditional covariance matrix and the unconditional covariance matrix, consider a specific case in which the sample design is simple random sampling with replacement with a fixed sample size $n_k > 1$ for province $k$ and $\hat{p}_k$ is the sample proportion in category $i$. The conditional covariance matrix of the vector of sampling errors for province $k$ is

$$V(e_k | u_k) = n_k^{-1}[\text{diag}(p_k) - p_k p'_k]$$

and, under an assumption that $E[e_k | u_k] = 0$, the unconditional variance is

$$\Sigma_{ee,k} = E[V(e_k | u_k)] = n_k^{-1}(1 - \psi)[\text{diag}(p_{T,k}) - p_{T,k} p'_{T,k}].$$

A direct estimator of the sampling variance is

$$\hat{\Sigma}_{ee,k} = \frac{1}{n_k - 1}[\text{diag}(\hat{p}_k) - \hat{p}_k \hat{p}'_k]. \quad (3.7)$$

Under simple random sampling with replacement, the estimator (3.7) is conditionally unbiased for $V(e_k | u_k)$ and unconditionally unbiased for $\Sigma_{ee,k}$. Also, for fixed $u_k$, $n_k \hat{\Sigma}_{ee,k} - n_k V(e_k | u_k)$ converges in probability to zero as $n_k$ increases. Because $u_k$ does not converge to zero, $\hat{\Sigma}_{ee,k}$ is not a consistent estimator of $\Sigma_{ee,k}$.

In the model, the true sampling variance is a covariance matrix with the property that columns sum to zero. In practice, the true sampling variance is unknown, and many possibilities are available for estimating $\Sigma_{ee,k}$. One approach (e.g., Singh et al., 2005) uses a model such as a generalized variance function to smooth the direct estimators of the sampling variances. The smoothed variances are then treated as the true sampling variances.

A convenient way to model the sampling variances is to assume that the conditional variance is of the form,

$$V(e_k | u_k) = d_k \frac{n_k}{n_k - 1}[\text{diag}(p'_k) - p_k p'_k],$$

where $d_k$ is a scale factor that depends on the sample design and the strata.
where $d_k$ represents a design effect for province $k$, and $n_k$ represents a realized sample size. Then

$$
\Sigma_{ee,k} = \frac{c_k}{n_k} \left[ \text{diag}(p_{T,k}) - p_{T,k}p_{T,k}' \right],
$$

(3.8)

where $c_k = d_k(1 - \psi)$, and the sampling variance is proportional to the variance of $u_k$. An estimator of $\Sigma_{ee,k}$ of (3.8) is obtained by estimating both $c_k$ and $p_{T,k}$. If one is concerned about misspecifying the model for the sampling variances but feels that the direct estimator of the sampling variance is too variable, one can use a smoothed estimator of the sampling variances for estimation and prediction and use the direct estimator of the sampling variance for estimation of the MSE. Estimation of the sampling variances is discussed further in Section 4.

If the multinomial model in (3.8) is not the true covariance matrix of the sampling errors, define the parameter $c_k$ to be

$$
c_k = E[n_k] \text{trace}\{\Sigma_{ee,k}^{(1)} \Gamma_{uu,k}^{(1)}\} (C - 1)^{-1},
$$

(3.9)

where $E[n_k]$ is the expected sample size in province $k$, and $\Sigma_{ee,k}^{(1)}$ and $\Gamma_{uu,k}^{(1)}$ are the $(C - 1) \times (C - 1)$ submatrices of $\Sigma_{ee,k}$ and $\Gamma_{uu,k}$ obtained by omitting the first row and column from $\Sigma_{ee,k}$ and $\Gamma_{uu,k}$, respectively, and $\Gamma_{uu,k}^{(1)}$ is assumed to be nonsingular. When the multinomial model (3.8) does not hold, the parameter $c_k$ defined in (3.9) has an interpretation as an average of generalized design effects (Rao and Scott, 1981).

### 3.4 Minimum Mean Squared Error Linear Predictors

For known $p_{T,ik}$, the minimum mean squared error linear predictor under the model (3.1) is

$$
p_{pred,k} = p_{T,k} + \psi \Gamma_{uu,k} (\psi \Gamma_{uu,k} + \Sigma_{ee,k})^{-1} (\hat{p}_k - p_{T,k}).
$$

(3.10)

The minimum mean squared error predictors in a province sum to one, but the predictors may not remain between zero and one. To obtain predictors that remain in the parameter space,
we consider convex combinations of the form, $\gamma_{ik}\widehat{p}_{ik} + (1 - \gamma_{ik})p_{T,ik}$, where $0 \leq \gamma_{ik} \leq 1$. The weight $\gamma_{ik}$ that minimizes the mean squared error, $E[(\gamma_{ik}\widehat{p}_{ik} + (1 - \gamma_{ik})p_{T,ik} - p_{ik})^2]$, is

$$\gamma_{ik} = \frac{\psi_{T,ik}(1 - p_{T,ik}) + \sigma^2_{e,ik}}{\psi_{T,ik}(1 - p_{T,ik})}, \quad (3.11)$$

where $\sigma^2_{e,ik}$ is the $i$th element of the unconditional covariance matrix $\Sigma_{ee,k}$, and the resulting predictor is

$$\tilde{p}_{ik}(\lambda_o, \psi, \sigma^2_{e,ik}) = p_{T,ik} + \gamma_{ik}(\widehat{p}_{ik} - p_{T,ik}). \quad (3.12)$$

If $\Sigma_{ee,k}$ has the form in (3.8), then the $i$th element of $p_{pred,k}$, defined in (3.10), is equal to the univariate predictor (3.12). Under the multinomial model (3.8), the optimal weight to assign to the direct estimator in the convex combination is

$$\gamma_k = \frac{\psi}{\psi + c_k n_k^{-1}}, \quad (3.13)$$

and the univariate predictor (3.12) is

$$\tilde{p}_{ik}(\lambda_o, \psi, c_k) = p_{T,ik} + \gamma_k(\widehat{p}_{ik} - p_{T,ik}). \quad (3.14)$$

### 3.5 Totals

As defined previously, $\widehat{M}_k$ denotes the direct estimator of the total in province $k$. Let $T_k$ be the expected value of $\widehat{M}_k$. Assume that $\widehat{M}_k$ is independent of $u_{ik}$ and $e_{ik}$ for $i = 1, \ldots, C$. The true total in category $i$ and province $k$ is the random variable

$$M_{ik} = (p_{T,ik} + u_{ik})T_k. \quad (3.15)$$

Because $\sum_{i=1}^{C} u_{ik} = 0$ and $\sum_{i=1}^{C} p_{T,ik} = 1$, the true province total is a fixed parameter;

$$\sum_{i=1}^{C} M_{ik} := M_k = T_k.$$

In contrast, the expected value of the true total in category $i$, aggregated across the provinces, is a function of the $u_{ik}$. The true category total is

$$M_i = \sum_{k=1}^{K} M_{ik} = \sum_{k=1}^{K} (p_{T,ik} + u_{ik})T_k,$$
and the expected value of $M_i$ is

$$E[M_i] = \sum_{k=1}^{K} p_{T,ik} T_k := T_i.$$  

Because

$$\sum_{k=1}^{K} T_k u_{ik} \neq 0,$$

(3.16)

$M_i$ does not equal $T_i$.

The direct estimators of the totals and the proportions are related through $\hat{M}_k = \hat{p}_k \hat{M}_k$, where $\hat{M}_k = (\hat{M}_{1k}, \ldots, \hat{M}_{Ck})'$, and $\hat{M}_{ik} = \hat{p}_{ik} \hat{M}_k$. Therefore, the vector of direct estimators of totals can be written as

$$\hat{M}_k = M_k + a_k,$$

where $a_k = u_k a_k + e_k \hat{M}_k + p_{T,k} a_k$, and $a_k = \hat{M}_k - T_k$. The vector $a_k$ represents the sampling error in the totals. By the assumption that $\hat{M}_k$ is uncorrelated with both $u_k$ and $e_k$, the direct estimators of the totals are unbiased for the true totals;

$$E[\hat{M}_k - M_k] = E[a_k] = 0.$$  

Also, given the assumption that $\hat{M}_k$ is uncorrelated with both $u_k u_k'$ and $e_k e_k'$, the variance of the vector of direct estimators of the totals is

$$V\{\hat{M}_k - M_k\} = V\{a_k\} = \psi \Gamma_{uu,k} V\{a_k\} + \Sigma_{ee,k}(T_k^2 + V\{a_k\}) + p_{T,k} p_{T,k}' V\{a_k\}.$$

Let $\hat{\Sigma}_{aa,k}$ denote a direct estimator of the sampling variance of the totals. If $\hat{\Sigma}_{aa,k}$ is conditionally unbiased for the conditional covariance matrix of $\hat{M}_k - M_k$ and if $E[a_k \mid u_k] = 0$, then $\hat{\Sigma}_{aa,k}$ is also unbiased for the unconditional covariance matrix;

$$E[\hat{\Sigma}_{aa,k}] = E[E[\hat{\Sigma}_{aa,k} \mid u_k]] = V\{\hat{M}_k - M_k\}.$$
3.6 Multiple Two-way Tables

The model described above is a model for a single two-way table defined by a cross-classification of provinces and three digit codes in a single two digit code. In the LFS application, predictors are obtained for multiple two-way tables: 31 two digit codes for four different years. Let $t$ index the two-way tables, where $t = 1, \ldots, D$. Let $\hat{p}_{k,t}$ and $\hat{M}_{k,t}$ denote the direct estimators for province $k$ and two-way table $t$. Let $\lambda_{o,t}$ and $\psi_{t}$ be the model parameters specific to table $t$. If one can justify an assumption that certain model parameters are constant across several two-way tables, then one can obtain more precise estimators by combining information across the two-way tables. For example, in the LFS application, predictors are constructed under an assumption that $\psi_{t} = \psi$ for two-way tables $t$ from different years. In the illustration with LFS data in Section 7, multiple years are not available, but an assumption that $\psi_{t} = \psi$ for two digit codes in a single one digit code is judged to be consistent with the observed data. It is also possible to obtain a more precise estimator of $c_k$ under an assumption that $c_{k,t} = c_k$ for several two-way tables. In the illustration of Section 7, 31 two digit codes are used to estimate a common $c_k$ for each province $k$. Use of several two-way tables to estimate a common parameter for the working variance model is similar to using a generalized variance function for the sampling covariance matrix (Valliant, 1987).

3.7 Summary

The quantities of interest are the true proportions, defined in (3.2), and the true totals, defined in (3.15), in a two-way table defined by three digit occupations and provinces. A model is specified for the direct estimators of the proportions. The interactions in the two-way table from the census are covariates in the nonlinear expectation function. The covariance matrix of the vector of random effects is proportional to a multinomial covariance matrix. In the model, the covariance matrix of the vector of sampling errors is unstructured. The specific case in which the covariance matrix of the vector of sampling errors is proportional to a multinomial covariance matrix is considered. The minimum mean squared error convex combinations of the direct estimators and the expected values depend on the ratios of the variances of the random...
effects to the variances of the sampling errors and are functions of the unknown parameters $\lambda_o$ and $\psi$ as well as the sampling variances. The data available for estimation and prediction include the direct estimators of the proportions, defined in (3.1), the corresponding estimators of the cell totals, the estimated sampling variances, and the Census table.
CHAPTER 4. Procedure

Predictors are desired that are more efficient than the direct estimators and that also preserve the direct estimators of the marginal totals. Optimal predictors under the model (3.1) are functions of the true $p_{T,ik}$, $\psi$ and the sampling variances. The multinomial model of (3.8) is used as a working model for estimation. We obtain estimates of $p_{T,ik}$ and $\psi$ using an iteratively re-weighted least squares procedure. An estimator of $c_k$ defined in (3.9) is derived under a Wishart model for the direct estimators of the sampling variances. Estimators of the optimal predictors are obtained under the working model for the sampling variances. A Beale adjustment is proposed to reduce the bias of the estimator of the weight assigned to the direct estimator in the optimal convex combination. The initial predictors, which are convex combinations of the direct estimators and estimators of $p_{T,ik}$ do not satisfy the benchmarking requirement. A raking operation is used to benchmark the predictors.

4.1 Estimators of Model Parameters

A generalized least squares estimator of $\lambda_o$ minimizes the quadratic form

$$Q(\lambda | \psi, \lambda_o) = \sum_{k=1}^{K} (\hat{p}_{k}^{(1)} - p_{T,k}^{(1)}(\lambda))' \Phi^{-1} (\hat{p}_{k}^{(1)} - p_{T,k}^{(1)}(\lambda)),$$

where $\hat{p}_{k}^{(1)} = (\hat{p}_{2k}, \ldots, \hat{p}_{Ck})'$, $p_{T,k}^{(1)}(\lambda) = (p_{T,2k}(\lambda), \ldots, p_{T,Ck}(\lambda))'$, $p_{T,ik}(\lambda)$ is defined in (3.3), $\Phi = (\Sigma_{uu,k}^{(1)} + \Sigma_{ee,k}^{(1)})$, $\Sigma_{uu,k}^{(1)}$ denotes the $(C-1) \times (C-1)$ submatrix of $\Sigma_{uu,k}$ obtained by omitting the first row and column, and $\Sigma_{ee,k}^{(1)}$ is defined analogously. The first category is omitted to avoid using generalized inverses and to ensure that $\Phi$ is positive definite. The covariance matrix $\Sigma_{uu,k}$ depends on the unknown parameters $\psi$ and $\lambda_o$. In practice, the covariance matrix of the vector of sampling errors $\Sigma_{ee,k}$ is unknown as well. One possibility
is to use the direct estimator of the sampling covariance matrix in estimated generalized least squares. If the direct estimator of the covariance matrix has a large variance, then the resulting estimated generalized least squares estimators may have large variances. An alternative is to use a working covariance structure for estimation. The mathematically convenient choice for a working covariance matrix is the covariance matrix introduced in (3.8), which is proportional to a multinomial covariance matrix.

Define the working covariance matrix by

$$\Sigma_{ee,k,w} = c_k n_k^{-1} \Gamma_{uu,k},$$

where $n_k$ is the realized sample size in province $k$ and $c_k$ is defined in (3.9). The quadratic form based on the working covariance structure simplifies to

$$Q_w(\lambda | \psi, \lambda_o, c) = \sum_{k=1}^{K} [\hat{p}^{(1)}_k - p_{T,k}^{(1)}(\lambda)]' \left[ (c_k/n_k + \psi) \Gamma_{uu,k}^{(1)} \right]^{-1} [\hat{p}^{(1)}_k - p_{T,k}^{(1)}(\lambda)],$$

(4.1)

where $c = (c_1, \ldots, c_K)'$. The derivative of $p_{T,k}^{(1)}(\lambda)$ with respect to $\lambda$ evaluated at $\lambda_o$ is $\Gamma_{uu,k}^{(1)} X_k^{(1)}$, where $X_k^{(1)} = (x_{2k}, \ldots, x_{Ck})'$, and $x_{ik} = (I[i = 2], \ldots, I[i = C])'$. It follows that the derivative of $Q_w(\lambda | \psi, \lambda_o, c)$ with respect to $\lambda$ evaluated at $\lambda_o$ is proportional to

$$S_w(\lambda_o | \psi, c) = K^{-1} \sum_{k=1}^{K} (c_k n_k^{-1} + \psi)^{-1} (X_k^{(1)})' [\hat{p}^{(1)}_k - p_{T,k}^{(1)}(\lambda_o)].$$

(4.2)

We use $S_w(\cdot | \psi, c)$ as a set of estimating equations to define the estimator of $p_{T,ik}$. Because $S_w(\cdot | \psi, c)$ is a function of the unknown $\psi$, we define estimators through an iterative procedure closely related to a general method outlined in Carroll and Ruppert (1988, pg. 69). The iteration uses estimated generalized least squares estimators of $\lambda_o$ and $\psi$. An outline of the procedure is presented here, but details are deferred to the following subsections.

1. Obtain an initial estimator of $\lambda_o$ using the maximum likelihood estimator under a multinomial model. Denote the initial estimator of $\lambda_o$ by $\hat{\lambda}^0$. Details of the procedure used to obtain the initial estimator of $\lambda_o$ are given in Section 4.1.2.
2. Compute an estimator of $c$ using the initial estimator of $\lambda_o$, and denote the estimator by $\hat{c}$ with $k^{th}$ element, $\hat{c}_k$. The estimator of $c_k$ is based on the maximum likelihood estimator under a Wishart distribution and is defined in Section 4.1.1.

3. Using the initial estimator of $\lambda_o$ from step 1 and the estimator of $c_k$ from step 2, compute an initial estimator of $\psi$. Denote the initial estimator of $\psi$ by $\hat{\psi}^0$. The initial estimator of $\psi$ depends on an approximation for the expected value of the squared difference between the $\hat{p}_{ik}$ and the initial estimator of $p_{T,ik}$.

4. Starting with $j = 1$, cycle through the following steps. Details of the procedure used to obtain the the initial estimator of $\psi$ are given in Section 4.1.3.
   
   i. Let $\hat{\lambda}^j$ satisfy $S_w(\lambda | \hat{\psi}^{j-1}, \hat{c}) = 0$.
   
   ii. Update the previous estimator, $\hat{\psi}^{j-1}$, to obtain $\hat{\psi}^j$. The method used to obtain $\hat{\psi}^j$ from $\hat{\lambda}^j$, $\hat{c}_k$, and $\hat{\psi}^{j-1}$ uses approximations for expected mean squares to define an estimated generalized least squares estimator. The weights used to form the new estimator of $\psi$ depend on both the current estimate of $\lambda_o$ and the previous estimate of $\psi$. The details of the procedure for updating an estimator of $\psi$ are described in section 4.1.3.
   
   iii. Set $\epsilon^j = \max\{|\hat{\lambda}^j - \hat{\lambda}^{j-1}|_\infty, |\hat{\psi}^j - \hat{\psi}^{j-1}|\}$, where $|v|_\infty$ is the maximum of the absolute values of the elements of the vector $v$. If $\epsilon^j < \epsilon$, then set $\hat{\psi} = \hat{\psi}^j$ and $\hat{\lambda} = \hat{\lambda}^j$. We use $\epsilon = 10^{-5}$. Otherwise, increase $j$ by 1, and return to step i.

Let $\hat{p}_{T,k} = p_{T,k}(\hat{\lambda})$ denote the estimate of $p_{T,k}$ obtained on convergence, where $p_{T,k}(\lambda) = (p_{T,1k}(\lambda), \ldots, p_{T,Ck}(\lambda))^T$. We refer to $\hat{p}_{T,k}$ as the synthetic estimator. The corresponding vector of synthetic estimators for the totals in province $k$ is $\hat{T}_k = \hat{p}_{T,k} \hat{M}_k$.

### 4.1.1 Estimator of $c_k$ and Sampling Variances

To derive an estimator of $c_k$, suppose $\hat{\Sigma}_{ee,k}^{(1)}$ has the distribution of $[n_k(n_k - 1)]^{-1}A_k$, where $A_k$ has a Wishart distribution with $n_k - 1$ degrees of freedom and $E[n_k \hat{\Sigma}_{ee,k}^{(1)}] = c_k \Gamma_{uu,k}^{(1)}$. 

If $\Gamma_{uu,k}^{(1)}$ is known, the maximum likelihood estimator (MLE) of $c_k$ under the Wishart model is

$$\tilde{c}_k = \text{trace}\{[\Gamma_{uu,k}^{(1)}]^{-1}\tilde{\Sigma}_{ee,k}^{(1)}n_k\}(C - 1)^{-1}.$$  

(See Anderson, 2003, Chapter 7.) To define an estimator of $c_k$, the unknown $\Gamma_{uu,k}$ is replaced by the initial estimator,

$$\hat{\Gamma}_{uu,k}^0 = \text{diag}(\hat{p}_{0T,k}^0) - (\hat{p}_{0T,k}^0)(\hat{p}_{0T,k}^0)' , \quad (4.3)$$

where $\hat{p}_{0T,k}^0$, the initial estimator of $p_{T,k}$, is defined in Section 4.1.2 below.

**Remark 1:** If multiple two-way tables are available and one is willing to assume that $c_{k,t} = c_k$ for each two-way table $t$, one can combine the estimators of $c_k$ to obtain a more precise estimator. Let $t$ index the two-way tables, where $t = 1, \ldots, D$. Assume that the model (3.1) holds for each two-way table with parameters $\psi_t$, $\lambda_{o,t}$, and $\Sigma_{ee,k,t}$ specific to table $t$. Let $\tilde{\Sigma}_{ee,k,t}$ and $n_{k,t}$ denote the direct estimator of the sampling covariance matrix and the realized sample size, respectively, for table $t$. Assume that for each two-way table $t$ in a set $A_c$, $\tilde{\Sigma}_{ee,k,t}$ has the distribution of $[n_{k,t}(n_{k,t} - 1)]^{-1}A_{k,t}$, where $A_{k,t}$ is a Wishart matrix with $n_{k,t} - 1$ degrees of freedom, and $E[n_{k,t}\tilde{\Sigma}_{ee,k,t}] = c_k\Gamma_{uu,k,t}^{(1)}$. Under an assumption that the two-way tables are independent, the log likelihood of $c_k$ is

$$L(c_k) = \sum_{t \in A_c} L_t(c_k),$$

where

$$L_t(c_k) = -\text{trace}\{c_k^{-1}[\Gamma_{uu,k,t}^{(1)}]^{-1}\tilde{\Sigma}_{ee,k,t}n_{k,t}(n_{k,t} - 1)\} - (n_{k,t} - 1)(C_t - 1)\log(c_k),$$

and $C_t$ is the number of categories in table $t$. The first derivative of $L_t(c_k)$ with respect to $c_k$ is

$$L_t(c_k) = c_k^{-2}\text{trace}\{[\Gamma_{uu,k,t}^{(1)}]^{-1}\tilde{\Sigma}_{ee,k,t}n_{k,t}(n_{k,t} - 1)\} - (n_{k,t} - 1)(C_t - 1)c_k^{-1},$$

and the maximum likelihood estimator of $c_k$ is

$$\tilde{c}_{k,D} = \frac{\sum_{t \in A_c} \text{trace}\{[\Gamma_{uu,k,t}^{(1)}]^{-1}\tilde{\Sigma}_{ee,k,t}n_{k,t}(n_{k,t} - 1)\}}{\sum_{t \in A_c} (n_{k,t} - 1)(C_t - 1)}. \quad (4.4)$$
Because the true $\Gamma_{uu,k,t}$ is unknown, we replace $\Gamma_{uu,k,t}$ in (4.4) with the initial estimator,

$$
\hat{\Gamma}^0_{uu,k,t} = \text{diag}(\hat{p}^0_{T,k,t}) - \hat{p}^0_{T,k,t}(\hat{p}^0_{T,k,t})'
$$

where $\hat{p}^0_{T,k,t}$ is the initial estimator of the expected value of the proportion for table $t$. The estimator of $c_k$ based on $D$ two-way tables is defined

$$
\hat{c}_k = \frac{\sum_{t \in A_c} \text{trace}\{[\hat{\Gamma}^{0(1)}_{uu,k,t}]^{-1}\hat{\Sigma}^{(1)}_{ee,k,t}n_{k,t}(n_{k,t} - 1)\}}{\sum_{t \in A_c} (n_{k,t} - 1)(C_t - 1)}.
$$

(4.5)

\[\square\]

Remark 2: Because the estimator of $c_k$ is a function of the estimator of $\hat{p}_{T,k}$, it is possible to update the estimator of $c_k$ after each step of the iteration. We keep the estimator of $c_k$ fixed at the initial value to simplify the computations. In our simulations, updating the estimator of $c_k$ had little effect on the efficiencies of the predictors.

Remark 3: In the application, if a direct estimator of a cell total is zero, then the corresponding direct estimator of the variance is also zero. When a direct estimator is zero, we modify the sampling covariance matrix in a way that ensures that the diagonal elements of the resulting covariance matrix are positive and that rows and columns sum to zero. As defined previously, $\hat{\Sigma}_{aa,k}$ denotes the direct estimator of the variance of $a_k$, the vector of sampling errors in the totals. If $\hat{M}_{ik} = 0$, then the $i^{th}$ row and column of $\hat{\Sigma}_{aa,k}$ are vectors of zeros. Under the working model, an estimator of the sampling variance of the total in cell $(i,k)$ is

$$
\hat{\sigma}_{aa,ik,md} = \hat{p}^0_{T,ik}(1 - \hat{p}^0_{T,ik})\hat{c}_kn_k^{-1}\hat{M}_k^2 + \hat{V}\{\hat{M}_k\}(\hat{p}^0_{T,ik})^2.
$$

(4.6)

If $\hat{M}_{ik} = 0$, then set the $i^{th}$ diagonal element of $\hat{\Sigma}_{aa,k}$ equal to $\hat{\sigma}_{aa,ik,md}$, and let $\hat{\Sigma}_{aa,k,md}$ denote the resulting covariance matrix. Because $\hat{\Sigma}_{aa,k,md}$ is a sum of two nonnegative definite matrices, $\hat{\Sigma}_{aa,k,md}$ is nonnegative definite. By a Taylor expansion, a direct estimator of the sampling variance of the vector of proportions for province $k$ is obtained from the direct estimator of the sampling variance of the totals through the operation,

$$
\hat{\Sigma}_{ee,k} = \hat{L}_k\hat{\Sigma}_{aa,k}\hat{L}_k',
$$

where $\hat{L}_k$ is the matrix of a linear approximation given in Appendix 1. If a direct estimator for a cell is zero, then define the modified direct estimator of the sampling variance for the
proportions in province $k$ by

$$
\tilde{\Sigma}_{ee,k,md} = \hat{L}_k \hat{\Sigma}_{aa,k,md} \hat{L}'_k.
$$

(4.7)

Remark 4: If the working model for the sampling covariance matrix is not the true covariance matrix, then inference based on the working covariance matrix can be biased. To protect against possible bias due to misspecification of the working covariance model, we use the modified versions of the direct estimators of the sampling covariance matrices defined in (4.6) and (4.7) for MSE estimation. We also use the modified direct estimators of the covariance matrices in the estimator of $\psi$, as described in Section 4.1.3 below.

4.1.2 Estimator of $\lambda_o$

The initial estimator of $\lambda_o$ is the maximum likelihood estimator under the assumption that the vector of direct estimators of totals in a province is a multinomial random vector with probability mass function

$$
P(\hat{M}_1 = x_{1k}, \ldots, \hat{M}_C = x_{Ck}) = \frac{\hat{M}_k!}{\prod_{i=1}^C x_{ik}!} \prod_{i=1}^C p_{T,k}^{x_{ik}},
$$

and the assumption that province estimators are independent. Let $\hat{p}_{T,k}^0 = p_{T,k}(\hat{\lambda}^0)$, where $\hat{\lambda}^0$ is the initial estimator of $\lambda$. If the estimator of $\theta_o$ is restricted to equal 1, then the initial estimator of $p_{T,k}$ is the SPREE estimator. Given the direct estimators of the marginal totals, the covariance matrix underlying the initial estimators is $\hat{M}_k^{-1} \Gamma_{uu,k}$.

The estimator of $\lambda_o$ at iteration $j$ (for $j \geq 1$), denoted $\hat{\lambda}^j$, satisfies

$$
S_w(\hat{\lambda}^j \mid \hat{\psi}^{j-1}, \hat{c}) = 0,
$$

(4.9)

where $\hat{p}_{T,k}^j = p_{T,k}(\hat{\lambda}^j)$, and $\hat{\psi}^{j-1}$ denotes the current estimate of $\psi$. (The procedure used to obtain the initial estimator of $\psi$ is described in Section 4.1.3.) Let $\tilde{\hat{p}}_{T,k}^j = p_{T,k}(\hat{\lambda}^j)$, and let $\hat{\Gamma}_{uu}^j = \text{diag}(\tilde{\hat{p}}_{T,k}^j) - \tilde{\hat{p}}_{T,k}^j (\tilde{\hat{p}}_{T,k}^j)'$. A linear approximation for $\hat{\lambda}^j$ is used to derive an estimator
of \( \psi \) and an estimator of the MSE. A justification for the linear approximation is given in Appendix 2.

Remark: If the covariance matrix of \( \hat{p}_k \) differs from \( \hat{M}_k^{-1}\Gamma_{uu,k} \), then the initial estimators are inefficient. Under the working model for the sampling variances, the covariance matrix of the vector of direct estimators of proportions is

\[
(c_kn_k^{-1} + \psi)\Gamma_{uu,k},
\]

where \( n_k \) is the realized sample size in province \( k \). If the working covariance matrix (4.10) is a better approximation to the true covariance matrix of \( \hat{p}_k \) than the multinomial covariance matrix \( \hat{M}_k^{-1}\Gamma_{uu,k} \), then an estimator of \( \lambda_o \) based on the working covariance matrix in (4.10) is more efficient than the initial estimator, \( \hat{\lambda}^{(0)} \).

4.1.3 Estimator of \( \psi \)

We update the estimator of \( \psi \) using approximations for expected mean squares. By a Taylor expansion,

\[
\hat{\lambda}^j - \lambda_o \approx H_{xx}^{-1} \sum_{k=1}^{K} (c_kn_k^{-1} + \psi)^{-1}(X_k^{(1)})'(\hat{p}_k^{(1)} - \hat{p}_T^{(1)}),
\]

(4.11)

where

\[
H_{xx} = \sum_{k=1}^{K} (X_k^{(1)})'(c_kn_k^{-1} + \psi)^{-1}\Gamma_{uu,k}^{(1)}X_k^{(1)},
\]

(4.12)

and \( X_k^{(1)} \) is a \( C \times C \) matrix with \( x_{ik} = (I[i = 2], \ldots, I[i = C], (\alpha\beta)_{ik}^{en}) \) in row \( i \). A justification for the linear approximation (4.11) is given in Appendix 2. It follows from (4.11) that

\[
\hat{p}_T^j - p_T \approx W_p'(\hat{p} - p_T),
\]

(4.13)

where \( W_p' = \Gamma_{uu}XX_{xx}^{-1}X'D_w^{-1} \), \( X = (X_1^T, \ldots, X_K^T)^T, X_k \) is a \( C \times C \) matrix obtained by adding a row of zeros to the first row of \( X_k^{(1)} \), and \( D_w = \text{diag}((c_1n_1^{-1} + \psi), \ldots, (c_Kn_K^{-1} + \psi)) \otimes I_C \). The notation \( I_d \) is used throughout to denote a \( d \times d \) identity matrix.
By the linear approximation (4.13), for \( j \geq 1 \),

\[
V\{\hat{\mathbf{p}} - \hat{\mathbf{p}}^j_T\} \approx \Sigma_a + \psi \Sigma_b,
\]

where

\[
\Sigma_a = (I_{CK} - W_p)' \Sigma_{ee} (I_{CK} - W_p),
\]

(4.14)

and

\[
\Sigma_b = (I_{CK} - W_p)' \Gamma_{uu} (I_{CK} - W_p).
\]

(4.15)

Letting \( \sigma_a \) and \( \sigma_b \) be vectors containing the diagonal elements of \( \Sigma_a \) and \( \Sigma_b \), respectively, an approximation for the expected value of the squared difference between the direct estimator and the estimator of \( \mathbf{p}_T \) is

\[
E[\{\hat{\mathbf{p}} - \hat{\mathbf{p}}^j_T\}^*]^2 \approx \sigma_a + \psi \sigma_b,
\]

where the notation \( \{v\}^* \) denotes the vector containing the squares of the elements of the vector \( v \). Under an assumption of normality,

\[
V[\{\hat{\mathbf{p}} - \hat{\mathbf{p}}^j_T\}^*] \approx 2\{\Sigma_a + \psi \Sigma_b\}^*^2,
\]

where the notation \( \{A\}^*^2 \) denotes the matrix containing the squares of the elements of the matrix \( A \). Let \( \hat{\Sigma}_a^j \) and \( \hat{\Sigma}_b^j \) denote estimators of \( \Sigma_a \) and \( \Sigma_b \) obtained by replacing the true \( \lambda_o \) and \( \psi \) by the estimators, \( \hat{\lambda}^j \) and \( \hat{\psi}^{j-1} \), and by replacing \( \Sigma_{ee} \) with the working covariance matrix, \( \hat{\Sigma}_{ee,k,w}^j = \hat{c}_k n_k^{-1} \hat{\Gamma}_{uu,k}^j \). Let \( \hat{\Sigma}_{a,dir}^j \) be the estimate of \( \Sigma_a \) obtained by replacing the true \( \lambda_o \) and \( \psi \) with the estimators \( \hat{\lambda}^j \) and \( \hat{\psi}^{j-1} \) and the unknown \( \Sigma_{ee} \) with the modified direct estimator defined in (4.7). Define the estimated generalized least squares estimator,

\[
\tilde{\psi}^j = \frac{\langle \hat{\sigma}_j^i \hat{\sigma}_b^{-1} \hat{V}_{\hat{\psi}^j-1} \{\hat{\mathbf{p}} - \hat{\mathbf{p}}^j_T\}^2 - \hat{\sigma}^i_{a,dir} \rangle}{\langle \hat{\sigma}_j^i \hat{\sigma}_b^{-1} \hat{V}_{\hat{\psi}^j-1} \hat{\sigma}_b^i \rangle},
\]

where

\[
\hat{V}_{\hat{\psi}^j-1} = 2\{\hat{\Sigma}_a^j + \hat{\psi}^{j-1} \hat{\Sigma}_b^j\}^*^2.
\]
and $\hat{\sigma}_a^j$, $\hat{\sigma}_b^j$, and $\hat{\sigma}_{a,dir}^j$ are the vectors of diagonal elements of $\hat{\Sigma}_a^j$, $\hat{\Sigma}_b^j$, and $\hat{\Sigma}_{a,dir}^j$, respectively.

The estimated generalized least squares estimator, $\hat{\psi}^j$, may be negative. Under an assumption of normality, the standard deviation of an estimator of a variance is proportional to the variance. If $\psi = 0$, then an estimator of the variance of a generalized least squares estimator of $\psi$ is

$$\hat{V}_2^j = [(\hat{\sigma}_b^j)'(2\{\hat{\Sigma}_a^j\})^{-2}(\hat{\sigma}_b^j)]^{-1}. \quad (4.16)$$

The estimator $\hat{V}_2^j$ is constructed under the assumption that the working model for the variance of $e_k$ is correct. Let $\xi^j = 0.5[\hat{V}_2^j]^{0.5}$, and define an updated estimator of $\psi$,

$$\tilde{\psi}^j = \max\{\hat{\psi}^j, \xi^j\}.$$

The linear approximation for the initial estimator of $\lambda_o$ (Appendix 3) differs from the linear approximation for $\hat{\lambda}_j$ for $j \geq 1$ (Appendix 2). As a consequence, the expected mean squares used to obtain the initial estimator of $\psi$ differ from $\Sigma_a$ and $\Sigma_b$. As explained in Section 4.1.2, the initial estimator of $\lambda_o$ is the maximum likelihood estimator under an assumption that the vectors of direct estimators for the provinces are independent multinomial random vectors with multinomial sample sizes $\{\hat{M}_1, \ldots, \hat{M}_K\}$. By a Taylor expansion of the multinomial score function,

$$\tilde{\mathbf{P}}_T - \mathbf{p}_T \approx W_{p,u}'D_{prov}\mathbf{u} + W_{p,d}'\mathbf{a}, \quad (4.17)$$

where $\tilde{\mathbf{p}}_T = (p_{T,1}(\hat{\lambda})', \ldots, p_{T,K}(\hat{\lambda})')'$, $\mathbf{a} = (a_1', \ldots, a_K')'$, $a_k = \hat{M}_k - M_k$,

$$W_{p,u}' = \Gamma_{uu}X(X'D_{prov}\Gamma_{uu}X)^{-1}X',$$

$$W_{p,d}' = W_{p,u}'[I_{CK} - \text{diag}(\mathbf{p}_T)(I_K \otimes J_C) \otimes J_C],$$

$D_{prov} = \text{diag}((T_1, \ldots, T_K)) \otimes I_C$, $\text{diag}(\mathbf{p}_T)$ is a diagonal matrix with the vector $\mathbf{p}_T$ on the diagonal, and $J_C$ is a $C$-dimensional column vector of ones. By the linear approximation (4.17),

$$V\{\tilde{\mathbf{p}} - \tilde{\mathbf{p}}_T^0\} \approx \Sigma_a^0 + \psi \Sigma_b^0, \quad (4.18)$$
\[
\begin{align*}
\Sigma_0^a &= W_d' \Sigma_{dd} W_d, \\
\Sigma_0^b &= (I_{CK} - W'_{p,u} D_{prov}) \Gamma_{uu}(I_{CK} - W_{p,u} D_{prov})', \\
W_d' &= [I_{CK}, -W'_{p,d}], \Sigma_{dd} \text{ is the covariance matrix of } (e', a')'. 
\end{align*}
\]

By the assumptions about the sampling errors described in Section 3.5,

\[
\Sigma_{dd} = \begin{pmatrix}
\Sigma_{ee} & \Sigma_{ee} D_{prov} \\
D_{prov} \Sigma_{ee} & V\{a\}
\end{pmatrix},
\]

where, and \( V\{a_k\} \) is the variance of \( a_k \). Let \( \hat{\Sigma}_0^b \) be the estimate of \( \Sigma_0^b \) obtained by evaluating \( \Sigma_0^b \) at the initial estimate \( \hat{\lambda}^0 \). Let \( \hat{\Sigma}_0^{dd} \) be an estimator of \( \Sigma_{dd} \) based on the working model for the sampling variances, and let \( \hat{\Sigma}_{dd,md} \) be an estimate based on the modified direct estimators of the variances. Specifically, the estimator of \( \Sigma_{dd} \) under the working model is

\[
\hat{\Sigma}_0^{dd} = \begin{pmatrix}
\hat{\Sigma}_{ee,w} & \hat{\Sigma}_{ee,w} D_{prov} \\
D_{prov} \hat{\Sigma}_{ee,w} & \hat{\Sigma}_{aa,md}
\end{pmatrix},
\]

where \( \hat{\Sigma}_{ee,w} \) is a block diagonal matrix with \( \hat{c}_k n_k^{-1} \hat{\Gamma}_{uu,k} \) as the \( k^{th} \) block, \( \hat{\Gamma}_{uu,k} = \text{diag}(\hat{p}_{T,k}^0) - (\hat{p}_{T,k}^0)(\hat{p}_{T,k}^0)'^T, D_{prov} = \text{diag}(\hat{M}_1, \ldots, \hat{M}_k') \otimes I_C \), and \( \hat{V}^0\{a\} \) is a block diagonal matrix with \( k^{th} \) block,

\[
\hat{V}^0\{a_k\} = \hat{c}_k n_k^{-1} \hat{\Gamma}_{uu,k} \hat{M}_k^2 + \hat{p}_{T,k}^0 \hat{p}_{T,k}^T \hat{V}\{\hat{M}_k\}.
\]

The estimator of \( \Sigma_{dd} \) constructed with the modified direct estimators of the variances is

\[
\hat{\Sigma}_{dd,md} = \hat{\Sigma}_0^{dd} = \begin{pmatrix}
\hat{\Sigma}_{ee,md} & \hat{\Sigma}_{ee,md} D_{prov} \\
D_{prov} \hat{\Sigma}_{ee,md} & \hat{\Sigma}_{aa,md}
\end{pmatrix},
\]

where \( \hat{\Sigma}_{ee,md} \) is a block diagonal matrix with \( \hat{\Sigma}_{ee,k,md} \) of (4.7) as the \( k^{th} \) block, and \( \hat{\Sigma}_{aa,md} \) is the modified direct estimator of the sampling variance for the totals with elements defined in (4.6).

Let \( \hat{\Sigma}_a^0 \) be the estimate of \( \Sigma_a^0 \) based on \( \hat{\Sigma}_{dd,k}^0 \), and let \( \hat{\Sigma}_{a,dir}^0 \) be the estimate of \( \Sigma_a^0 \) obtained from \( \hat{\Sigma}_{dd,k,md}^0 \). Defining \( \hat{\sigma}_b^0, \hat{\sigma}_a^0, \) and \( \hat{\sigma}_{a,dir}^0 \) to be the diagonal elements of \( \hat{\Sigma}_b^0, \hat{\Sigma}_a^0, \) and \( \hat{\Sigma}_{a,dir}^0 \),
respectively, the initial estimate of $\psi$ is
\[
\tilde{\psi}^0 = \max\{\tilde{\psi}_0^0, \xi^0\},
\] (4.19)
where
\[
\tilde{\psi}_0^0 = \left(\hat{\sigma}_b^0\right)' \left[2\left(\hat{\Sigma}_a^0 + \xi^0\hat{\Sigma}_b^0\right)^{1/2}\right]^{-1} \left[\{\hat{p}_T - \hat{p}_T^0\}^2 - \hat{\sigma}_{a,dir}^0\right],
\] (4.20)
\[
\xi^0 = 0.5[\hat{V}_2\{\tilde{\psi}^0\}]^{0.5}, \text{ and } \hat{V}_2\{\tilde{\psi}^0\} = \left(\hat{\sigma}_b^0\right)' \left[2\left(\hat{\Sigma}_a^0 + \xi^0\hat{\Sigma}_b^0\right)^{1/2}\right]^{-1} \left[\{\hat{p}_T - \hat{p}_T^0\}^2 - \hat{\sigma}_{a,dir}^0\right]^{-1}.
\]

Remark 1: Both the estimator of the working covariance model and the direct estimator of the sampling covariance matrix are used in the estimator of $\psi$. The estimator of the working model is used to estimate $2\{\Sigma_a + \psi\Sigma_b\}^{1/2}$ to construct an estimated covariance matrix for the estimated generalized least squares estimator. The direct estimator is used to define the responses, $\lfloor\{\hat{p}_T - \hat{p}_T^0\}^2 - \hat{\sigma}_{a,dir}^0\rfloor$ and $\lfloor\{\hat{p}_T - \hat{p}_T^j\}^2 - \hat{\sigma}_{a,dir}^j\rfloor$. The estimator of the working model is used for the covariance matrix for two main reasons. One, the estimator of the working model has a smaller variance than the direct estimator of the sampling covariance matrix. Two, the direct estimator of the covariance matrix is highly correlated with the response, which introduces a bias in the estimator of $\psi$. The direct estimator is used to form the response because use of the working model will lead to a bias in the estimator of $\psi$ if the working model for the sampling variances is misspecified.

If several two-way tables are available and one is willing to assume that the two-way tables were generated from a common value of $\psi$, then one can improve the estimator of $\psi$. As before, let $t = 1, \ldots, D$ index the different two-way tables. Suppose that $\psi_t = \psi$ for each table $t$ in a set $A_\psi$. Let $\hat{\Sigma}_{a,t}^j$, $\hat{\Sigma}_{b,t}^j$, and $\hat{\Sigma}_{a,dir,t}^j$ denote versions of $\hat{\Sigma}_a^j$, $\hat{\Sigma}_b^j$, and $\hat{\Sigma}_{a,dir}^j$ specific to table $t$. Let $\tilde{\psi}_D^{j-1}$ denote an estimator of the common value of $\psi$ obtained in iteration $j - 1$. Define $\hat{V}_{\tilde{\psi}_t}^{j-1} = 2\{\hat{\Sigma}_{a,t}^j + \tilde{\psi}_D^{j-1}\hat{\Sigma}_{b,t}^j\}^{1/2}$. To estimate the common value of $\psi$, at the $j^{th}$ iteration ($j \geq 1$), replace $\tilde{\psi}_D^{j-1}$ by
\[
\tilde{\psi}_D^j = \frac{\sum_{t \in A_\psi} (\hat{\sigma}_{b,t}^j)' \hat{V}_{\tilde{\psi}_t}^{j-1} \lfloor\{\hat{p}_t - \hat{p}_T^j\}^2 - \hat{\sigma}_{a,dir,t}^j\rfloor}{\sum_{t \in A_\psi} (\hat{\sigma}_{b,t}^j)' \hat{V}_{\tilde{\psi}_t}^{j-1} (\hat{\sigma}_{b,t}^j)},
\] (4.21)
and replace $\xi_j$ by $\xi_D = 0.5|\hat{V}_{2,D}^j|^{0.5}$, where
\[
\hat{V}_{2,D}^j = \left\{ \sum_{t \in A_\psi} (\hat{\sigma}_{b,t}^j)^2 \left( 2\{\hat{\Sigma}_{a,t}^j\}^2 \right)^{-1} (\hat{\sigma}_{b,t}^j)^{-1} \right\}. 
\]
Then, the updated estimator of $\psi$ in the $j^{th}$ step is $\hat{\psi}_D^j = \max\{\tilde{\psi}_D^j, \xi_D^j\}$. Let the final estimator of the common value of $\psi$ be
\[
\hat{\psi} = \max\{\tilde{\psi}_D, \xi_D\}, 
\]
where $\tilde{\psi}_D$ and $\xi_D$ are the values of $\hat{\psi}_D^j$ and $\xi_D^j$ obtained on the last iteration. An initial estimator of $\psi$ based on multiple two-way tables is defined analogously using the linear approximation in (4.17) to define $\hat{\Sigma}_{a,t}^0$, $\hat{\Sigma}_{b,t}^0$, and $\hat{\Sigma}_{a,dir,t}^0$.

**Remark 2:** To assess the validity of an assumption that the value of $\psi$ is constant across the two-way tables, a test of the null hypotheses, $H_0 : \psi_t = \psi$ for $t \in A_\psi$ is desired. Define the test statistic,
\[
S_\psi = \sum_{t \in A_\psi} \frac{\left( \hat{\sigma}_{b,t}^j \hat{V}_{\psi,\hat{\psi}_t}^{-1} (\{\hat{p}_t - \hat{p}_{T,t}\}^2 - \hat{\sigma}_{a,dir,t} - \hat{\sigma}_{b,t} \hat{\psi}_D^j) \right)^2}{\hat{\sigma}_{b,t}^j \hat{V}_{\psi,\hat{\psi}_t}^{-1} 2\{\hat{\Sigma}_{a,t,dir} + \hat{\psi}_D^j \hat{\Sigma}_{b,t}\}^2 \hat{V}_{\psi,\hat{\psi}_t}^{-1} \hat{\sigma}_{b,t}}, 
\]
where $\hat{\Sigma}_{a,t}^0$, $\hat{\Sigma}_{b,t}^0$, and $\hat{\Sigma}_{a,dir,t}^0$ are evaluated at the final estimators of $\lambda_0$ and $\psi$. We compare the test statistic $S_\psi$ to a $\chi^2$ distribution with $|A_\psi| - 1$ degrees of freedom, where $|A_\psi|$ is the number of two-way tables used to estimate $\psi$. We also use $S_\psi$ to construct bootstrap tests of the null hypothesis of a common value of $\psi$.

**Remark 3:** It is not necessary to use the same two-way tables to estimate $c_k$ and $\psi$. For example, in the application, the two-way tables from 31 two-digit codes are used to estimate $c_k$. Because the data are not consistent with an assumption that all of the two-way tables share a common value of $\psi$, a smaller subset of two-way tables is used to estimate $\psi$.

**Remark 4:** The iteration is used to improve the precisions of the estimators. In simulation studies not discussed in this thesis, the estimator of the random component based on one step of the iteration is more efficient than the initial estimator, but iterating to convergence does lead to further gains. Additional problems with the iteration is that we do not have an overall objective function or a proof that the iteration will converge. In future work, we would consider terminating the iteration after one step.
Remark 5: In the estimators of $\psi$ in (4.21) and (4.20), the estimator of the full covariance matrix, $2\{\Sigma_a + \psi\Sigma_b\}^{\ast}2$ is used as the weight matrix. The covariance matrix is of dimension $CK \times CK$ and can be close to singular. An alternative is to use the diagonal elements of the weight matrix. In our simulations, the use of the full covariance matrix leads to a more efficient estimator of $\psi$. It may be preferable to use a diagonal weight matrix to avoid unstable estimators of the inverse.

4.2 Predictors of True Proportions and Totals

We desire predictors of the true proportions that have small mean squared errors and also preserve the direct estimators of the marginal totals. We define an initial predictor that estimates the minimum mean squared error convex combination of $\hat{p}_{ik}$ and $p_{T,ik}$ under the assumption that the sampling covariance matrix has the form $\Sigma_{ee,k} = c_k n_k^{-1} \Gamma_{uu,k}$. We use a Beale ratio estimator of the weight assigned to the direct estimator in the optimal convex combination. We then benchmark the initial predictors to the direct estimators of the marginal totals. In this section, we first define predictors using a standard estimator of the optimal convex combination. We then define predictors based on the Beale ratio estimator. Although we use the working covariance matrix to construct the predictors, we use the modified direct estimator of the covariance matrix defined in (4.7) to compute the MSE to protect against bias due to misspecification of the working covariance model.

4.2.1 Initial Predictors of Proportions

To ensure that the predictors remain in the parameter space, we consider a predictor of the form $\gamma_{ik}\hat{p}_{ik} + (1 - \gamma_{ik})p_{T,ik}$, where $0 \leq \gamma_{ik} \leq 1$. The predictor that minimizes the mean squared error, $E[(\gamma_{ik}\hat{p}_{ik} + (1 - \gamma_{ik})p_{T,ik} - p_{ik})^2]$, is

$$\hat{p}_{ik}(\lambda_o, \psi, \sigma^2_{e,ik}) = p_{T,ik} + \gamma_{ik}(\hat{p}_{ik} - p_{T,ik}),$$

defined in (3.12), where $\gamma_{ik} = \psi p_{T,ik}(1 - p_{T,ik})(\psi p_{T,ik}(1 - p_{T,ik}) + \sigma^2_{e,ik})^{-1}$ is defined in (3.11). Isaki et al. (2000) compare univariate predictors calculated with the diagonal elements of an
estimated covariance matrix to the empirical BLUP. In their simulation study, the univariate predictors have smaller MSE’s than predictors based on the full estimated covariance matrix.

The predictor (3.12) depends on unknown parameters. To calculate the predictor, we use the EGLS estimator of $\psi$ and the synthetic estimator of $p_{T,ik}$. In practice, the sampling variances are also unknown. If the variances of the direct estimators of the sampling variances are large, then predictors calculated under the working model for the sampling variances may be more efficient than predictors calculated with the direct estimators of the sampling variances. We construct predictors under the working model for the sampling variances. An initial predictor for cell $(i,k)$ as

$$\hat{P}_{\text{pred},ik} = \hat{p}_{T,ik} + \hat{\gamma}_k(\hat{p}_{ik} - \hat{p}_{T,ik}),$$

(4.24)

where

$$\hat{\gamma}_k = \frac{\hat{\psi}}{\hat{\psi} + \hat{c}_kn_k^{-1}}.$$  

(4.25)

By construction, the initial predictor in (4.24) is between the direct estimator and the synthetic estimator, and the initial predictors in a single province sum to 1.

### 4.2.2 Beale Predictors of Proportions

The weight $\hat{\gamma}_k$ used to construct the initial predictor is a ratio of two estimators. The bias of $\hat{\gamma}_k$ affects both the efficiency of the predictor and the accuracy of the MSE estimator. We use a Beale ratio estimator in an attempt to reduce the MSE of the estimator of $\gamma_k$.

To motivate the Beale estimator of $\gamma_k$, we first suppose that the sampling variance $\sigma^2_{e,ik}$ is known. Let $\gamma_{ik}$ be as defined in (3.11), and define the estimator

$$\hat{\gamma}_{ik,o} = \frac{\hat{\psi}\hat{p}_{T,ik}(1 - \hat{p}_{T,ik})}{\hat{\psi}\hat{p}_{T,ik}(1 - \hat{p}_{T,ik}) + \sigma^2_{e,ik}}.$$  

The estimator $\hat{\gamma}_{ik,o}$ is a ratio of two estimators, and the bias of $\hat{\gamma}_{ik,o}$ can be important (Tin, 1965). A Beale ratio estimator (Tin, 1965) of $\gamma_{ik,o}$ is

$$\hat{\gamma}_{ik,o,B} = \frac{\hat{\psi}\hat{p}_{T,ik}(1 - \hat{p}_{T,ik})(\hat{\psi}\hat{p}_{T,ik}(1 - \hat{p}_{T,ik}) + \sigma^2_{e,ik}) + V\{\hat{\psi}\hat{p}_{T,ik}(1 - \hat{p}_{T,ik})\}}{(\hat{\psi}\hat{p}_{T,ik}(1 - \hat{p}_{T,ik}) + \sigma^2_{e,ik})^2 + V\{\hat{\psi}\hat{p}_{T,ik}(1 - \hat{p}_{T,ik})\}}.$$  

(4.26)
Under regularity conditions, the order of the bias of the Beale estimator for \( \hat{\gamma}_{ik,o} \) is smaller than the order of the bias of \( \tilde{\gamma}_{ik,o} \).

In the LFS application, the sampling variances are estimated, and predictors are calculated under the working model. Under the working model, \( \sigma^2_{e,ik} = n^{-1}_k c_k p T,ik (1 - p T,ik) \), and

\[
\tilde{\gamma}_{ik} = \gamma_k = \frac{\psi}{\psi + c_k n_k^{-1}}. 
\]

Under the working model, the Beale estimator of the true \( \gamma_k \) is

\[
\gamma_{k,B1} = \frac{\hat{\psi}(\hat{\psi} + \hat{c}_k n_k^{-1}) + C\{\hat{\psi}, \hat{c}_k n_k^{-1} + \hat{\psi}\}}{(\hat{\psi} + \hat{c}_k n_k^{-1})^2 + V\{\hat{\psi} + \hat{c}_k n_k^{-1}\}}, 
\]

and under an assumption that \( \hat{\psi} \) and \( \hat{c}_k \) are uncorrelated, the Beale estimator is

\[
\gamma_{k,B} = \frac{\hat{\psi}(\hat{\psi} + \hat{c}_k n_k^{-1}) + V\{\hat{\psi}\}}{(\hat{\psi} + \hat{c}_k n_k^{-1})^2 + V\{\hat{\psi}\} + V\{\hat{c}_k n_k^{-1}\}}. 
\]

Computation of the Beale estimators requires an estimator of the covariance matrix of \( (\hat{\psi}, \hat{c})' \). The variances of the variance estimators depend on fourth moments. Given the limited data, we use the fourth moments of the normal distribution to derive estimators of the variances of the variance estimators.

We denote an estimator of the variance of an estimator of \( \psi \) based on a single two-way table by \( \hat{V}_1(\hat{\psi}) \) and define the variance estimator by

\[
\hat{V}_1(\hat{\psi}) = (\hat{\sigma} b' \hat{V}_1^{-1} \hat{\sigma}_b)^{-2} (\hat{\sigma} b' \hat{V}_1^{-1}[2\{\hat{\Sigma}_{a,dir} + \hat{\psi}\hat{\Sigma}_b\}^{*2}]\hat{V}_1^{-1} \hat{\sigma}_b). \tag{4.27}
\]

If multiple two-way tables are used to estimate \( \psi \), then the estimator of the variance of \( \hat{\psi}_D \) of (4.22) is

\[
\hat{V}(\hat{\psi}) = \frac{\sum_{t \in A_{\psi}} \hat{\sigma}_{b,t}' \hat{V}_{\hat{\psi},t}^{-1}[2\{\hat{\Sigma}_{a,dir,t} + \hat{\psi}\hat{\Sigma}_{b,t}\}^{*2}]\hat{V}_{\hat{\psi},t}^{-1} \hat{\sigma}_{b,t}}{(\sum_{t \in A_{\psi}} \hat{\sigma}_{b,t}' \hat{V}_{\hat{\psi},t}^{-1} \hat{\sigma}_{b,t})^2}. \tag{4.28}
\]

To derive an estimator of the variance of \( \hat{c}_k \), suppose the working model for the sampling covariance matrix holds and that \( \Gamma_{uu,k} \) is known. If a \( (C - 1) \times (C - 1) \) submatrix of the direct estimator of the covariance matrix of the vector of sampling errors for province \( k \) has a Wishart
distribution with \( n_k \) degrees of freedom, then the variance of the maximum likelihood estimator of \( c_k \) is

\[
V\{\hat{c}_{k,mle}\} = \frac{2c_k^2}{(C - 1)(n_k - 1)}.
\]

A variance estimator that relies only mildly on the assumptions underlying \( V\{\hat{c}_{k,mle}\} \) can be obtained if multiple two-way tables are used to estimate \( c_k \). Let \( \hat{c}_{k,t} \) denote the estimator of \( c_{k,t} \) for two-way table \( t \). The estimator (4.31) is a generalized least squares estimator under the model,

\[
\hat{c}_{k,t} \sim \left( c_k, \frac{\sigma^2_{c,k}}{(C_t - 1)(n_{k,t} - 1)} \right)
\]

where \( \sigma^2_{c,k} \) is a constant assumed to be common to all two way tables in province \( k \). An estimator of the variance of the generalized least squares estimator, \( \hat{c}_k \), is

\[
\hat{V}(\hat{c}_k) = \frac{\hat{\sigma}^2_{c,k}}{\sum_{t \in A_c} (C_t - 1)(n_{k,t} - 1)}.
\]

where

\[
\hat{\sigma}^2_{c,k} = (|A_c| - 1)^{-1} \sum_{t \in A_c} (C_t - 1)(n_{k,t} - 1)(\hat{c}_{k,t} - \hat{c}_k)^2,
\]

\( A_c \), defined previously, is the set of two-way tables used to estimate \( c_k \), and \(|A_c|\) is the number of two-way tables in the set \( A_c \). If a single two-way table is used to estimate \( c_k \), then a simple variance estimator derived under the Wishart model is

\[
\hat{V}_1(\hat{c}_k) = \frac{2c_k^2}{(n_{k,t} - 1)(C_t - 1)}.
\]

We use the estimators of the variances of \( \hat{c}_k \) and \( \hat{\psi} \) defined in (4.30) and (4.28), respectively, to define Beale estimators of \( \gamma_k \) which are used to form the predictors. Let

\[
\hat{\gamma}_{k,B} = \frac{\hat{\psi}(\hat{c}_k n_k^{-1} + \hat{\psi}) + \hat{V}(\hat{\psi})}{(\hat{c}_k n_k^{-1} + \hat{\psi})^2 + \hat{V}(\hat{\psi}) + n_k^{-2}\hat{V}(\hat{c}_k)}.
\]

If only one two-way table is used to estimate \( \psi \) or \( c_k \), then the estimators of the variances in (4.31) or (4.27) are used instead. The initial predictor of the proportion calculated with the Beale estimator of \( \gamma_k \) is denoted by \( \hat{p}_{pred,ik,B} \) and is defined by

\[
\hat{p}_{pred,ik,B} = \hat{\gamma}_{k,B} \hat{p}_{ik} + (1 - \hat{\gamma}_{k,B}) \hat{p}_{T,ik}.
\]
The model for $\hat{c}_{k,t}$ in (4.29) is motivated by the moments of the Wishart distribution but holds under less restrictive distributional assumptions. As a consequence, the estimator (4.30) may be an unbiased variance estimator if the direct estimator of the covariance matrix of the sampling errors is not a Wishart matrix. In the simulations, the direct estimator of the sampling variance is not a Wishart matrix and the working model for the sampling variance does not hold, but the assumptions underlying the variance estimator (4.30) are satisfied.

The estimator of the variance of the estimator of $\psi$ based on the fourth moments of the normal distribution is biased in general. An alternative estimator of the variance of $\hat{\psi}$ uses the squares of the residuals, $[\{\hat{p} - \hat{p}_T\}^2 - \hat{\sigma}_{a,dir} - \hat{\sigma}_b \hat{\psi}]$, to estimate the fourth moments directly. In our simulations, the estimator based on the normal moments had a positive bias for the variance of $\hat{\psi}$, and the estimator based on the squared residuals had a negative bias.

### 4.2.3 Benchmarked Predictors

An initial predictor of the total in category $i$ and province $k$ is

$$\hat{M}_{pred,ik,B} = \hat{p}_{pred,ik,B} \hat{M}_k.$$  \hspace{1cm} (4.34)

The initial predictors of (4.34) are not benchmarked to the direct estimators of the category totals aggregated across provinces; that is,

$$\sum_{k=1}^{K} \hat{p}_{pred,ik,B} \hat{M}_k \neq \hat{M}_i.$$  \hspace{1cm} (4.34)

Because the table with $\hat{M}_{pred,ik,B}$ as the entry for category $i$ and province $k$ does not necessarily satisfy the desired benchmarking property, we use a final raking operation to benchmark the predictors. Let $\hat{M}_{ik,B}$ denote the predictor of the total in category $i$ and province $k$ in the raked table. Define the proportions corresponding to the benchmarked predictors of the totals by

$$\hat{p}_{ik,B} = \frac{\hat{M}_{ik,B}}{\hat{M}_k}.$$  \hspace{1cm} (4.35)

for $i = 1, \ldots, C$ and $k = 1, \ldots, K$. Unlike the initial univariate predictor $\hat{p}_{pred,ik,B}$, the benchmarked proportion in (4.35) does not necessarily fall between the direct estimator and
the synthetic estimator. However, the benchmarked predictors satisfy the marginal restrictions;
\[ \sum_{i=1}^{C} \hat{M}_{ik,B} = \hat{M}_{ik} \text{, and } \sum_{k=1}^{K} \hat{M}_{ik,B} = \hat{M}_{i}. \]

The subscript “B” is used to indicate that the predictors defined above are calculated with the Beale estimator of \( \gamma_k \). Let \( \hat{M}_{pred,ik} = \hat{\gamma}_{ik,ik} \hat{M}_{ik} \), where \( \hat{\gamma}_{pred,ik} \), defined in (4.24), is the initial predictor calculated with the standard estimator \( \hat{\gamma}_k \) defined in (4.25). Let \( \hat{M}_{ik} \) and \( \hat{\gamma}_{ik} \) denote the benchmarked predictors obtained from \( \hat{M}_{pred,ik} \).

**Remark:** Raking is not the only way to obtain benchmarked predictors that remain in the parameter space. Wang and Fuller (2008) review several benchmarking procedures for the linear model. We considered adapting three of the methods in Wang and Fuller (2008) to obtain benchmarked predictors that remain in the parameter space and preserve the marginal totals. One alternative benchmarking method is to change the estimating equation used to define the estimator of \( p_{T,ik} \) along the lines of the You et al. (2002 a) procedure so that the initial predictors satisfy benchmarking. A second possibility is to adapt the method of Pfeffermann and Barnard (1991) by finding the predictor that minimizes a quadratic form subject to the benchmarking restrictions. In a simulation, the adaptation of the Pfeffermann and Barnard (1991) method is at least as efficient as raking. Because the model for \( p_{T,k} \) is nonlinear and the predictors of the proportions are required to remain between zero and one, the adaptation of the Pfeffermann and Barnard (1991) method involves minimizing a nonlinear function subject to nonlinear constraints. Convergence problems were encountered when applying the method to the LFS data. A third option is to augment the matrix of covariates in the model for \( p_{T,ik} \) with category indicators multiplied by \( \hat{c}_k n_k^{-1} \hat{M}_{ik} \) and use the score function defined in (4.2) to estimate \( p_{T,ik} \). Based on the results of Wang and Fuller (2008) and Pfeffermann and Barnard (1991) for a linear model, we do not expect any method to be uniformly optimal. Because raking is simple to compute and has good efficiency in simulations, we use raking to benchmark the LFS predictors.

We present algorithms for two augmented model predictors.

**Augmented Model Predictors**

We describe two versions of augmented model predictors for a generalization of the
model (3.1) in which

\[ p_{T,ik} = \frac{\exp[x'_{ik}\lambda_0]}{1 + \sum_{j=2}^{C} \exp[x'_{jk}\lambda_0]}, \]  

(4.36)

and \( x_{ik} \) is an arbitrary vector of covariates. In the model (3.1), \( x_{ik} \) contains indicators for three digit categories and census interactions. Given estimators \( \hat{\psi}, \hat{c}_k, \) and \( \hat{\lambda} \) of \( \psi, c_k \) and \( \lambda_0 \), the initial predictor can be written as

\[ \hat{p}_{\text{pred},k}(\hat{\psi}, \hat{c}_k, \hat{\lambda}) = \hat{p}_k - (1 - \hat{\gamma}_k)(\hat{p}_k - p_{T,k}(\hat{\lambda})), \]

where \( p_{T,k}(\lambda) = (p_{T,1k}(\lambda), \ldots, p_{T,Ck}(\lambda))' \). The predictors \( \{\hat{p}_{\text{pred},k}(\hat{\psi}, \hat{c}_k, \hat{\lambda}) : k = 1, \ldots, K\} \) are benchmarked if they satisfy,

\[ \sum_{k=1}^{K} \hat{p}_{\text{pred},k}(\hat{\psi}, \hat{c}_k, \hat{\lambda})\hat{p}_k = \sum_{k=1}^{K} \{\hat{p}_k\hat{p}_k + (1 - \hat{\gamma}_k)(\hat{p}_k - p_{T,k}(\hat{\lambda}))\hat{p}_k\} = \hat{p}_i, \]

where \( \hat{p}_k = (\sum_{k=1}^{K} \hat{M}_k)^{-1}\hat{M}_k \), and \( \hat{p}_i \) is the \( C \)-dimensional vector of direct estimators of national category proportions. Because \( \sum_{k=1}^{K} \hat{p}_k\hat{p}_k = \hat{p}_i \), the benchmarking restriction can be expressed as,

\[ K^{-1} \sum_{k=1}^{K} \frac{\hat{c}_kn_k^{-1}\hat{p}_k}{\hat{c}_kn_k^{-1} + \hat{\psi}} X_{\text{cat},k}'(\hat{p}_k - p_{T,k}(\hat{\lambda})) = 0, \]

(4.37)

where \( X_{\text{cat},k} \) is the matrix containing the last \( C - 1 \) columns of the \( C \times C \) identity matrix. An estimated generalized least squares estimator of \( \lambda_0 \) under the working model for \( \Sigma_{ee,k} \) satisfies

\[ S_w(\hat{\lambda} \mid \hat{c}_k, \hat{\psi}) = K^{-1} \sum_{k=1}^{K} (\hat{c}_kn_k^{-1} + \hat{\psi})^{-1}(X_k^{(1)})'(\hat{p}_k^{(1)} - p_{T,k}^{(1)}(\hat{\lambda})) = 0, \]

(4.38)

where \( X_k^{(1)} \) contains the last \( C - 1 \) rows of the matrix of covariates \( X_k = (x_{1k}, \ldots, x_{Ck})' \), and \( p_{T,k}^{(1)}(\hat{\lambda}) \) contains the last \( C - 1 \) elements of \( p_{T,k}^{(1)}(\hat{\lambda}) \). By the form of \( S_w(\hat{\lambda} \mid \hat{c}_k, \hat{\psi}) \) and the expression for the benchmarking restriction in (4.37), the initial predictors satisfy the benchmarking property if the matrix \( \hat{c}_kn_k^{-1}\hat{p}_k X_{\text{cat},k} \) is a submatrix of \( X_k \). If \( \hat{c}_kn_k^{-1}\hat{p}_k X_{\text{cat},k} \) is not in the original model for the mean, then a way to achieve the benchmarking restriction is to augment the original matrix of covariates with \( \hat{c}_kn_k^{-1}\hat{p}_k X_{\text{cat},k} \) and construct predictors under the expanded model. Two ways to obtain augmented model predictors are described below.

The first five steps are essentially the same as one iteration of the procedure outlined in Section
4.1. The sixth step of each procedure is the critical part of the augmented model approach. The two augmented model predictors differ in how the parameters of the augmented model are estimated. In the first method, the parameter \( \lambda \) and the parameters of the augmented model are simultaneously estimated. In the second augmented model approach, the estimate of \( \lambda \) is held fixed at the estimate obtained under the initial model (4.36) when the model is expanded to achieve benchmarking. The efficiencies of the two augmented model predictors are compared to the efficiencies of the raked predictors through simulation in Section 6.7.

**Steps to Compute Augmented Model Predictor 1:**

1. Estimate \( \lambda \) using the score function corresponding to the multinomial likelihood in (6.2) in which the vector of direct estimators of totals in province \( k \) is assumed to be a multinomial random vector with a multinomial sample size of \( \hat{M}_k \). The initial estimator \( \hat{\lambda}^0 \) of \( \lambda \) satisfies
   \[
   S_M(\hat{\lambda}^0) = 0,
   \]
   where
   \[
   S_M(\lambda) = K^{-1} \sum_{k=1}^{K} \hat{M}_k(X_k^{(1)})'(\hat{p}_k^{(1)} - p_{T,k}(\lambda)).
   \]

2. Obtain an estimator \( \hat{c}_k \) using the method of Section 4.1.1 with \( \hat{\Gamma}^{(0)}_{uu,k} \) evaluated at \( \hat{\lambda}^{(0)} \) from step 1.

3. Obtain an initial estimator of \( \psi \) using the linear approximation for \( p_{T,k}(\hat{\lambda}^{(0)}) \) in (4.17). The initial estimator, \( \hat{\psi}^{(0)} \), has the form in (4.19), where the matrices \( \hat{\Sigma}^{(0)}_a \) and \( \hat{\Sigma}^{(0)}_b \) needed for the EGLS estimator of (4.20) are evaluated at the estimator of \( \lambda \) obtained in step 1.

4. Define an updated estimator of \( \lambda \) to be a solution to the estimating equations (4.38), where \( \hat{c}_k \) and \( \hat{\psi}^{(0)} \), obtained in steps 1 and 2, respectively, are used to estimate the variances. The updated estimator \( \hat{\lambda} \) satisfies
   \[
   S_w(\hat{\lambda} | \hat{c}_k, \hat{\psi}^{(0)}) = 0,
   \]
   where
   \[
   S_w(\lambda | \hat{c}_k, \hat{\psi}^{(0)}) = K^{-1} \sum_{k=1}^{K} \frac{1}{\hat{c}_ kn_k^{-1} + \hat{\psi}^{(0)}} (X_k^{(1)})'(\hat{p}_k^{(1)} - p_{T,k}(\lambda)).
   \]

5. Update an estimator of \( \psi \) using the method of section 4.1.3 with the matrices \( \hat{\Sigma}_a \) and \( \hat{\Sigma}_b \) evaluated at the estimator \( \hat{\lambda} \) obtained in step 4. Denote the updated estimator of \( \psi \) by \( \hat{\psi} \) (or \( \hat{\psi}_D \) if multiple 2-way tables are used to estimate \( \psi \)).
6. Define $X_{k,aug} = X_{cat,k}n_k^{-1}\hat{p}_k$ and $X_{k,aug,1} = (X_k, X_{k,aug})$. The score function for the augmented model is

$$S_{w,aug1}(\beta | \hat{c}, \hat{\psi}) = K^{-1} \sum_{k=1}^{K} (\hat{c}_k n_k^{-1} + \hat{\psi})^{-1} (X_{k,aug}^{(1)})' (\hat{p}_k^{(1)} - p_{T,k}^{(1)}(\beta)),$$

(4.39)

where $\beta = (\lambda', \delta')'$, $\delta$ is a $C - 1 \times 1$ vector of parameters associated with $X_{k,aug}$, $p_{T,k}^{(1)}(\beta) = (p_{T,2k}(\beta), \ldots, p_{T,Ck}(\beta))'$,

$$p_{T,ik}(\beta) = \frac{\exp[x_{ik}' \lambda + x_{ik,aug}']}{1 + \sum_{j=2}^{C} \exp[x_{jk}' \lambda + x_{jk,aug}']},$$

and $x_{ik,aug}$ is the row of $X_{k,aug}$ associated with category $i$ for $i = 1, \ldots, C$. Define $\hat{\beta}$ to be a root of the score function $S_{w,aug1}(\cdot | \hat{c}, \hat{\psi})$.

7. Define the augmented model predictor,

$$\hat{p}_{pred,ik,aug1} = \hat{p}_{ik} - (1 - \hat{\gamma}_k)(\hat{p}_{ik} - p_{T,ik}(\hat{\beta})).$$

(4.40)

**Steps to Compute Augmented Model Predictor 2:**

1-5 Complete steps 1-5 above.

6. Given the estimators of $\lambda_o$, $c_k$ and $\psi$ from step 5 above, define the function of $\delta$,

$$S_{w,aug2}(\delta | \hat{c}, \hat{\psi}, \hat{\lambda}) = K^{-1} \sum_{k=1}^{K} (\hat{c}_k n_k^{-1} + \hat{\psi})^{-1} (X_{k,aug}^{(1)})' (\hat{p}_k^{(1)} - p_{T,k}^{(1)}(\delta | \hat{\beta})),$$

(4.41)

where $X_{k,aug}^{(1)}$ denotes the matrix consisting of the last $C - 1$ rows of $X_{k,aug}$, $p_{T,k}(\delta | \hat{\beta})^{(1)} = (p_{T,2k}(\delta | \hat{\beta}), \ldots, p_{T,Ck}(\delta | \hat{\beta}))'$, and

$$p_{T,ik}(\delta | \hat{\lambda}) = \frac{\exp[x_{ik}' \hat{\lambda} + x_{ik,aug}']}{1 + \sum_{j=2}^{C} \exp[x_{jk}' \hat{\lambda} + x_{jk,aug}']},$$

Let $\hat{\delta}_{aug2}$ satisfy $S_{w,aug2}(\delta | \hat{c}, \hat{\psi}, \hat{\lambda}) = 0$.

7. Define the augmented model predictor,

$$\hat{p}_{pred,ik,aug2} = \hat{p}_{ik} - (1 - \hat{\gamma}_k)(\hat{p}_{ik} - p_{T,ik}(\hat{\delta}_{aug2} | \hat{\lambda})).$$

(4.42)
Remark 1: A comparison of the score functions (4.39) and (4.41) to the benchmarking restriction in (4.37) shows that the augmented model predictors \( \{\hat{p}_{\text{pred},ik,aug1}, \hat{p}_{\text{pred},ik,aug2} : i = 1, \ldots, C; k = 1, \ldots, K \} \) are benchmarked to the direct estimators of the marginal totals.

Remark 2: If \( \hat{c}_k n_k^{-1} \hat{M}_k \) is constant across provinces, then the initial predictors constructed with the estimator of \( \lambda \) that solves the score function, \( S_w(\hat{\lambda} | \hat{c}_k, \hat{\psi}) \), are benchmarked provided indicators for three digit codes are included in the vectors of covariates. If the sampling procedure and design based estimators are such that \( \hat{c}_k \) is constant across provinces and the sample size \( n_k \) is proportional to \( \hat{M}_k \), then the condition for benchmarking is satisfied.
CHAPTER 5. MSE Estimators

Alternative estimators of the MSE are proposed in this subsection. An estimator of the MSE of the initial predictor of the proportion is developed using a method related to the Prasad-Rao (1990) approach. The special case in which the sampling covariance matrix is proportional to the covariance matrix of the model random effects is considered. An estimator of the MSE of the initial predictor of the total is derived using an approach analogous to the derivation used for the proportions. The MSE estimators appropriate for the initial predictors do not account for the raking operation, and difficulties associated with estimating the MSE of the benchmarked predictors are discussed. A bootstrap estimator of the MSE is proposed.

5.1 MSE Estimators based on Taylor Linearization

Closed form estimators of the MSE’s of the initial predictors are obtained using Taylor linearization. First, we define estimators of the MSE’s of the proportions and consider the situation in which the sampling covariance matrix is proportional to a multinomial covariance matrix. Second we define MSE estimators for totals and discuss issues involved with using a linear approximation for raking.

5.1.1 MSE Estimators for Proportions

To obtain an approximation for the estimator of the MSE of the predictor, define the predictor with known $\gamma_k$ by

$$\hat{p}_{ik}(\gamma_k) = \hat{p}_{T,ik} + \gamma_k(\hat{p}_{ik} - \hat{p}_{T,ik}),$$

where $\gamma_k = \psi(c_k n_k^{-1} + \psi)^{-1}$. Let $\hat{p}(\gamma)$ be the vector with elements $\{\hat{p}_{ik}(\gamma_k) : i = 1, \ldots, C; k = 1, \ldots, K\}$, let $D_\gamma = \text{diag}(\gamma_1, \ldots, \gamma_K) \otimes I_C$, and let $p$ be the vector of true proportions defined in
(3.2) listed in order with elements for a province grouped together. By the linear approximation for the synthetic estimator in (4.13),

\[ \tilde{p}(\gamma) - p \approx D_\gamma (u + e) - u + (I_{CK} - D_\gamma) W'_p (u + e), \]

where \( W'_p \) is defined after (4.13). Letting \( D_1 = D_\gamma + (I_{CK} - D_\gamma) W'_p \) and \( D_2 = D_1 - I_{CK} \), an approximation for the MSE of the vector predictor \( \tilde{p}(\gamma) \) is

\[ MSE_1 = D_1 \Sigma_{ee} D'_1 + \psi D_2 \Gamma_{uu} D'_2. \]  

(5.1)

The MSE in (5.1) can be written as

\[ MSE_1 = MSE_{11} + (I_{CK} - D_\gamma) V_{syn}(I_{CK} - D_\gamma) \]

\[ + (I_{CK} - D_\gamma) W'_p (\psi \Gamma_{uu}) (D_\gamma - I_{CK}) + (D_\gamma - I_{CK}) (\psi \Gamma_{uu}) W_p (I_{CK} - D_\gamma) \]

\[ + (I_{CK} - D_\gamma) W'_p \Sigma_{ee} D_\gamma + D_\gamma \Sigma_{ee} W_p (I_{CK} - D_\gamma), \]

where

\[ MSE_{11} = D_\gamma \Sigma_{ee} D_\gamma + \psi(I_{CK} - D_\gamma) \Gamma_{uu} (I_{CK} - D_\gamma), \]  

(5.3)

and

\[ V_{syn} = \Gamma_{uu} X (H^{-1}) X' D_w^{-1} (\Sigma_{ee} + \psi \Gamma_{uu}) D_w^{-1} X H^{-1} X' \Gamma_{uu} \]  

(5.4)

is the approximate variance of \( \hat{p}_T \). The first term, \( MSE_{11} \), in (5.2) is the MSE of the predictor in (3.12) calculated with the true parameters. The diagonal element of \( MSE_{11} \) associated with category \( i \) and province \( k \) is

\[ MSE_{11,ik} = \gamma_k^2 \sigma_{e,ik}^2 + \psi(1 - \gamma_k)^2 p_{T,ik}(1 - p_{T,ik}). \]  

(5.5)

The second line in (5.2) accounts for the variance of the synthetic estimator. The last two terms in (5.2) account for covariances between the error in the predictor of \( u \) calculated with the unknown parameters and the error in the synthetic estimator.

An estimator of the MSE of \( \tilde{p}(\gamma) \) is obtained by evaluating the unknown parameters in \( MSE_1 \) at estimators. Let

\[ \tilde{MSE}_1 = \tilde{D}_1 \tilde{\Sigma}_{ee,ma} \tilde{D}'_1 + \tilde{\psi} \tilde{D}_2 \tilde{\Gamma}_{uu} \tilde{D}'_2, \]  

(5.6)
where $\hat{D}_1$, $\hat{D}_2$, $\hat{\Sigma}_{ee,md}$, and $\hat{\Gamma}_{uu}$ are estimators of the corresponding population quantities in (5.1). The diagonal elements of $D_\gamma$ needed for the estimators of $D_1$ and $D_2$ are evaluated at the Beale estimators, $\{\hat{\gamma}_{k,B} : k = 1, \ldots, K\}$. The estimator $\hat{\Sigma}_{ee,md}$ is a modified direct estimator of the sampling covariance matrix, which is a block diagonal matrix with $k$-th block defined in (4.7). Let $\hat{MSE}_{1,ik}$ denote the diagonal element of $\hat{MSE}_1$ associated with category $i$ and province $k$.

To estimate the MSE of $\hat{p}_{\text{pred},ik,B}$, a term for the variance of $\hat{\gamma}_{k,B}$ is added to $\hat{MSE}_{1,ik}$. The estimator $\hat{\gamma}_{k,B}$ is estimated using the Prasad-Rao approach (Rao, 2003, pg. 103-105).

The partial derivatives of $\gamma_k$ with respect to $\psi$ and $c_k$, evaluated at $\psi$ and $c_k$, are

$$\frac{\partial \gamma_k}{\partial \psi} = \frac{c_k n_k^{-1}}{(\psi + c_k n_k^{-1})^2},$$

(5.7)

and

$$\frac{\partial \gamma_k}{\partial c_k} = -\frac{\psi n_k^{-1}}{(c_k n_k^{-1} + \psi)^2},$$

(5.8)

respectively. Let

$$g_{3,ik} = \left[ \left( \frac{\partial \gamma_k}{\partial \psi} \right)^2 V\{\hat{\psi}\} + \left( \frac{\partial \gamma_k}{\partial c_k} \right)^2 V\{\hat{c}_k\} \right] \left( \sigma_{e,ik}^2 + \psi p_{T,ik} (1 - p_{T,ik}) \right),$$

(5.9)

where $V\{\hat{\psi}\}$ and $V\{\hat{c}_k\}$ denote the variances of $\hat{\psi}$ and $\hat{c}_k$, respectively, and $V\{u_{ik} + e_{ik}\} = \sigma_{e,ik}^2 + \psi p_{T,ik} (1 - p_{T,ik})$. From (5.1) and (5.9), an approximation for the MSE of $\hat{p}_{\text{pred},ik,B}$ is

$$MSE_{2,ik} = MSE_{1,ik} + g_{3,ik}.$$  

(5.10)

The approximation (5.10) ignores covariances between the synthetic estimators and the estimator of $\psi$. The approximation also assumes that $E[(\hat{p}_T - p_T)(\hat{p}_T - p_T)'(\hat{p}_T - p_{T,ik})]$ is negligible.

In the linear model with normal errors, an estimator of the leading term in the MSE has a negative bias (Rao, 2003, pg. 104). Let

$$\hat{MSE}_{11,ik} = \hat{\gamma}_{k,B}^2 \sigma_{e,ik,md}^2 + (1 - \hat{\gamma}_{k,B})^2 \psi p_{T,ik} (1 - p_{T,ik})$$

(5.11)

be an estimator of the leading term, $MSE_{11,ik}$, defined in (5.5), where $\sigma_{e,ik,md}^2$ is the diagonal element of the modified estimator of the sampling variance associated with category $i$ and
province $k$. In analogy with the theory for the linear model with normal errors, we use a Taylor expansion to approximate the bias of (5.11) for the leading term defined in (5.5). The approximation assumes $E[\hat{\gamma}_k^2 \sigma^2_{e,ik,md}] = E[\hat{\gamma}_k^2] \sigma^2_{e,ik}$. The second derivative of $MSE_{11,ik}$ with respect to $\psi$ is

$$\frac{\partial^2 MSE_{11,ik}}{\partial \psi^2} = 2\left( \left( \frac{\partial \gamma_k}{\partial \psi} \right)^2 + \gamma_k \frac{\partial^2 \gamma_k}{\partial \psi^2} \right) \sigma^2_{e,ik}$$

$$+ \left( \frac{\partial^2 \gamma_k}{\partial \psi^2} \right)^2 \left( \frac{c_k n_k^{-1} \psi}{(\psi + c_k n_k^{-1})^3} \right) - \left( \frac{\partial \gamma_k}{\partial \psi} \right)^2 \left( \frac{c_k n_k^{-1} \psi}{(\psi + c_k n_k^{-1})} \right)^2 (1 - p_{T,ik})$$

where

$$\frac{\partial^2 \gamma_k}{\partial \psi^2} = -2 \frac{c_k n_k^{-1}}{(\psi + c_k n_k^{-1})^3}.$$

The second derivative of $MSE_{11,ik}$ with respect to $c_k$ is

$$\frac{\partial^2 MSE_{11,ik}}{\partial c_k^2} = \frac{6 n_k^{-2} \psi^2}{(\psi + c_k n_k^{-1})^4} \sigma^2_{e,ik}$$

$$- 2 \left( \rho^2 c_k n_k^{-2} \psi \right) \left( \frac{\psi n_k^{-1}}{(c_k n_k^{-1} + \psi)^2} \right)^2 - 2 \left( \rho^2 c_k n_k^{-2} \psi \right) \left( \frac{\psi n_k^{-1}}{(c_k n_k^{-1} + \psi)^2} \right)^2 (1 - p_{T,ik})$$

and the second derivative of $MSE_{11,ik}$ with respect to $p_{T,ik}$ is

$$\frac{\partial^2 MSE_{11,ik}}{\partial p_{T,ik}^2} = -2(1 - \gamma_k)^2 \psi.$$

An approximation for the bias of $\hat{MSE}_{11,ik}$ for $MSE_{11,ik}$ is $-g_{4,ik}$, where

$$g_{4,ik} = -0.5 \frac{\partial MSE_{11,ik}}{\partial \psi^2} V\{\hat{\psi}\} - 0.5 \frac{\partial^2 MSE_{11,ik}}{\partial c_k^2} V\{\hat{c}_k\} - 0.5 \frac{\partial^2 MSE_{11,ik}}{\partial p_{T,ik}^2} V\{\hat{p}_{T,ik}\}$$

An estimator of the MSE of $\hat{p}_{pred,ik}$ is

$$\hat{MSE}_{2,ik,B} = \hat{MSE}_{1,ik} + \hat{g}_{3,ik} + \hat{g}_{4,ik},$$

where $\hat{MSE}_{1,ik}$ is the diagonal element of $\hat{MSE}_1$, defined in (5.6), and $\hat{g}_{3,ik}$ and $\hat{g}_{4,ik}$ are estimators of $g_{3,ik}$ and $g_{4,ik}$, respectively. We use the estimators defined in (4.28) and (4.30) to estimate the variances of the estimators of $\psi$ and $c_k$ needed for $\hat{g}_{3,ik}$ and $\hat{g}_{4,ik}$.

**Remark:** The leading term in the MSE of the predictor calculated with $\hat{\gamma}_{ik,o,B}$ defined in (4.26) is

$$g_{1,ik,o} = \hat{\gamma}_{ik,o} \sigma^2_{e,ik} = \frac{\psi p_{T,ik}(1 - p_{T,ik})}{\psi p_{T,ik}(1 - p_{T,ik}) + \sigma^2_{e,ik}} \sigma^2_{e,ik}$$
An estimator of the leading term obtained by replacing $\gamma_{ik,o}$ with the Beale estimator $\hat{\gamma}_{ik,o,B}$ has a bias of smaller order than $K^{-1}$. Because the sampling variances are estimated and the predictors are calculated under the working model, the leading term in the MSE of the initial predictor of the proportion for cell $(i,k)$ has the form in (5.5). Evaluating $MSE_{11,ik}$ of (5.5) at the Beale estimator of $\gamma_k$ does not necessarily lead to a bias reduced estimator if the sampling variances are estimated or if the working model does not hold. Wang and Fuller (2003) use a Beale ratio estimator of the leading term in the MSE approximation under an assumption that the direct estimators of the sampling variances have chi-squared distributions and are independent of the estimator of the random effects variance. We construct estimators of the MSE by evaluating $MSE_{11,ik}$, $g_3,ik$ and $g_4,ik$ at the Beale estimator, $\hat{\gamma}_{k,B}$. The resulting estimator of the MSE is defined in (5.14).

5.1.1.1 Sampling Covariance Proportional to Multinomial Covariance

The mean squared errors of the predictors of the proportions for the special case in which the working model for the sampling covariance matrix is assumed to be correct is considered in this section. Under an assumption that the covariance matrix of the vector of proportions in province $k$ is

$$\Sigma_{ee,k,w} = \frac{c_k}{n_k} \Gamma_{uu,k}, \quad (5.15)$$

the estimator

$$\hat{\Sigma}_{ee,k,w} = \frac{\hat{c}_k}{n_k} \hat{\Gamma}_{uu,k} \quad (5.16)$$

is a consistent estimator of the covariance matrix for the proportions. As discussed in section 3.5, the optimal convex combination of $\hat{p}_{ik}$ and the true mean $p_{T,ik}$ is equal to the $i^{th}$ element of the vector predictor defined in (3.10). The components of the MSE of the predictors of the proportions simplify under the assumption that $\Sigma_{ee,k}$ has the form in (5.15).

- The leading term, $MSE_{11}$ of (5.2) is

$$MSE_{11} = \text{block-diag}(\frac{c_1}{n_1} \Gamma_{uu,1}, \ldots, \frac{c_K}{n_K} \Gamma_{uu,K}). \quad (5.17)$$
An estimator of the $i^{\text{th}}$ diagonal element of $\gamma_k \frac{c_k}{n_k} \Gamma_{uu,k}$ evaluated at the Beale estimator of $\gamma_k$ is

$$\hat{MSE}_{11,ik,B} = \hat{\gamma}_{t,B} c_k n_k^{-1} \hat{p}_{t,ik} (1 - \hat{p}_{t,ik}).$$  \tag{5.18}$$

Assuming $E[\hat{MSE}_{11,ik,B}] \approx E[\hat{\gamma}_{k,B} c_k n_k^{-1} p_{t,ik}(1 - p_{t,ik})]$, use of the Beale estimator of $\gamma_k$ leads to an approximately unbiased estimator of the leading term.

- The covariance between $\gamma_k (\hat{p}_k - p_{t,k})$ and $\hat{p}_{t,k} - p_{t,k}$ in (5.2) can be written as

$$\left( D_n \psi \hat{V} V^T - p_{t,k} \psi \hat{V} \right) \hat{X} \left( I_{n} - D_n \psi \hat{V} \right),$$

where $V$ is a block-diagonal matrix with $\Sigma_{ee,k} + \psi \Gamma_{uu,k}$ as the $k^{\text{th}}$ block. When $\Sigma_{ee,k} = \Sigma_{ee,k,w}$ of (5.16), $D_n \psi \hat{V} V^T = \psi \Gamma_{uu} \hat{V}$, and the covariance is zero.

- The approximate variance of $\hat{p}_{t,k}$ defined in (5.4) simplifies to

$$V_{syn,w} = \Gamma_{uu} \hat{X} (H_{XX})^{-1} \hat{X} \Gamma_{uu}$$

under the assumption that $\Sigma_{ee,k}$ has the form in (5.15).

- Under the working model, the $g_{3,ik}$ of (5.9) simplifies to

$$g_{3,ik,w} = p_{t,ik}(1 - p_{t,ik}) \left[ c_k^2 n_k^{-2} V\{\hat{\psi}\} + \psi^2 n_k^{-2} V\{\hat{c}_k\} \right].$$

The diagonal element of (5.17) associated with category $ik$ is

$$MSE_{11,ik} = \frac{\hat{\psi}}{\psi + c_k n_k^{-1}} p_{t,ik}(1 - p_{t,ik}) c_k n_k^{-1}. $$ \tag{5.19}$$

The second partial derivatives of $MSE_{11,ik}$ with respect to $\psi$, $c_k$, and $p_{t,ik}$ are

$$\frac{\partial^2 MSE_{11,ik}}{\partial \psi^2} = -2 \frac{c_k^2 n_k^{-2}}{(\psi + c_k n_k^{-1})^3} p_{t,ik}(1 - p_{t,ik}),$$

$$\frac{\partial^2 MSE_{11,ik}}{\partial c_k^2} = -2 \frac{\psi c_k n_k^{-1}}{(\psi + c_k n_k^{-1})^3} p_{t,ik}(1 - p_{t,ik}),$$

and

$$\frac{\partial^2 MSE_{11,ik}}{\partial p_{t,ik}^2} = -2 \gamma_k c_k n_k^{-1},$$

where $\Gamma_{uu,k}$ is the $k^{\text{th}}$ diagonal element of $\Gamma_{uu}$.
respectively. Then, the approximation for the bias of the estimator of the leading term is $-g_{4,ik,w}$, where

$$g_{4,ik,w} = \frac{c_k^2 n_k^{-2}}{(\psi + c_k n_k^{-1})^3} p_{T,ik} (1 - p_{T,ik}) V\{\hat{\psi}\} + \frac{c_k^2 n_k^{-2}}{(\psi + c_k n_k^{-1})^3} p_{T,ik} (1 - p_{T,ik}) V\{\hat{c}_k\} + \gamma_k c_k n_k^{-1} V\{\hat{p}_{T,ik}\}.$$

If the constant $c_k$ is assumed known or if the variance of $\hat{c}_k$ is negligible, then

$$g_{4,ik,w} = 2 g_{3,ik,w} + \gamma_k c_k n_k^{-1} V\{\hat{p}_{T,ik}\}.$$ 

### 5.1.2 MSE Estimators for Totals

Estimators of the MSE's of the initial predictors of the totals are obtained using an approach analogous to that used for the proportions. Let

$$\hat{M}_{pred,ik}(\gamma_k) = \hat{p}_{ik}(\gamma_k) \hat{M}_k,$$

and let $\hat{M}_{pred}(\gamma)$ be the vector with elements $\hat{M}_{pred,ik}(\gamma_k)$ listed in an order with provinces grouped together. Then,

$$\hat{M}_{pred}(\gamma) - M = D_{\gamma}(\hat{M} - M) + (I_{CK} - D_{\gamma})(\hat{T} - M) \quad (5.20)$$

$$\approx D_{\gamma} a - (I_{CK} - D_{\gamma}) D_{prov} u + (I_{CK} - D_{\gamma}) D_{prov} W_p(u + e) + \text{diag}(p_T)[(I_K \otimes J'_C) \otimes J_C] a$$

$$= D_{1,M} a + D_{2,M} u + D_{3,M} e,$$

where,

$$D_{1,M} = D_{\gamma} + (I_{CK} - D_{\gamma}) \text{diag}(p_T)[(I_K \otimes J'_C) \otimes J_C],$$

$$D_{2,M} = (I_{CK} - D_{\gamma}) D_{prov}(W'_p - I_{CK}),$$

$$D_{3,M} = (I_{CK} - D_{\gamma}) D_{prov} W'_p,$$
and, as defined following (4.17), \( a = \hat{M} - M \) and \( D_{\text{prov}} = \text{diag}(\hat{M}_1, \ldots, \hat{M}_K) \otimes I_K \). An estimator of the MSE based on (5.20) is

\[
\hat{\text{MSE}}_{1,T} = \hat{D}_{1,M} \hat{\Sigma}_{aa,md} \hat{D}'_{1,M} + \hat{\psi} \hat{D}_{2,M} \hat{\Gamma}_{uu} \hat{D}'_{2,M} + \hat{D}_{3,M} \hat{\Sigma}_{ee,md} \hat{D}'_{3,M} \quad (5.21)
\]

where \( \hat{\Sigma}_{ee,md} \) and \( \hat{\Sigma}_{aa,md} \), defined in (4.7) and (4.6), respectively, are the modified versions of the direct estimators of the sampling variances in the proportions and totals, respectively. The estimator (5.21) does not account for variability due to estimation of \( \gamma_k \). To account for the variance of \( \hat{\gamma}_k \), we use a Taylor expansion analogous to the one used for the proportions, and define

\[
g_{3,ik,T} = \left( \frac{\partial \gamma_k}{\partial \psi} \right)^2 V\{\hat{\psi}\} + \left( \frac{\partial \gamma_k}{\partial c_k} \right)^2 V\{\hat{c}_k\} \left( V\{a_{ik}\} + T_k^2 \hat{\psi} p_{T,ik}(1 - p_{T,ik}) \right). \quad (5.22)
\]

The leading term in the MSE for the totals is

\[
\text{MSE}_{11,ik,T} = \gamma_k^2 V\{a_{ik}\} + (1 - \gamma_k)^2 T_k^2 \hat{\psi} p_{T,ik}(1 - p_{T,ik}). \quad (5.23)
\]

An estimator of the bias of an estimator of \( \text{MSE}_{11,ik,T} \) is obtained through a second order Taylor expansion. The second derivative of \( \text{MSE}_{11,ik,T} \) with respect to \( \psi \) is obtained by replacing \( \sigma^2_e,ik \) in (5.12) by \( V\{a_{ik}\} \) and multiplying the second line in (5.12) by \( T_k^2 \). The second derivatives of \( \text{MSE}_{11,ik,T} \) with respect to \( \psi \) and \( c_k \) are obtained similarly. The second derivative of \( \text{MSE}_{11,ik,T} \) with respect to \( T_k \) is \( 2(1 - \gamma_k)^2 \hat{\psi} p_{T,ik}(1 - p_{T,ik}) \). Then, an estimator of the bias of an estimator of the leading term (5.23) is obtained by multiplying the second derivatives by the corresponding variances, as in (5.13). Letting \( \hat{g}_{4,ik,T} \) be the resulting estimator of the bias of an estimator of (5.23), an estimator of the MSE of the predictor of the total in category \( i \) and province \( k \) is given by

\[
\hat{\text{MSE}}_{2,ik,T,B} = \hat{\text{MSE}}_{1,ik,T} + \hat{g}_{3,ik,T} + \hat{g}_{4,ik,T}; \quad (5.24)
\]

where \( \hat{\text{MSE}}_{1,ik,T} \) is the diagonal element of \( \hat{\text{MSE}}_{1,T} \), and \( \hat{g}_{3,ik,T} \) is an estimator of \( g_{3,ik,T} \) defined in (5.22). The subscript \( B \) indicates that the estimator is evaluated at the Beale estimator of \( \gamma_k \).
5.1.3 Estimating the MSE After Raking

The estimators of the MSE defined in the previous subsections do not account for the raking operation used to benchmark the predictors. Estimation of the effect of raking on the MSE of the predictors is challenging because the raking procedure applies a nonlinear transformation to the initial univariate predictors. Dick (1995) states that a Taylor approximation can be used but cautions that the Taylor approximation assumes that the small area sample sizes are large, which may not be realistic in applications. Instead of using the Taylor approximation, Dick (1995) suggests multiplying an estimator of the MSE of the initial predictors by the squared ratio of the raked predictors to the initial predictors to estimate the effect of raking on the MSE of the predictor. Pfeffermann et al. (2006) assert that the double bootstrap can be used to obtain an accurate estimator of the MSE of the raked predictors but do not provide empirical or theoretical justification for using the bootstrap to account for the MSE after raking. Zhang and Chambers (2004) use the bootstrap to estimate the MSE of the raked predictors conditional on the model random effects. The MSE estimator of Zhang and Chambers (2004) estimates the MSE with respect to the distribution generated by repeated sampling from a fixed finite population.

Under an assumption that the province sample sizes, \( n_k \), increase, a linear approximation for the error in the raked predictor is

\[
\hat{M} - M \approx \hat{M} - M - (I_{CK} - \tilde{A}\tilde{X}') (I_{CK} - \hat{D}_\gamma)(\hat{M} - \hat{T}),
\]

where \( \tilde{X} \) denotes the \( CK \times (C-1) \) matrix with \( \tilde{x}_{ik}' = (I[i = 2], \ldots, I[i = C], I[k = 1], \ldots, I[k = K]) \) in the row for cell \((i,k)\), and \( \tilde{A} = \text{diag}(M)\tilde{X}(\tilde{X}'\text{diag}(M)\tilde{X})^{-1} \). The linear approximation is obtained by a second order Taylor expansion of the objective function defining the raking operation (Deville and Sarndal, 1993). 1, 1993).

The matrix \((I_{CK} - \tilde{A}\tilde{X}')\) contains the derivatives associated with the linear approximation and is a function of the vector of random effects \(u\). Because we do not assume that the variance of \(u\) converges to zero, we do not have a formal justification for using the linearization in (5.25) to define an MSE estimator. Therefore, we use simulation to estimate the MSE of the raked predictor.
5.2 Bootstrap Estimator of MSE

As an alternative to Taylor approximations, we propose a simulation based estimator of the MSE of the predictor. The method is an adaptation of the moment-matching, or “wild,” bootstrap (Wu, 1986). The wild bootstrap (Wu, 1986) uses a bootstrap distribution with moments that are consistent for the moments of the model. Wu (1986) introduced the wild bootstrap for linear regression. Liu (1988) recommends using a three point distribution to generate bootstrap samples. Shao and Tu (1995, pg. 343) discuss applications of the moment-matching bootstrap in generalized linear models.

Applications of the moment-matching bootstrap in small area estimation include Gonzalez-Manteiga (2007) and Hall and Maiti (2006). Gonzalez-Manteiga use the moment-matching bootstrap to estimate the MSE of small area predictors based on a logistic mixed model and compare bootstrap estimators to MSE estimators based on Taylor approximations. Hall and Maiti (2006) suggest the moment-matching bootstrap as a way to obtain non-parametric estimators of the MSE in the context of a unit-level mixed linear model. They specify bootstrap distributions to match nonparametric estimates of the first, second, and fourth moments. Use of consistent estimators of the fourth moments ensures that the bootstrap estimator of the variance of the estimator of the random effects variance is consistent. Hall and Maiti (2006) recommend using a three point distribution or a t distribution to generate the bootstrap distribution and explain how to specify parameters to match the first, second and fourth moments.

5.2.1 Bootstrap Distributions

In our application of the bootstrap, the true proportions are generated from Dirichlet distributions. We use the notation \( p_k \sim \text{Dirichlet}_C(p_{T,k}, \omega_o) \) to mean that the probability density function of \( p_k^{(1)} \), the last \( C - 1 \) elements of \( p_k \), is given by

\[
P(v_{2k}, \ldots, v_{(C)k}) = \left[ \frac{\Gamma(\omega_o)}{\prod_{i=1}^{C} \Gamma(\omega_{ik})} \right] \prod_{i=1}^{C} v_{ik}^{\omega_{ik}-1},
\]

where \( \omega_{ik} = p_{T,ik} \omega_o, \ v_{ik} > 0 \ (i = 2, \ldots, C), \ \sum_{i=2}^{C} v_{ik} < 1, \ v_{1k} = 1 - \sum_{i=2}^{C} v_{ik}, \) and \( \Gamma(a) = \int_0^\infty t^{a-1}e^{-t}dt \). If \( p_k \sim \text{Dirichlet}_C(p_{T,k}, \omega_o) \), then \( E[p_k] = p_{T,k} \), and \( V[p_k] = (\omega_o + 1)^{-1} \Gamma_{uu,k} \).
The LFS provides a jackknife estimate of the sampling variance in the scale of totals. Let $\hat{\Sigma}_{aa,k,md}$ denote the jackknife estimate of the sampling variance of the vector of totals for province $k$ after modifying the covariance matrix for zero estimates as described in (4.6). The proposed bootstrap data generating method was developed under an assumption that

$$E[\hat{\Sigma}_{aa,k,md}] = V\{\hat{M}_k - M_k\}.$$ 

As a consequence, in the bootstrap procedure described below, a vector of direct estimators, $\hat{M}_k^*$, and a vector of true totals, $M_k^*$, are generated so that the variance of $\hat{M}_k^* - M_k^*$ under the bootstrap distribution is approximately equal to $\hat{\Sigma}_{aa,k,md}$.

The sampling errors for the totals are generated by converting correlated normal random variables to Gamma random variables. This method produces bootstrap versions of the direct estimators of totals and proportions that remain in the parameter space. A problem with this data generating method is that the method does not preserve the desired second moments. The data generating method preserves the desired variances for the totals but distorts the correlations in the scale of totals. The distortion of the covariance matrix resulting from the bootstrap data generating procedure is discussed further in Appendix 4.

In the LFS application, the direct estimator of the sampling variance of the proportions is obtained from the direct estimator of the sampling variance of the totals using the linear approximation of Appendix 1. The variance of the proportions generated in the bootstrap does not equal the direct estimator of the sampling variance obtained from the linear approximation. The relationship between the variance of the proportions generated in the bootstrap and the direct estimator of the variance is discussed further in the Section 5.2.4 entitled “Moments of Bootstrap Distributions” and in Appendix 4.

A bootstrap version of the direct estimator of the covariance matrix is required to construct bootstrap MSE estimators. If one is willing to ignore variability in the direct estimators of the sampling variances, then one can use the original direct estimators of the sampling variances to construct the bootstrap MSE estimators. We generate bootstrap versions of the direct estimators of the sampling covariance matrices from Wishart distributions. The mean of the Wishart distribution is the direct estimator of the sampling covariance matrix, and the
degrees of freedom are \([\lfloor n_k \hat{e}_k^{-1} - 1 \rfloor : k = 1, \ldots, K]\), where \(n_k\) is the realized sample size in province \(k\).

### 5.2.1.1 Steps to simulate from bootstrap distributions

The bootstrap data generating method consists of three steps.

1. For \(k = 1, \ldots, K\), let \( p_k^* \sim \text{Dirichlet}(\hat{p}_{T,k}, \hat{\omega}_o) \), where \( \hat{\omega}_o = \hat{\psi}^{-1} - 1 \), and set \( M_k^* = p_k^* \hat{M}_k \). The variables \( p_k^* \) and \( M_k^* \) are bootstrap versions of the vectors of true proportions and totals, respectively, for province \(k\).

2. Let \( \hat{\Sigma}_{aa,k,md} \) be the estimator of the sampling variance of the vector of totals in province \(k\) defined in Section 3, and let \( \hat{R}_{D,k} \) be the corresponding correlation matrix. Let \( \hat{V}_{D,ik} \) be the \(i\)th diagonal element of \( \hat{\Sigma}_{aa,k,md} \). For \(k = 1, \ldots, K\), generate \( Z_k^* \sim N(0, \hat{R}_{D,k}) \), and let \( W_{ik}^* = \Phi(Z_{ik}) \), where \( Z_k = (Z_{1k}, \ldots, Z_{Ck})' \), and \( \Phi \) is the standard normal CDF. Set

\[
\alpha_{ik}^* = \frac{ (M_{ik}^*)^2 }{ \hat{V}_{D,ik} } \tag{5.28}
\]

and

\[
\beta_{ik}^* = \frac{ M_{ik}^* }{ \hat{V}_{D,ik} } \tag{5.29}
\]

where \( M_{ik}^* \) is defined in step 1. Let \( G_{ik}^* \) denote the CDF of a Gamma(\(\alpha_{ik}^*, \beta_{ik}^*\)) distribution, where Gamma\((a, b)\) denotes a Gamma distribution with probability density function, \( f(x) \propto x^{a-1} \exp(-xb) \), mean \( ab^{-1} \), and variance \( ab^{-2} \). The bootstrap versions of the direct estimators of the totals in province \(k\) are \( \{\hat{M}_{1k}^*, \ldots, \hat{M}_{Ck}^*\} \), where \( \hat{M}_{ik}^* \) satisfies \( G_{ik}^*(\hat{M}_{ik}^*) = W_{ik}^* \). The bootstrap versions of the direct estimators of the proportions are \( \{\hat{p}_{ik}^* : i = 1, \ldots, C; k = 1, \ldots, K\} \), where \( \hat{p}_{ik}^* = \hat{M}_{ik}^* (\hat{M}_{ik}^*)^{-1} \), and \( \hat{M}_k^* = \sum_{i=1}^C \hat{M}_{ik}^* \).

3. Let \( df_k = \max\{[n_k \hat{e}_k^{-1} - 1], 1\} \), where \(n_k\) is the realized sample size in province \(k\), and the notation \([a]\) denotes the integer closest to \(a\). Then, let \( \hat{\Sigma}_{ee,k}^{(1)*} = [df_k(df_k + 1)]^{-1} B^{(1)*} \),
where $B^{(1)*}$ is a Wishart matrix with $df_k$ degrees of freedom, and

$$E[\hat{\Sigma}_{ee,k}^{(1)*} | \hat{\Sigma}_{ee,k,md}] = \hat{\Sigma}_{ee,k,md}.$$  \hfill (5.30)

where the notation $A^{(1)}$ denotes the matrix obtained by omitting the first row and column of the matrix $A$. Let $\hat{\Sigma}_{ee,k}^*$ be the symmetric $C \times C$ matrix such that the $(C-1) \times (C-1)$ submatrix formed from the last $C-1$ rows and columns of $\hat{\Sigma}_{ee,k}^*$ is equal to $\hat{\Sigma}_{ee,k}^{(1)*}$, and the rows and columns of $\hat{\Sigma}_{ee,k}^*$ sum to zero. To create a bootstrap version of the direct estimator of the sampling variance of the totals, let $\hat{\Sigma}_{ee,k,md}^*$ be the symmetric $C \times C$ matrix such that the $(C-1) \times (C-1)$ submatrix formed from the last $C-1$ rows and columns of $\hat{\Sigma}_{ee,k}^*$ is equal to $\hat{\Sigma}_{ee,k,md}^{(1)*}$, and the rows and columns of $\hat{\Sigma}_{ee,k}^*$ sum to zero. To create a bootstrap version of the direct estimator of the sampling variance of the totals, let $\hat{V}^*\{\hat{M}_k\} = \hat{c}_k n_k^{-1} (\hat{M}_k^*)^2$. Then, define the bootstrap version of the direct estimator of the variance of $a_k$ by

$$\hat{V}^*\{a_k\} = \hat{\Sigma}_{ee,k}^* (\hat{M}_k^*)^2 + \hat{p}_k^* (\hat{p}_k^*)' \hat{V}^*\{\hat{M}_k\}.$$  

5.2.2 Definitions of Bootstrap MSE Estimators

Given the generated $\hat{M}_{ik}^*$, $\hat{p}_{ik}^*$, $\hat{V}^*\{a_k\}$, and $\hat{\Sigma}_{ee,k}^*$, bootstrap versions of estimators of the model parameters and predictors are calculated using the procedure of Section 4. Let $\hat{p}_{ik,B}$, $\hat{\psi}^*$, and $\hat{c}_k^*$ be the bootstrap versions of the estimators of the model parameters, and let $\hat{p}_{ik,B}$ and $\hat{M}_{ik,B}$ be bootstrap predictors calculated with the Beale estimator,

$$\hat{\gamma}_{ik,B} = \frac{\hat{\psi}^* (\hat{\psi}^* + \hat{c}_k^* n_k^{-1}) + \hat{V}^* (\hat{\psi})}{(\hat{\psi}^* + \hat{c}_k^* n_k^{-1})^2 + \hat{V}^* (\hat{\psi}) + n_k^{-2} \hat{V}^* (\hat{\psi})},$$

where $\hat{V}^* (\hat{\psi})$ and $\hat{V}^* (\hat{c}_k)$ are the estimates of the variances of $\hat{\psi}$ and $\hat{c}_k$ defined in (4.28) and (4.30), respectively, calculated with the bootstrap data. For each bootstrap sample, the bootstrap versions of the leading terms in the MSE approximation are

$$\hat{MSE}_{11,ik}^* = (\hat{\gamma}_{ik,B}^*)^2 \hat{\sigma}_{e,ik,md}^2 + (1 - \hat{\gamma}_{ik,B}^*)^2 \hat{\psi}^* \hat{p}_{T,ik}^* (1 - \hat{p}_{T,ik}^*),$$

$$\hat{MSE}_{11,ik,T}^* = (\hat{\gamma}_{ik,B}^*)^2 \hat{V}^*\{a_{ik}\} + (1 - \hat{\gamma}_{ik,B}^*)^2 \hat{\psi}^* \hat{p}_{T,ik}^* (1 - \hat{p}_{T,ik}^*) (\hat{M}_k^*)^2,$$

where $\hat{\sigma}_{e,ik}^2$ and $\hat{V}^*\{a_{ik}\}$ are the diagonal elements of $\hat{\Sigma}_{ee,k}^*$ and $\hat{V}^*\{a_k\}$, respectively, corresponding to category $i$, and $\hat{\Sigma}_{ee,k}^*$ and $\hat{V}^*\{a_k\}$ are generated in step 3.

To define a bootstrap estimator of the MSE of the predictor of the proportion, let $E_*$ denote expectation with respect to the bootstrap distribution defined by steps (1-3) above.
Define the bootstrap estimator of the MSE of the predictor of the proportion by

$$
\text{MSE}_{1,ik}^{bs} = \begin{cases} 
\text{MSE}_{1,ik}^{bs} - b_{ik}^{bs} & \text{if } \text{MSE}_{1,ik}^{bs} > b_{ik}^{bs} \\
2\text{MSE}_{1,ik}^{bs}[1 + \exp\{2b_{ik}^{bs}(\text{MSE}_{1,ik}^{bs})^{-1}\}]^{-1} & \text{otherwise,}
\end{cases}
$$

(5.31)

where

$$
\text{MSE}_{1,ik}^{bs} = E[(\hat{p}_{ik,B}^* - p_{ik}^*)^2],
$$

(5.32)

$$
b_{ik}^{bs} = E[\hat{MSE}_{11,ik}^{*}] - \hat{MSE}_{11,ik},
$$

(5.33)

and $\hat{MSE}_{11,ik} = \hat{\gamma}_{k,B}^2\hat{\sigma}_{e,ik}^2 + (1 - \hat{\gamma}_{k,B})^2\hat{\psi}\hat{p}_{T,ik}(1 - \hat{p}_{T,ik})$, as defined in (5.11). A bootstrap estimator of the MSE of the predictor of the total for three digit code $i$ and province $k$ is defined analogously by replacing $\hat{p}_{ik,B}^*$, $p_{ik}^*$, $\hat{MSE}_{11,ik}^{*}$, and $\hat{MSE}_{11,ik}$ with $\hat{M}_{ik,B}^*$, $M_{ik}^*$, and $\hat{MSE}_{11,ik}^{*}$, and $\hat{MSE}_{11,ik,T}$, and $\hat{MSE}_{11,ik,T}$, where

$$
\hat{MSE}_{11,ik,T} = \hat{\gamma}_{k,B}^2V(a_{ik}) + (1 - \hat{\gamma}_{k,B})^2\hat{\psi}\hat{p}_{T,ik}(1 - \hat{p}_{T,ik})\hat{M}_k^2.
$$

(5.34)

**Remark 1:** A heuristic justification for the form of the estimator in (5.31) is as follows. Defining $\hat{p}_{ik}^*(\hat{\lambda}, \hat{\psi}, \hat{c}_k) = \hat{\gamma}_{k,B}\hat{\lambda}_{ik}^* + (1 - \hat{\gamma}_{k,B})\hat{\psi}\hat{p}_{T,ik}(1 - \hat{p}_{T,ik})$, one can express $\text{MSE}_{1,ik}^{bs}$ as

$$
\text{MSE}_{1,ik}^{bs} = E_*[(p_{ik,B}^* - p_{ik}^*)^2] + E_*[(\hat{p}_{ik}^*(\hat{\lambda}, \hat{\psi}, \hat{c}_k) - p_{ik}^*)^2]
$$

$$
+ 2E_*[(\hat{p}_{ik,B}^* - \hat{p}_{ik}^*(\hat{\lambda}, \hat{\psi}, \hat{c}_k))(\hat{p}_{ik}^*(\hat{\lambda}, \hat{\psi}, \hat{c}_k) - p_{ik}^*)].
$$

(5.35)

If $E_*[(\hat{p}_{ik}^* - p_{ik}^*)^2] = \Sigma_{ee,k}$, then $E_*[(\hat{p}_{ik}^*(\hat{\lambda}, \hat{\psi}, \hat{c}_k) - p_{ik}^*)^2] = \hat{MSE}_{11,ik}$. A bootstrap estimator of the bias of $\hat{MSE}_{11,ik}$ for $\text{MSE}_{11,ik}$ is $E_*[\hat{MSE}_{11,ik}^{*}] - \hat{MSE}_{11,ik}$. A Taylor expansion of the second line in (5.31) around $E_*[\hat{MSE}_{11,ik}^{*}] - \hat{MSE}_{11,ik} = 0$ shows that as the bootstrap estimate of the bias approaches zero, the difference between the two estimators in (5.31) approaches zero.

**Remark 2:** Hall and Maiti (2006) and Chatterjee et al. (2007) suggest bias corrections that guarantee positive MSE estimates. The function in the second line of (5.31) is suggested in Chatterjee et al. (2007). A Taylor expansion of the second line in (5.31) around $E_*[\hat{MSE}_{11,ik}^{*}] - \hat{MSE}_{11,ik} = 0$ shows that as the bootstrap estimate of the bias approaches zero, the difference between the two estimators in (5.31) approaches zero.

The definition of the bias corrected bootstrap MSE estimator in (5.31) differs from the bias correction in Chatterjee et al. (2007) in two main ways. One, the bias defined in
(5.37) is an estimate of the bias of \( \hat{MSE}_{11,ik} \) for the MSE of a predictor constructed with the true parameters. In contrast, Chatterjee et al. (2007) use the double bootstrap to obtain an estimate of the bias of an estimator of the MSE of a predictor constructed with the unknown parameters. A second difference is that we only apply the multiplicative bias correction in the second line of (5.31) when the additive bias estimate in the first line is negative. Chatterjee et al. (2007) use the function in the second line of (5.31) to define an MSE estimator regardless of the sign of \( \hat{MSE}^{bs}_{1,ik} - b^{bs}_{ik} \). Because our estimator of the bias is an estimator of the bias of the estimator of the leading term, we prefer the additive bias correction and only use the multiplicative bias correction when the additive bias correction leads to a negative MSE estimate. The multiplicative bias correction is not needed in the LFS application of Section 7. The multiplicative bias correction is only needed in one of the four simulation scenarios of Section 6, where the multiplicative bias correction is used in less than 0.1% of the MC samples.

5.2.3 Computation of Bootstrap MSE Estimators

We approximate the expected values defined in (5.31) with Monte Carlo averages as follows:

1. For \( b = 1, \ldots, B \),
   
   i. Generate \( \hat{M}_{ik}^{*(b)}, \hat{M}^{*(b)}_{ik}, \hat{\Sigma}^{*(b)}_{ee,k}, \) and \( \hat{V}^{*(b)}\{a_k^*\} \) using the method described in steps 1-3 above.

   ii. Use the bootstrap data generated in step i to compute estimators of the model parameters, predictors, and estimators of the leading terms in the MSE approximation. Denote the true proportion, predictor, and estimator of the MSE of the leading term obtained in the \( b^{th} \) bootstrap sample by \( \hat{p}_{ik}^{*(b)}, \hat{p}_{ik}^{*(b)}, \) and \( \hat{MSE}_{11,ik}^{*(b)} \), respectively.

2. The bootstrap estimator of the MSE of the predictor of proportion for three digit code
i and province k is defined,

\[
\hat{\text{MSE}}_{1,ik}^{bs} = \begin{cases} 
\hat{\text{MSE}}_{1,ik}^{bs} - \hat{b}_{ik}^{bs} & \text{if } \hat{\text{MSE}}_{1,ik}^{bs} > \hat{b}_{ik}^{bs} \\
2\hat{\text{MSE}}_{1,ik}[1 + \exp\{2\hat{b}_{ik}^{bs}(\hat{\text{MSE}}_{1,ik}^{bs})^{-1}\}]^{-1} & \text{otherwise},
\end{cases}
\]  

(5.36)

where

\[
\hat{\text{MSE}}_{1,ik}^{bs} = \frac{1}{B} \sum_{b=1}^{B} (\hat{p}_{ik,B}^* - \hat{p}_{ik})^2,
\]

and

\[
\hat{b}_{ik}^{bs} = \frac{1}{B} \sum_{b=1}^{B} \hat{\text{MSE}}_{1,ik}^{*,(b)} - \hat{\text{MSE}}_{11,ik}^{*}.
\]  

(5.37)

A bootstrap estimator of the MSE of the total is obtained analogously by substituting \(\hat{p}_{ik,B}, \hat{p}_{ik}^{*,(b)}, \hat{\text{MSE}}_{11,ik}^{*,(b)}, \) and \(\hat{\text{MSE}}_{11,ik}^{*}\) with \(\hat{p}_{ik,B}, \hat{M}_{ik,B}, \hat{M}_{ik}^{*,(b)}, \hat{\text{MSE}}_{11,ik,T}^{*,(b)}\), and \(\hat{\text{MSE}}_{11,ik,T}\), respectively. Let \(\hat{\text{MSE}}_{ik,T}^{bs}\) denote the resulting estimator of the MSE of the total for three digit code i and province k.

### 5.2.4 Moments of Bootstrap Distributions of Proportions and Totals

The bootstrap versions of the true values satisfy

\[
E_*[p_k^*] = \hat{\text{P}}_{T,k},
\]

\[
V_*\{p_k^*\} = \hat{\text{P}}_{uu,k}\hat{\psi},
\]

and \(E_*\) and \(V_*\) denote the conditional mean and variance, respectively, given the original data. Under the distribution generated in step 2,

\[
E_*[\hat{M}_{ik}^* \mid M_{ik}^*] = M_{ik}^*,
\]

and

\[
V_*\{\hat{M}_{ik}^* \mid M_{ik}^*\} = \hat{V}_{D,ik}.
\]
The distribution generated in step 2 has the desired means and variances in the scale of totals. The method does not preserve the covariances between sampling errors for totals for two different categories;

\[ C_s\{\hat{M}^*_ik, \hat{M}^*_jk | M^*_ik, M^*_jk\} \neq \hat{V}_{D,ij,k}, \]

where \( \hat{V}_{D,ij,k} \) denotes the covariance in row \( i \) and column \( j \) of \( \hat{\Sigma}_{aa,md} \), and \( \hat{\Sigma}_{aa,md} \) is defined in (4.6). Erhardt and Czado (2009) propose a way to generate multivariate count data with a covariance matrix approximately equal to an arbitrary prespecified covariance matrix. Their algorithm is implemented in the R package “corcounts.” Their method requires solving a system of equations, which can be slow and is not guaranteed to converge. Therefore, we prefer the simple but approximate method described in step 2. In Appendix 4, a particular \( \hat{\Sigma}_{aa,md} \) is compared to the empirical covariance matrix of \( \hat{M}^*_k - M^*_k \). In the example of Appendix 4, the covariances of \( \hat{\Sigma}_{aa,md} \) have the same order and sign as the covariances of the empirical covariance matrix of \( \hat{M}^*_k - M^*_k \).

The variance of \( \hat{p}^*_k - p^*_k \) under the bootstrap distribution does not equal \( \hat{\Sigma}_{ee,md} \), where, as defined in (4.7),

\[ \hat{\Sigma}_{ee,md} = \hat{L}_k \hat{\Sigma}_{aa,md} \hat{L}_k', \]

and the matrix \( \hat{L}_k \) is obtained by evaluating the matrix \( L_k \) of the linear approximation (A.2) that converts totals to proportions at the direct estimators. One reason that \( V_s\{\hat{p}^*_k | p^*_k\} \neq \hat{\Sigma}_{ee,md} \) is related to the linear approximation used to obtain \( \hat{\Sigma}_{ee,md} \) from \( \hat{\Sigma}_{aa,md} \). In Appendix 4, it is argued that for small \( \psi \), an approximation for the \( V_s\{\hat{p}^*_k | p^*_k\} \) is

\[ \hat{\Sigma}_{ee,k} = \hat{L}_{T,k} \hat{\Sigma}_{aa,md} \hat{L}_{T,k}', \quad (5.38) \]

where \( \hat{L}_{T,k} \) is obtained by evaluating \( L_k \) (the matrix of the linear approximation of Appendix 1) at the synthetic estimator \( \hat{T}_{ik} = \hat{p}_{T,ik} \hat{M}_k \). (In contrast, \( \hat{\Sigma}_{ee,md} \) is obtained by evaluating the matrix of the linear approximation at the direct estimator.) In Appendix 4, an alternative data generating procedure is proposed such that \( V_s\{\hat{p}^*_k - p^*_k\} \) is closer to \( \hat{\Sigma}_{ee,k,md} \).

One implication of the difference between \( \hat{\Sigma}_{ee,k,md} \) of (4.7) and \( \hat{\Sigma}_{ee,k,md} \) of (5.38) is that the bias estimate, \( \hat{b}_{ik}^b \), defined in (5.37), is not an estimate of the bias of the bootstrap estimator.
of the leading term. The bootstrap estimator of the leading term is \( E_\ast[(\hat{p}^\ast_{ik}(\hat{\lambda}, \hat{\psi}, \hat{c}_k) - p^\ast_{ik})^2] \).

Because \( V_\ast\{\hat{p}^\ast_k - p^\ast_k\} \neq \hat{\Sigma}_{ee,k,md} \), the bootstrap estimator of the leading term is not equal to \( \hat{MSE}_{11,ik} \).

Because the totals satisfy \( E_\ast[(\hat{M}^\ast_{ik} - M^\ast_{ik})^2] = \hat{V}_{D,ik} \), the bootstrap estimator of the leading term for the totals is the same as the Taylor estimator of the leading term for the totals. That is,

\[
\hat{MSE}_{11,ik,T} = E_\ast[(\hat{\gamma}_{k,B}\hat{M}^\ast_{ik} + (1 - \hat{\gamma}_{k,B})\hat{T}_{ik} - M_{ik})^2],
\]

where \( \hat{MSE}_{11,ik,T} \) is defined in (5.34). As a consequence, the bootstrap estimator of the bias of the estimator of the leading term in the MSE for the totals is an estimator of the bias of either the bootstrap or the Taylor estimator of the leading term.

**5.2.5 Bootstrap Distributions of Direct Estimators of Sampling Variances**

The role of the direct estimator of the sampling variance in the bootstrap MSE estimator is analogous to the role of the direct estimator of the sampling variance in the MSE estimator of Wang and Fuller (2003). In constructing an initial estimate of leading term in the MSE, the direct estimators, \( \hat{\Sigma}_{aa,k,md} \) and \( \hat{\Sigma}_{ee,k,md} \), are treated as consistent estimators of the sampling variances. The direct estimators are then assumed to have Wishart distributions for estimation of the bias of the initial estimator of the leading term and for estimation of the variance of \( \hat{\gamma}_{k,B} \).

In step 3 of the bootstrap data generating procedure, the direct estimator of the sampling covariance matrix is assumed to have a Wishart distribution that is independent of the bootstrap estimators of the proportions and totals. The independent Wishart model is natural for linear models with normally distributed errors. Other distributions for the direct estimators of the sampling variances may be preferable in situations where estimated means are related to estimated variances. One example of an alternative distribution for \( \hat{\Sigma}_{ee,k}^\ast \) is to generate \( \hat{\Sigma}_{ee,k}^{\ast(1)} \) from a Wishart distribution with mean \( \hat{c}_k n_k^{-1}[\text{diag}(\hat{p}_k) - \hat{p}_k(\hat{p}_k)'] \). This possibility was motivated by a linear relationship between the elements of \( n_k^{-1}[\text{diag}(\hat{p}_k) - \hat{p}_k(\hat{p}_k)'] \) and the
direct estimators of the sampling variances of the proportions observed in the LFS. A different possibility is suggested in Section 5.2.7 below.

The bootstrap versions of the direct estimators of the sampling variances affect the MSE estimators in three ways. One, the bootstrap version of the estimator of \( c_k \) is constructed with the bootstrap version of the direct estimator of the sampling covariance matrix. The synthetic estimators depend on the estimator of the sampling covariance matrix only through the estimator of \( c_k \). The direct estimator of the sampling covariance matrix affects the estimator of \( \psi \) through the estimator of \( \Sigma_a \) defined in (4.14). Consequently, we expect a change in the distribution used for the bootstrap version of the direct estimator of the sampling covariance matrix to have a more noticeable impact on the bootstrap estimator of the variance of \( \tilde{c}_k \) than on the variances of the other model parameters. The estimators of the sampling variances are used directly in the estimators of the leading terms in the MSE approximation. We expect the choice of the distribution for \( \tilde{\Sigma}_{ee,k}^* \) to impact the MSE estimator most directly through the estimator of the bias of the estimator of the leading term. In limited simulation studies (not discussed in this thesis), changing the distribution used to generate \( \tilde{\Sigma}_{ee,k}^* \) has a substantial impact on the bootstrap estimates of the variance of \( \tilde{c}_k \) but does not lead to significant differences in the MC biases of the MSE estimators. A more detailed investigation may be warranted.

In the LFS application, the direct estimators of the sampling variances of the direct estimators of the province two digit totals are linearly related to the squares of the direct estimators of the province two digit totals. (See Section 7.4.) The bootstrap versions of the direct estimators of the variances of the province two digit totals are defined in step 3 in an attempt to approximate the relationship between the means and the variances observed in the LFS. The choice of the bootstrap distribution used for the sampling variances does not directly affect the bootstrap estimator of the leading term in the MSE because the original direct estimators of the sampling covariance matrices are used to generate the bootstrap versions of the direct estimators of the totals and proportions.
5.2.6 Details on Implementation

Remark 1: When \( \hat{\beta}_{T,ik} \) is close to zero, \( \alpha_{ik}^* \) of (5.28) is often close to zero. In our implementation of the bootstrap data generating procedure, we use the R function, “qgamma” to compute the quantile function of the gamma distribution. When \( \alpha_{ik}^* \) is sufficiently small, the “qgamma” function returns a missing value. The numerical problems arise when the generated \( M_{ik}^* \) is close to zero. To avoid generating missing values, we set \( \hat{M}_{ik}^* = M_{ik}^* \) if the R function returns a missing value. By setting the bootstrap version of the direct estimator of the total equal to \( M_{ik}^* \), we preserve the property that \( E_s[\hat{M}_{ik}^* | M_{ik}^*] = M_{ik}^* \). The modification, however, distorts the variance. The R function “qgamma” is more prone to numerical problems for smaller values of \( \hat{\beta}_{T,ik} \) and larger values of \( \hat{\psi} \). The fractions of samples that lead to numerical problems observed in the simulations and in the LFS example are reported in Section 6.2 and Section 7.3, respectively.

Several alternatives to the ad hoc procedure of setting \( \hat{M}_{ik}^* = M_{ik}^* \) in cases where a missing value is returned are possible and may lead to closer approximations of the first and second moments. If the value of \( \alpha_{ik}^* \) is close to zero, and the value of \( \beta_{ik}^* \) is moderate, then the resulting gamma distribution puts large probabilities near zero. As a consequence, one natural alternative is to set \( \hat{M}_{ik}^* = 0 \) if the R function returns a missing value. A potential danger in modifying the generated variable if the R function returns a missing value instead of using a prespecified upper bound for \( \alpha_{ik}^* \) is that the R function might return a missing value for a reason other than \( \alpha_{ik}^* \) being too small, and we would not detect the cause of the problem. An alternative to setting the bootstrap direct estimator equal to the bootstrap true value if the R function returns a missing value is to set a lower bound for \( \alpha_{ik}^* \). For example, we experimented with a procedure in which we set \( \hat{M}_{ik}^* = M_{ik}^* \) if \( \alpha_{ik}^* < 0.01 \). The choice of 0.01 as the lower bound was selected on the basis of informal experiments with a two-way table in which the probability in the cell for three digit code 4 in the province representing Quebec is 0.0404. This procedure assumes that any value of \( \alpha_{ik}^* > 0.01 \) will not lead to numerical problems. Instead of setting the bootstrap direct estimator equal to the true value, one could define \( \hat{M}_{ik}^* \) to satisfy \( G_{ik}^*(\hat{M}_{ik}^*) = W_{ik}^* \), where \( G_{ik}^* \) is the cdf of a gamma distribution with
\( \alpha_{ik}^* = 0.01 \) and \( \beta_{ik}^* = 0.01(M_{ik}^*)^{-1} \) if the original \( \alpha_{ik}^* \) is smaller than 0.01. Using this method, 
\[ E_*(\hat{M}_{ik}^* | M_{ik}^*, \hat{V}_{D,ik}) = M_{ik}^*, \text{ and } V_*(\hat{M}_{ik}^* | M_{ik}^*, \hat{V}_{D,ik}) = 100(M_{ik}^*)^2. \]

Another alternative is to generate \( \hat{M}_{ik}^* \) from a gamma distribution with mean \( \hat{T}_{ik} \) and variance \( \hat{\psi}_p \hat{p}_{T,ik}(1 - \hat{p}_{T,ik}) \hat{M}_{ik}^2 + \hat{V}_{D,ik} \), where \( \hat{T}_{ik} \) is the synthetic estimator of the total for three digit code \( i \) and province \( k \).

**Remark 2:** An iterative algorithm for calculating the synthetic estimators based on a bootstrap data set may not converge. We use the following procedure to obtain bootstrap estimators if numerical problems arise when the estimation procedure is applied to a bootstrap data set. The initial estimators are defined as the solutions to the Poisson score function. If the iterative procedure used to obtain the initial estimators based on a bootstrap data set does not converge, then define the bootstrap version of the initial estimator by,

\[
\hat{\psi}_{0*}^{(1)} = \exp(x_{ik}' \hat{\lambda}_1^{0*}) \sum_{i=1}^C \exp(x_{ik}' \hat{\lambda}_1^{0*}),
\]

where

\[
\hat{\lambda}_1^{0*} = \hat{\lambda} + \left( \sum_{k=1}^K (X_k^{(1)})' \hat{T}_{uu,k} \hat{M}_k X_k^{(1)} \right)^{-1} \sum_{k=1}^K (X_k^{(1)})' \hat{M}_k (\hat{p}_{k}^{*-(1)} - \hat{p}_{T,k}^{(1)}),
\]

\( \hat{p}_{k}^{*-(1)} = (\hat{p}_{1,k}^{*-(1)}, \ldots, \hat{p}_{C-1,k}^{*-(1)})' \), and \( \hat{p}_{T,k} \) is the vector of synthetic estimators for province \( k \) based on the original data. After obtaining the initial values, it is still possible for the iteratively reweighted least squares procedure used to define the final estimator of \( p_{T,ik} \) to fail. If the iteratively reweighted least squares algorithm does not converge, then the bootstrap version of the estimator of \( p_{T,ik} \) is

\[
\hat{\psi}_{T,ik}^{*-(1)} = \exp(x_{ik}' \hat{\lambda}_1^*) \sum_{i=1}^C \exp(x_{ik}' \hat{\lambda}_1^*),
\]

where

\[
\hat{\lambda}_1^* = \hat{\lambda} + \left( \sum_{k=1}^K (X_k^{(1)})' \hat{T}_{uu,k} (\hat{c}_k n_k^{-1} + \hat{\psi})^{-1} X_k^{(1)} \right)^{-1} \sum_{k=1}^K (X_k^{(1)})' (\hat{c}_k n_k^{-1} + \hat{\psi})^{-1} (\hat{p}_{k}^{*-(1)} - \hat{p}_{T,k}^{(1)}).
\]

For a given estimator of \( p_{T,ik} \), the estimator of \( \psi \) can be calculated in one step. Nonetheless, numerical problems can arise in the estimator of \( \psi \) based on the bootstrap data. For example, the bootstrap estimator of \( \{\Sigma_a\}^{*2} \) or \( \{\Sigma_a + \xi \Sigma_b\}^{*2} \) may be singular, where the notation \( \{A\}^{*2} \)
denotes the matrix containing the squares of the elements of the matrix $A$. If one can not compute the estimator of $\psi$ using the bootstrap data, then define the bootstrap version of the estimator of $\psi$ by

$$\hat{\psi}^* = \max\{\xi, \tilde{\psi}^*\}$$

where $\xi$ is the lower bound defined following (4.16) obtained in the last step of the iteration with the original data, and

$$\tilde{\psi}^* = \frac{\hat{\sigma}_b'[2(\hat{\Sigma}_a + \hat{\psi}\hat{\Sigma}_b)]^{-1}[\{\tilde{\mathbf{p}}^* - \hat{\mathbf{p}}^*_T\}^2 - \hat{\sigma}_{a,\text{dir}}]}{\hat{\sigma}_b[2(\hat{\Sigma}_a + \hat{\psi}\hat{\Sigma}_b)]^{-1/2}}.$$ 

$\hat{\mathbf{p}}^* = ((\hat{\mathbf{p}}^*_1)', \ldots, (\hat{\mathbf{p}}^*_K)')'$, $\hat{\mathbf{p}}_T = ((\hat{\mathbf{p}}^*_{T,1})', \ldots, (\hat{\mathbf{p}}^*_{T,K})')'$, the matrices $\hat{\Sigma}_a$ and $\hat{\Sigma}_b$ are based on the original data, and the notation $\{A\}^2$ denotes the matrix containing the squares of the elements of $A$. Shao and Tu (1996, pg. 343) suggest using one step of an iterative algorithm, as above, to reduce computational burden when implementing the bootstrap. Lohr and Rao (2009) suggest using two steps of a Newton-Raphson algorithm starting at the original estimator.

**Remark 3:** If a direct estimator for a cell is zero, then $\hat{\Sigma}_{ee,k,md}^{(1)}$ is singular. If $\hat{\Sigma}_{ee,k,md}^{(1)}$ is singular, then the mean of the Wishart distribution used to generate $\hat{\Sigma}_{ee,k}^*$ in step 3 is defined by

$$\hat{\Sigma}_{ee,k,md,0}^{(1)} = \frac{n_k}{n_k + 2} \hat{\Sigma}_{ee,k,md}^{(1)} + \frac{2}{n_k + 2} \hat{\Sigma}_{uu,k}^{(1)} \hat{\gamma}_{k}.$$ 

A justification for the form of $\hat{\Sigma}_{ee,k,md,0}^{(1)}$ is as follows. If $\hat{\mathbf{V}}$ is proportional to a Wishart matrix with mean $\mathbf{V}$, then an improved estimator of the inverse of $\mathbf{V}$ is $(\hat{\mathbf{V}} + 2n_k^{-1}\mathbf{V})^{-1}$. Wang and Fuller (2003) use an estimator of the form $(\hat{\mathbf{V}} + 2n_k^{-1}\mathbf{V})^{-1}$ for estimated generalized least squares estimation of the parameters in a regression model. Because we are using the covariance matrix for variance estimation, we rescale the covariance matrix so that the trace of the modified estimator, $\hat{\Sigma}_{ee,k,md,0}^{(1)}$, is equal to the trace of the modified direct estimator, $\hat{\Sigma}_{ee,k,md}^{(1)}$. 
5.2.7 Simplifications to Bootstrap MSE Estimator when Working Model for Sampling Variances Holds

The bootstrap data generating procedure described in steps 1-3 of Section 5.2.1 is for an unstructured estimate of the sampling covariance matrix. If the working model for the sampling covariance matrix of (5.15) is assumed to hold, then a simpler method can be used to generate the bootstrap versions of the direct estimators of the proportions and totals. Let the estimator of the sampling covariance matrix be

$$\hat{\Sigma}_{ee,k,w} = \frac{c_k}{n_k} \left[ \text{diag}(\hat{p}_T,k) - \hat{p}_T,k \hat{p}_T,k' \right].$$

Under the assumption that $\hat{\Sigma}_{ee,k,w}$ is a consistent estimator of $\Sigma_{ee,k} = E\left[V\{e_k \mid u_k\}\right]$, the bootstrap versions of the direct estimators can be generated to preserve the covariance structure.

1a. Let $p^*_k$ and $M^*_k$ be generated according to the method in step 1 of Section 5.2.1 titled “Bootstrap Distributions.”

2a. For $k = 1, \ldots, K$, let $d_k > 1$ be an integer such that $n_k(1 - \hat{\psi})(d_k\hat{c}_k)^{-1} > 1$. For $l = 1, \ldots, d_k$, let $\hat{p}^*_l,k \sim \text{Dirichlet}(p^*_k, n_k(1 - \hat{\psi})(d_k\hat{c}_k)^{-1} - 1)$. Let $\hat{p}^*_k = d_k^{-1}\sum_{l=1}^{d_k} \hat{p}^*_l,k$ be the bootstrap version of the direct estimator of the proportion. To construct bootstrap versions of the direct estimators of the totals, let $\hat{M}^*_k$ have a gamma distribution with mean $\hat{M}_k$ and variance $\hat{V}\{\hat{M}_k\}$, where $\hat{V}\{\hat{M}_k\}$ is the direct estimator of the variance of the direct estimator of the two digit total in province $k$, and set $\hat{M}^*_k = \hat{p}^*_k\hat{M}_k$.

3a. Let the bootstrap version of the direct estimator of $V\{e_k\}$ be

$$\hat{\Sigma}_{ee,k}^* = \frac{1}{d_k - 1} \sum_{l=1}^{d_k} (\hat{p}^*_l,k - \hat{p}^*_k)(\hat{p}^*_l,k - \hat{p}^*_k)',$$

where $\hat{p}^*_l,k$ are generated in step 2a. Define the bootstrap version of the direct estimator of the sampling covariance matrix for the vector of totals in province $k$ by

$$\hat{\Sigma}_{aa,k}^* = (\hat{M}^*_k)^2\hat{\Sigma}_{ee,k}^* + (\hat{p}^*_k)(\hat{p}^*_k)'\hat{V}^*\{a_{ik}\}.$$

Remark: The direct estimators of the proportions generated through steps 1a - 3a satisfy

$$E_*[\hat{p}^*_k \mid \hat{p}^*_l,k : l = 1, \ldots, d_k] = p^*_k,$$
and
\[
V_*\{\hat{p}_{k}^* \mid p_{l,k}^* : l = 1, \ldots, d_k\} = \frac{\hat{c}_k}{(1 - \hat{\psi})n_k} \left[ \text{diag}(p_{l,k}^*) - p_{l,k}^*(p_{l,k}^*)' \right].
\]

It follows that
\[
E_*[\hat{p}_{k}^*] = \hat{p}_{T,k}^*.
\]

and
\[
V_*\{\hat{p}_{k}^* \} = \frac{\hat{c}_k}{n_k} \left[ \text{diag}(\hat{p}_{T,k}^*) - \hat{p}_{T,k}^*(\hat{p}_{T,k}^*)' \right].
\]

Remark: The method in step 3a used to generate the bootstrap version of the direct estimator of the sampling covariance matrix of the proportions differs from step 3 of Section 5.2.1 for the general sampling covariance matrix. The covariance matrix obtained in step 3a is not a Wishart matrix. The estimated variances from step 3a are not independent of the estimated means obtained in step 2a. The number of proportions used to form the variance estimate, \(d_k\), has an interpretation as a degrees of freedom. In some simulations not reported in this thesis, we used \(d_k = n_k \hat{c}_k^{-1}\). Other choices for \(\hat{\Sigma}_{ee,k}^*\) are possible. For example, in one of the simulation scenarios in Section 6, \(\hat{\Sigma}_{ee,k}^* = \hat{c}_kn_k^{-1}[\text{diag}(\hat{p}_{k}^*) - (\hat{p}_{k}^*)(\hat{p}_{k}^*)']\).

Remark: The method in step 2a used to generate the bootstrap versions of the proportions and totals preserves the working covariance structure. In contrast, the variance of \(\hat{p}_{ik}^* - p_{ik}^*\) generated through step 2 of Section 5.2.1 is not equal to the direct estimator of the variance of sampling error in the proportion.

5.3 Benefits and Drawbacks of Taylor and Bootstrap MSE Estimators

In the previous two subsections, two estimators of the MSE’s of the predictors of the proportions and totals were discussed. In Section 5.1, a closed form approximation for the MSE of the “initial predictor” (the predictor that is not benchmarked to the direct estimators of the marginal totals) is derived using Taylor linearizations. We refer to the closed form estimator of the MSE of the initial predictor as the “Taylor” MSE estimator. In Section 5.2, a simulation based MSE estimator related to the wild bootstrap procedure (Wu, 1986) is
proposed. We refer to the MSE estimators of Section 5.2 as bootstrap MSE estimators. Each MSE estimator has benefits and drawbacks.

One advantage of the Taylor MSE estimator relative to the bootstrap MSE estimator is that the Taylor MSE estimator respects the specified covariance structures for both the random effects $u$ and the sampling errors $e$. In contrast, the data generating procedure used for the bootstrap distorts the specified covariance structure for the direct estimators of the proportions and totals. As illustrated in Appendix 4, the distortion of the variances of the proportions can be substantial, especially in small provinces. The distributions suggested in Section 5.2 to generate the bootstrap data are chosen to preserve the specified covariance of $u_k$ and to produce bootstrap versions of the direct estimators that remain in the parameter space. We did not study the effects of specific distributional assumptions in detail.

A second advantage of the Taylor MSE estimator is computational simplicity. The bootstrap MSE estimator requires additional computing time, and computation for a large range of samples leads to potential convergence problems. For example, when $\hat{p}_{T,ik}$ is sufficiently small, the shape parameter of the gamma distribution used to generate the direct estimators is often small enough that the R quantile function returns a missing value. The frequencies with which the R function returns missing values in the simulations and the LFS application are reported in Sections 6.2 and 7.3, respectively.

A third advantage of the Taylor procedure involves the estimated variance of the estimator of $\hat{c}_k$, which affects the MSE estimate through the estimate of the variance of $\hat{\gamma}_k$ and the estimate of the bias of the estimate of the leading term. The Taylor estimator of the variance of $\hat{c}_k$ relies less on the assumption that the sampling variances have Wishart distributions than does the bootstrap variance estimator.

A potentially important benefit of the bootstrap procedure is that the bootstrap MSE estimator accounts for the raking operation, while the Taylor MSE estimator does not. In Section 6, the contribution of raking to the MSE of the predictor is investigated empirically. In situations where raking leads to a substantial increase in the MSE of the predictor, the Taylor MSE estimator has the potential to have a substantial negative bias.
The bootstrap estimator of the variance of the estimator of $\psi$ uses the fourth moments of the gamma distribution instead of the fourth moments of the normal distribution. The gamma seems to be more appropriate for count data. Also, the bootstrap accounts for the effect of the lower bound.

A final benefit of the bootstrap is that one does not need to derive separate expressions for the MSE’s of the predictors of the totals and the proportions. The bootstrap produces estimates for both totals and proportions from a single set of bootstrap samples. Similarly, the bootstrap furnishes estimates of covariances between prediction errors for two different cells. This allows us to construct intervals for differences without needing to derive expressions for the covariances. This attribute was used in Section 7.
CHAPTER 6. Simulation

The small area prediction procedures are evaluated through simulation. The models and parameters used in the simulation are based on preliminary analyses of the LFS data for the two digit codes A1 and E0. Two distributions for generating the sampling errors are used: one represents a simple random sample and the other represents a two stage design. The models used for the simulation are described and the results presented.

6.1 Simulation Models

To generate vectors of true proportions with the desired first and second moments, let \( \omega_o = \psi^{-1} - 1 \) and \( \omega_{ik} = p_{T,ik}\omega_o \). Then, the vector of true proportions \( p_k \) is generated from a Dirichlet distribution with probability density function

\[
P(v_2, \ldots, v_{Ck}) = \left[ \frac{\Gamma(\omega_o)}{\prod_{i=1}^{C} \Gamma(\omega_{ik})} \right] \prod_{i=1}^{C} v_{ik}^{\omega_{ik}-1},
\]

where \( v_{ik} > 0 \) (\( i = 2, \ldots, C \)), \( \sum_{i=2}^{C} v_{ik} < 1 \), \( v_{1k} = 1 - \sum_{i=2}^{C} v_{ik} \), and \( \Gamma(a) = \int_0^\infty t^{a-1}e^{-t}dt \). We use the notation \( p_k \sim \text{Dirichlet}_C(p_{T,k}, \omega_o) \) to indicate that \( p_k \) has a distribution specified by (6.1) with \( (\omega_{1k}, \ldots, \omega_{Ck})' = p_{T,k}\omega_o \). Under the simulation model, \( E[p_k] = p_{T,k} \), and \( V\{p_k\} = \Gamma_{uu,k}\psi \).

The direct estimators of the province two digit totals are generated to have the coefficients of variation observed for the two digit code A1 in the LFS. The LFS estimates of the province totals for the two digit code A1 serve as the true province two digit totals \( \{T_1, \ldots, T_K\} \) in the simulation. Let \( (CV)_k \) denote the LFS estimate of the coefficient of variation of the direct estimator of the two digit total for province \( k \). Let \( \tilde{n}_k \) be a realization from a Poisson distribution with a mean of \( n_k \), where \( n_k = (CV)_k^{-2} \). The \( n_k \) and \( \tilde{n}_k \) represent
expected and realized sample sizes, respectively, for province $k$. (Samples with $\tilde{n}_k = 0$ are not used in the simulation study. The smallest expected sample size is $n_k = 13$, and the probability that $\tilde{n}_k = 0$ when $n_k = 13$ is $2.26 \times 10^{-6}$.) Define the direct estimator of the province two digit total by $\hat{M}_k = T_k \tilde{n}_k n_k^{-1}$, where $T_k$ denotes the true value.

We consider two simulation models for the sampling errors. The first is intended to represent a simple random sample. The second is used in Thomas and Rao (1987) to simulate from distributions that have properties of a two stage sample.

1. Simple Random Sample

Given $\tilde{n}_k$ and $p_k$, the direct estimator of the total in three digit code $i$ and province $k$ depends on a multinomial random vector, $(\tilde{M}_{1k}^{(d)}, \ldots, \tilde{M}_{Ck}^{(d)})' = \tilde{M}_k$, with probability mass function,

$$P(\tilde{M}_{1k}^{(d)} = \tilde{m}_{1k}, \ldots, \tilde{M}_{mk}^{(d)} = \tilde{m}_{Ck}) = \tilde{n}_k!\left(\prod_{i=1}^{C} \tilde{m}_{ik}!\right)^{-1} \prod_{i=1}^{C} (p_{ik})^{\tilde{m}_{ik}}, \quad (6.2)$$

where $\tilde{m}_{1k}, \ldots, \tilde{m}_{Ck}$ are non-negative integers that sum to $\tilde{n}_k$. We use the notation $\tilde{M}_k^{(d)} \sim \text{Multinomial}_{C}(p_k, \tilde{n}_k)$ if $\tilde{M}_k^{(d)}$ has the probability mass function given in (6.2). For $\tilde{n}_k > 0$, the direct estimator of the proportion in the simulation is $\hat{p}_{ik} = \tilde{n}_k^{-1} \tilde{M}_{ik}^{(d)}$. The corresponding direct estimator of the total, $\tilde{M}_{ik}$, is $\tilde{M}_{ik} \hat{p}_{ik}$, where $\tilde{M}_k$ is generated according to the method described above. The direct estimator of the sampling covariance matrix of the vector of proportions in province $k$ in the simulation is

$$\hat{\Sigma}_{ee,k,srs} = \tilde{n}_k^{-1}[\text{diag}(\hat{p}_k) - \hat{p}_k \hat{p}_k'], \quad (6.3)$$

and the direct estimator of the sampling covariance matrix of the vector of totals is

$$\hat{\Sigma}_{aa,k,srs} = \frac{\tilde{M}_k^2}{\tilde{n}_k} \text{diag}(\hat{p}_k). \quad (6.4)$$

A Taylor approximation can be used to relate the direct estimators of the sampling variances of the totals and proportions. More specifically,

$$\hat{\Sigma}_{aa,k,srs} = \hat{\Sigma}_{ee,k,srs} \tilde{M}_k^2 + \hat{p}_k \hat{p}_k' \hat{V} \{\tilde{M}_k\},$$

and

$$\hat{\Sigma}_{ee,k,srs} = (\hat{L}_k)' \hat{\Sigma}_{aa,k,srs} \hat{L}_k,$$
where $\hat{V}\{\hat{M}_k\} = \hat{M}_k^2 \tilde{n}_k^{-1}$, and $\hat{L}_k$ is the matrix of the linear approximation in (A.2) of Appendix 1 for the operation that converts totals to proportions.

**Remark: Properties of SRS Simulation Model**

By properties of multinomial random variables, the direct estimators are conditionally unbiased for the true proportions: $E[\hat{p}_k | \tilde{n}_k, p_k] = p_k$. The conditional covariance matrix of the vector of sampling errors for a province is

$$V\{\hat{p}_k - p_k | p_k, \tilde{n}_k\} = \tilde{n}_k^{-1}[\text{diag}(p_k) - p_k p_k'] := V_{srs}, \quad (6.5)$$

and the unconditional variance is

$$E[V\{\hat{p}_k - p_k | p_k, \tilde{n}_k\}] = E[\tilde{n}_k^{-1}](1 - \psi)[\text{diag}(p_{T,k}) - p_{T,k} p_{T,k}']. \quad (6.6)$$

It follows that the working model for the sampling covariance matrix is the true covariance under the SRS simulation model, and $c_k = n_k E[\tilde{n}_k^{-1}](1 - \psi)$.

Under the model for the simulation,

$$V\{\hat{M}_k - M_k | p_k\} = E[V\{\hat{M}_k - M_k | p_k, \tilde{n}_k\} | p_k] + V\{E[\hat{M}_k - M_k | p_k, \tilde{n}_k] | p_k\} = \frac{T_k^2}{n_k} \text{diag}(p_k).$$

Because $E[\hat{M}_k - M_k | p_k] = 0$,

$$V\{\hat{M}_k - M_k\} = \frac{T_k^2}{n_k} \text{diag}(p_{T,k}).$$

To see that $\hat{\Sigma}_{aa,k,srs}$ is unbiased for $V\{\hat{M}_k - M_k\}$, note that $\hat{M}_k^2 \tilde{n}_k^{-1} \hat{p}_{ik} = n_k \tilde{n}_k^{-2} T_k^2 \hat{p}_{ik}$, and

$$E[\tilde{n}_k \hat{p}_{ik}] = E[E[\tilde{n}_k \hat{p}_{ik} | \tilde{n}_k]] = n_k p_{T,ik}. \quad (6.7)$$

Conditional on $\tilde{n}_k$, the optimal predictor under the Dirichlet-multinomial model used for the simulation is

$$E[p_k | \tilde{n}_k, \hat{p}_k] = \frac{V\{u_{ik}\}}{V\{u_{ik}\} + V\{e_{ik} | \tilde{n}_k\} \hat{p}_{ik}} \hat{p}_{ik} + \frac{V\{e_{ik} | \tilde{n}_k\}}{V\{u_{ik}\} + V\{e_{ik} | \tilde{n}_k\} p_{T,ik}}, \quad (6.7)$$

where $V\{u_{ik}\} = \psi p_{T,ik} (1 - p_{T,ik})$, and $V\{e_{ik} | \tilde{n}_k\} = \tilde{n}_k^{-1} (1 - \psi) p_{T,ik} (1 - p_{T,ik})$. Optimal predictors for exponential quadratic variance function models are linear (Ghosh and Lahiri,
1987 and Ghosh and Maiti, 2004). The univariate predictor defined in (4.24) with \( \hat{\gamma}_k = \hat{\psi}(\psi + \hat{c}_k n_k^{-1})^{-1} \) is an estimator of the optimal predictor defined in (6.7) because \( \hat{c}_k \) is an estimator of \( (1 - \psi) \).

2. Two Stage Sampling Design

The LFS uses a stratified two stage cluster sample, and the estimated design effects are not constant in each province. We use the model in Thomas and Rao (1987) to generate sampling errors such that the covariance matrix of the vector of sampling errors in each province differs from the multinomial covariance matrix. The method involves simulating multinomial random vectors from \( \tilde{r}_k \) clusters, where each cluster probability is a realization of a mixture of two Dirichlet distributions. The parameters defining the mixture distributions are functions of the vector of true proportions \( p_k \), chosen so that the direct estimators of the proportions are conditionally unbiased for the true proportions.

Let \( n_{c,k} \) denote the number of units in each cluster in province \( k \). Set \( \tilde{r}_k = \max\{\tilde{n}_k n_{c,k}^{-1}, 2\} \), where \( \tilde{r}_k \) represents the number of clusters sampled in province \( k \), and \( \tilde{n}_k \) is generated as described above. Generate \( \tilde{r}_k \) cluster probabilities from a mixture of two Dirichlet distributions as follows. Given \( p_k \), define

\[
p_{ik}^A = \frac{p_{ik}}{1 + \delta_{i,k}},
\]

for \( i = 1, \ldots, C - 1 \), where \( \delta_{i,k} > 0 \), and define \( p_{Ck}^A = 1 - \sum_{i=1}^{C-1} p_{ik}^A \). Let \( p_k^A = (p_{1k}^A, \ldots, p_{Ck}^A)' \), and set \( \pi_{1,k} = \alpha_k p_{Ck}^A (p_{Ck}^A)^{-1} \), where \( 0 < \alpha_k < 1 \). Because \( p_{Ck} < p_{Ck}^A \), \( 0 < \pi_{1,k} < 1 \). Then, define

\[
p_k^B = \frac{1}{1 - \pi_{1,k}} (p_k - \pi_{1,k} p_k^A)
\]

so that \( \pi_{1,k} p_k^A + (1 - \pi_{1,k}) p_k^B = p_k \). For \( l = 1, \ldots, \tilde{r}_k \), generate \( p_{k,l} \sim \text{Dirichlet}_C(p_k^A B_{k,l} + p_k^B (1 - B_{k,l}), \tau_k) \), where \( P(B_{k,l} = 1) = \pi_{1,k} = 1 - P(B_{k,l} = 0) \).

Then, generate \( (\tilde{n}_{1k,l}, \ldots, \tilde{n}_{Ck,l})' \sim \text{Multinomial}_C(p_{k,l}, n_c) \), and set \( \tilde{p}_{ik,l} = n_{c,k}^{-1} \tilde{n}_{ik,l} \). The direct estimator of the proportion is \( \tilde{p}_{ik} = \tilde{r}_k^{-1} \sum_{l=1}^{\tilde{r}_k} \tilde{p}_{ik,l} \). The direct estimators of the province two digit totals are generated using the method described above, and the direct estimator of the total in three digit code \( i \) and province \( k \) is \( \hat{M}_{ik} = \hat{M}_{k} \hat{p}_{ik} \). The direct estimator of the
covariance matrix of the vector of sampling errors of the proportions for province $k$ used in the simulation is

$$
\hat{\Sigma}_{ee,k,2st} = \tilde{r}_k^{-1}(\tilde{r}_k - 1)^{-1} \sum_{l=1}^{\tilde{r}_k} (\hat{p}_{k,l} - \tilde{p}_k)(\hat{p}_{k,l} - \tilde{p}_k)',
$$

(6.8)

where $\hat{p}_k = (\hat{p}_{1k}, \ldots, \hat{p}_{Ck})'$, and $\tilde{p}_{k,l} = (\tilde{p}_{1k,l}, \ldots, \tilde{p}_{Ck,l})'$. Thomas and Rao (1987) use the estimator (6.8). The direct estimator of the sampling covariance matrix of the vector of totals for province $k$ is based on a Taylor approximation and is

$$
\hat{\Sigma}_{aa,k,2st} = \hat{\Sigma}_{ee,k,2st} \hat{M}_k^2 + \hat{p}_k \hat{p}_k' \hat{M}_k^2 \tilde{n}_k,
$$

(6.9)

where $\hat{M}_k^2 \tilde{n}_k^{-1}$ is an estimator of the variance of the direct estimator of the total in province $k$.

**Remark: Properties of Sampling Errors Under the Two Stage Design**

The parameters defining the mixture distribution are chosen so that $E[\hat{p}_k | p_k] = p_k$. Conditional on the vector of true proportions, the covariance matrix of the vector of sampling errors in province $k$ is

$$
V\{e_k | p_k\} = \frac{n_{c,k} + \tau_k}{1 + \tau_k} \text{diag}(p_k) - p_k p_k' + \frac{(n_{c,k} - 1)\tau_k}{1 + \tau_k} \pi_{1,k}(p_k^A - p_k)(p_k^A - p_k)' + \frac{(n_{c,k} - 1)\tau_k}{1 + \tau_k} (1 - \pi_{1,k})(p_k^B - p_k)(p_k^B - p_k)'
$$

(6.10)

$$
:= V_{2st,k}
$$

(6.12)

(Thomas and Rao, 1987). Define the generalized design effects to be the eigenvalues of $[V_{srs}^{(1)}]^{-1} V_{2st}^{(1)}$, where the notation $A^{(1)}$ denotes the matrix obtained by omitting the first row and column of the matrix $A$. Given $p_k$, the two distinct eigenvalues are

$$
d_{1,k} = (n_{c,k} + \tau_k)(1 + \tau_k)^{-1}
$$

and

$$
d_{2,k} = [(n_{c,k} - 1)\tau_k\delta_k](1 + \tau_k)^{-1} + (n_{c,k} + \tau_k)(1 + \tau_k)^{-1},
$$

(6.13)

where $\delta_k = \pi_{1,k}(1 - \pi_{1,k})(p_k^A - p_k^B)' \tilde{\Gamma}_{uu,k}(p_k^A - p_k^B)$ (Rao and Scott, 1981), and $\tilde{\Gamma}_{uu,k} = \text{diag}(p_k) - p_k p_k'$. The design effect $d_{1,k}$ has multiplicity $C - 2$, and $d_{2,k}$ has multiplicity 1. The
sum of the two multiplicities is $C - 1$ because $\hat{\Gamma}_{uu,k}$ is a singular matrix with rank $C - 1$. The average of the design effects is

$$d_k = \frac{(n_{c,k} + \tau_k)(1 + \tau_k)^{-1} + [(n_{c,k} - 1)\tau_k\delta_k][(1 + \tau_k)(C - 1)]^{-1}}.$$

The parameter $\tau_k$ has an interpretation as a measure of similarity among units in the same cluster. As $\tau_k$ decreases, the variability among the cluster probability vectors increases, and the design effects decrease to one (all other parameters held fixed). If $n_{c,k} = 1$, then the model represents a simple random sample, and the design effects are both equal to 1. For a fixed value of $n_k$, the design effects increase as $n_{c,k}$ increases. The parameters $\alpha$ and $\delta_{1,k}, \ldots, \delta_{C,k}$ determine the difference between the two design effects.

### 6.1.1 Parameters for the Simulation

The parameters defining the expected values of the proportions and marginal totals are based on an initial analysis of the data for the two digit code A1 (specialist managers). The $\{p_{T,ik} : i = 1, \ldots, C; k = 1, \ldots, K\}$ used in the simulation are given in Table 6.2, along with the expected values of the province two digit totals and the coefficients of variation of the direct estimators of the province two digit totals. The expected sample sizes are the inverses of the squares of the coefficients of variation. Two values of $\psi$ are used in the simulation: $\psi = 0.02$ and $\psi = 0.003$. The parameters used to generate the sampling errors for the simulation representing a two stage sample are given in Table 6.3. The design effect $d_{1,k}$ does not depend on $p_k$, and the $\{d_{1,k} : k = 1, \ldots, K\}$ used in the simulation are in Table 6.3. Because $d_{2,k}$ is a nonlinear function of the true proportions, the mean of $d_{2,k}$ is difficult to derive analytically. The design effects $\{d_{2,k} : k = 1, \ldots, K\}$ and the averages of the design effects, $\{\bar{d}_k : k = 1, \ldots, K\}$ in the last two rows of Table 6.3 are obtained by evaluating the function defining $d_{2,k}$ in (6.13) at the fixed parameter $p_{T,k}$.

Under the working model for the sampling variance, the covariance matrix of the vector of sampling errors in province $k$ is

$$\Sigma_{ee,k,w} = c_k n_k^{-1} \Gamma_{uu,k},$$
where \( c_k = \text{trace}\{[\Gamma_{uu,k}]^{-1}n_k\Sigma_{ee,k}\} \). We obtain a MC approximation for the true \( c_k \) under the 2-stage simulation model by replacing the unknown \( \Sigma_{ee,k} \) in the definition of \( c_k \) with the MC variance of \( e_k \). The MC approximation for \( c_k \) is given in the last row of Table 6.4. The first four rows of Table 6.4 have the ratios of \( c_k p_{T,ik}(1 - p_{T,ik})n_k^{-1} \) to the MC variances of \( e_{ik} \) \((i = 1, \ldots, 4; k = 1, \ldots, 10)\). In the provinces other than British Columbia and Alberta, the working model variances are larger than the sampling variances in the first three categories, and the working model variances are smaller than the sampling variances in the last category. In British Columbia and Alberta, the model for the sampling errors represents a cluster sample design with a constant design effect of 1.4. Because the working model is true in British Columbia and Alberta, the ratio of \( c_k p_{T,ik}(1 - p_{T,ik})n_k^{-1} \) to the MC variance of \( e_{ik} \) is close to 1 in each cell.

The optimal weight, \( \gamma_{ik} = p_{T,ik}(1 - p_{T,ik})\psi(\sigma^2_{e,ik} + p_{T,ik}(1 - p_{T,ik})\psi)^{-1} \), that leads to the minimum MSE convex combination of the direct estimator and \( p_{T,ik} \) depends on the ratios of the variances of \( \{u_{ik} : i = 1, \ldots, C; k = 1, \ldots, K\} \) to the variances of \( \{e_{ik} : i = 1, \ldots, C; k = 1, \ldots, K\} \). Under the simple random sampling simulation model, the working model for the sampling errors holds, so the optimal weight, conditional on \( \tilde{n}_k \), is equal to \( \gamma_k \), where \( \gamma_k = \psi(\psi + c_k n_k^{-1})^{-1} \) for \( i = 1, \ldots, C \). Table 6.5 shows the weights \( \gamma_k \) under the simple random sampling model with \( c_k = (1 - \psi) \). For the simulation model representing a two stage design, the ratios of the variances of \( \{u_{ik} : i = 1, \ldots, C; k = 1, \ldots, K\} \) to the variances of \( \{e_{ik} : i = 1, \ldots, C; k = 1, \ldots, K\} \) are not constant in a province, and the optimal \( \gamma_{ik} \) differs from \( \gamma_k \). To compare the optimal predictors to predictors under the working model, we compute a Monte Carlo approximation for the optimal \( \gamma_{ik} \) with \( \psi = 0.02 \). The Monte Carlo approximation for the true \( \gamma_{ik} \) in Table 6.6 is the ratio of the Monte Carlo variance of \( u_{ik} \) to the Monte Carlo variance of \( \tilde{p}_{ik} \). The \( \gamma_k \) in the last row of Table 6.6 is \( \psi(\psi + \bar{d}_k(1 - \psi)n_k^{-1})^{-1} \), where \( \bar{d}_k \) is in the last row of Table 6.3. Because of the nonconstant design effects, the optimal weight in category 4 is smaller than the optimal weights in the other three categories in provinces other than British Columbia and Alberta. In British Columbia and Alberta, \( d_{1,k} = d_{2,k} \), and the differences among the Monte Carlo approximations for the optimal weights reflect simulation
variability.

<table>
<thead>
<tr>
<th></th>
<th>PE</th>
<th>NF</th>
<th>NB</th>
<th>NS</th>
<th>SK</th>
<th>MB</th>
<th>BC</th>
<th>AB</th>
<th>QC</th>
<th>ON</th>
</tr>
</thead>
<tbody>
<tr>
<td>(i = 1)</td>
<td>0.407</td>
<td>0.376</td>
<td>0.334</td>
<td>0.353</td>
<td>0.362</td>
<td>0.386</td>
<td>0.348</td>
<td>0.325</td>
<td>0.378</td>
<td>0.327</td>
</tr>
<tr>
<td>(i = 2)</td>
<td>0.171</td>
<td>0.137</td>
<td>0.157</td>
<td>0.094</td>
<td>0.148</td>
<td>0.117</td>
<td>0.165</td>
<td>0.142</td>
<td>0.185</td>
<td>0.187</td>
</tr>
<tr>
<td>(i = 3)</td>
<td>0.286</td>
<td>0.276</td>
<td>0.334</td>
<td>0.334</td>
<td>0.348</td>
<td>0.332</td>
<td>0.323</td>
<td>0.377</td>
<td>0.364</td>
<td>0.362</td>
</tr>
<tr>
<td>(i = 4)</td>
<td>0.136</td>
<td>0.211</td>
<td>0.175</td>
<td>0.219</td>
<td>0.141</td>
<td>0.165</td>
<td>0.164</td>
<td>0.156</td>
<td>0.073</td>
<td>0.124</td>
</tr>
<tr>
<td>(N_k)</td>
<td>995</td>
<td>3695</td>
<td>6140</td>
<td>7885</td>
<td>8625</td>
<td>10725</td>
<td>40260</td>
<td>47680</td>
<td>80235</td>
<td>187705</td>
</tr>
</tbody>
</table>

Table 6.1 Census proportions and 2-digit totals

<table>
<thead>
<tr>
<th></th>
<th>PE</th>
<th>NF</th>
<th>NB</th>
<th>NS</th>
<th>SK</th>
<th>MB</th>
<th>BC</th>
<th>AB</th>
<th>QC</th>
<th>ON</th>
</tr>
</thead>
<tbody>
<tr>
<td>(i = 1)</td>
<td>0.386</td>
<td>0.374</td>
<td>0.351</td>
<td>0.363</td>
<td>0.363</td>
<td>0.376</td>
<td>0.358</td>
<td>0.345</td>
<td>0.372</td>
<td>0.346</td>
</tr>
<tr>
<td>(i = 2)</td>
<td>0.155</td>
<td>0.140</td>
<td>0.149</td>
<td>0.117</td>
<td>0.144</td>
<td>0.129</td>
<td>0.152</td>
<td>0.142</td>
<td>0.161</td>
<td>0.162</td>
</tr>
<tr>
<td>(i = 3)</td>
<td>0.335</td>
<td>0.332</td>
<td>0.361</td>
<td>0.363</td>
<td>0.367</td>
<td>0.359</td>
<td>0.355</td>
<td>0.382</td>
<td>0.376</td>
<td>0.374</td>
</tr>
<tr>
<td>(i = 4)</td>
<td>0.124</td>
<td>0.154</td>
<td>0.140</td>
<td>0.157</td>
<td>0.126</td>
<td>0.135</td>
<td>0.135</td>
<td>0.132</td>
<td>0.092</td>
<td>0.118</td>
</tr>
<tr>
<td>(T_k)</td>
<td>770</td>
<td>2243</td>
<td>4538</td>
<td>6091</td>
<td>5622</td>
<td>9220</td>
<td>35052</td>
<td>31323</td>
<td>72926</td>
<td>142835</td>
</tr>
<tr>
<td>CV of (\hat{M}_k)</td>
<td>0.277</td>
<td>0.236</td>
<td>0.186</td>
<td>0.171</td>
<td>0.143</td>
<td>0.129</td>
<td>0.112</td>
<td>0.112</td>
<td>0.094</td>
<td>0.068</td>
</tr>
<tr>
<td>(n_k)</td>
<td>13</td>
<td>18</td>
<td>29</td>
<td>34</td>
<td>49</td>
<td>60</td>
<td>80</td>
<td>80</td>
<td>113</td>
<td>215</td>
</tr>
</tbody>
</table>

Table 6.2 Expected values of proportions and province two digit totals,
CV’s of direct estimators of province two digit totals, expected province sample sizes \((n_k)\). \(n_k\) is the inverse of the squared CV for province \(k\).
### Table 6.3 Parameters for 2-stage simulation model

<table>
<thead>
<tr>
<th>PE</th>
<th>NF</th>
<th>NB</th>
<th>NS</th>
<th>SK</th>
<th>MB</th>
<th>BC</th>
<th>AB</th>
<th>QC</th>
<th>ON</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n_{c,k}$</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>$\tau_k$</td>
<td>30</td>
<td>30</td>
<td>50</td>
<td>50</td>
<td>30</td>
<td>40</td>
<td>4</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>$\alpha_k$</td>
<td>0.950</td>
<td>0.950</td>
<td>0.950</td>
<td>0.950</td>
<td>0.950</td>
<td>1.000</td>
<td>1.000</td>
<td>0.900</td>
<td>0.900</td>
</tr>
<tr>
<td>$\delta_{1,k}$</td>
<td>1.650</td>
<td>1.650</td>
<td>1.150</td>
<td>1.150</td>
<td>1.150</td>
<td>1.650</td>
<td>1.000</td>
<td>1.000</td>
<td>1.250</td>
</tr>
<tr>
<td>$\delta_{2,k}$</td>
<td>1.950</td>
<td>1.950</td>
<td>1.550</td>
<td>1.550</td>
<td>1.550</td>
<td>1.950</td>
<td>1.000</td>
<td>1.000</td>
<td>1.400</td>
</tr>
<tr>
<td>$\delta_{3,k}$</td>
<td>1.750</td>
<td>1.750</td>
<td>1.250</td>
<td>1.250</td>
<td>1.250</td>
<td>1.750</td>
<td>1.000</td>
<td>1.000</td>
<td>1.350</td>
</tr>
<tr>
<td>$d_{1,k}$</td>
<td>1.032</td>
<td>1.032</td>
<td>1.020</td>
<td>1.020</td>
<td>1.032</td>
<td>1.024</td>
<td>1.400</td>
<td>1.400</td>
<td>1.500</td>
</tr>
<tr>
<td>$d_{2,k}$</td>
<td>1.416</td>
<td>1.413</td>
<td>1.201</td>
<td>1.194</td>
<td>1.210</td>
<td>1.409</td>
<td>1.400</td>
<td>1.400</td>
<td>1.811</td>
</tr>
<tr>
<td>$\bar{d}_k$</td>
<td>1.160</td>
<td>1.159</td>
<td>1.080</td>
<td>1.078</td>
<td>1.092</td>
<td>1.153</td>
<td>1.400</td>
<td>1.400</td>
<td>1.604</td>
</tr>
</tbody>
</table>

### Table 6.4 Ratio of $c_kp_{T,ik}(1 - p_{T,ik})n_k^{-1}$ to MC variance of $e_{ik}$. The $c_k$, in the last row, are MC approximations. 2-stage sampling error model, $\psi = 0.02$

<table>
<thead>
<tr>
<th>PE</th>
<th>NF</th>
<th>NB</th>
<th>NS</th>
<th>SK</th>
<th>MB</th>
<th>BC</th>
<th>AB</th>
<th>QC</th>
<th>ON</th>
</tr>
</thead>
<tbody>
<tr>
<td>$i = 1$</td>
<td>1.088</td>
<td>1.048</td>
<td>1.033</td>
<td>1.051</td>
<td>1.046</td>
<td>1.101</td>
<td>0.978</td>
<td>0.995</td>
<td>1.094</td>
</tr>
<tr>
<td>$i = 2$</td>
<td>1.087</td>
<td>1.122</td>
<td>1.070</td>
<td>1.027</td>
<td>1.058</td>
<td>1.120</td>
<td>1.018</td>
<td>0.997</td>
<td>1.071</td>
</tr>
<tr>
<td>$i = 3$</td>
<td>1.083</td>
<td>1.087</td>
<td>1.060</td>
<td>1.056</td>
<td>1.072</td>
<td>1.087</td>
<td>1.025</td>
<td>1.004</td>
<td>1.076</td>
</tr>
<tr>
<td>$i = 4$</td>
<td>0.839</td>
<td>0.828</td>
<td>0.884</td>
<td>0.906</td>
<td>0.881</td>
<td>0.811</td>
<td>0.982</td>
<td>1.003</td>
<td>0.856</td>
</tr>
<tr>
<td>$c_k$</td>
<td>1.275</td>
<td>1.253</td>
<td>1.115</td>
<td>1.108</td>
<td>1.107</td>
<td>1.156</td>
<td>1.417</td>
<td>1.434</td>
<td>1.630</td>
</tr>
<tr>
<td>$\bar{d}_k$</td>
<td>1.607</td>
<td>1.607</td>
<td>1.607</td>
<td>1.607</td>
<td>1.607</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 6.3 Parameters for 2-stage simulation model

Table 6.4 Ratio of $c_kp_{T,ik}(1 - p_{T,ik})n_k^{-1}$ to MC variance of $e_{ik}$. The $c_k$, in the last row, are MC approximations. 2-stage sampling error model, $\psi = 0.02$
\[ \psi = 0.02 \quad 0.210 \quad 0.269 \quad 0.372 \quad 0.410 \quad 0.500 \quad 0.550 \quad 0.620 \quad 0.620 \quad 0.698 \quad 0.814 \]

\[ \psi = 0.003 \quad 0.038 \quad 0.051 \quad 0.080 \quad 0.093 \quad 0.128 \quad 0.153 \quad 0.194 \quad 0.194 \quad 0.254 \quad 0.393 \]

Table 6.5 Optimal weights \( \gamma_k \) under SRS model

<table>
<thead>
<tr>
<th></th>
<th>PE</th>
<th>NF</th>
<th>NB</th>
<th>NS</th>
<th>SK</th>
<th>MB</th>
<th>BC</th>
<th>AB</th>
<th>QC</th>
<th>ON</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \gamma_{1k} )</td>
<td>0.176</td>
<td>0.234</td>
<td>0.344</td>
<td>0.394</td>
<td>0.478</td>
<td>0.532</td>
<td>0.521</td>
<td>0.532</td>
<td>0.603</td>
<td>0.739</td>
</tr>
<tr>
<td>( \gamma_{2k} )</td>
<td>0.185</td>
<td>0.243</td>
<td>0.357</td>
<td>0.385</td>
<td>0.484</td>
<td>0.539</td>
<td>0.542</td>
<td>0.521</td>
<td>0.596</td>
<td>0.736</td>
</tr>
<tr>
<td>( \gamma_{3k} )</td>
<td>0.182</td>
<td>0.241</td>
<td>0.349</td>
<td>0.395</td>
<td>0.484</td>
<td>0.520</td>
<td>0.530</td>
<td>0.524</td>
<td>0.598</td>
<td>0.738</td>
</tr>
<tr>
<td>( \gamma_{4k} )</td>
<td>0.144</td>
<td>0.197</td>
<td>0.318</td>
<td>0.360</td>
<td>0.439</td>
<td>0.466</td>
<td>0.529</td>
<td>0.540</td>
<td>0.539</td>
<td>0.706</td>
</tr>
</tbody>
</table>

\( \gamma_k = E_{MC}[u_{ik}^2](V_{MC}\{\hat{p}_{ik}\})^{-1} \); \( \gamma_k = \psi(\psi + c_k n_k)^{-1} \)

Table 6.6 MC approximations for optimal \( \gamma_{ik} \) and working \( \gamma_k \) under 2-stage model, \( \psi = 0.02 \)

6.2 Details of Computing and Estimation

The two sampling error models (SRS and 2-stage) and the two values of \( \psi \) (0.02 and 0.003) lead to four simulation models. We generate 5000 data sets from each of the four models. The empirical properties of the predictors are based on the 5000 samples. Because the bootstrap MSE estimator requires more time to compute, the empirical properties the MSE estimators are based on a subset of 2000 data sets.

The model described above is used to generate data for one two-way table representing a single two digit code. In the LFS, two-way tables from four independent years of data are combined to improve the precision on the estimator of \( \psi \). In the simulation, four iid sets of true values and direct estimators are generated from the models above. The estimators of \( c_k \)
and $\psi$ are constructed with the four two-way tables as shown in (4.5) and (4.22) of Section 4.1.1 and Section 4.1.3, respectively. (In the notation of Section 4, $|A_\psi| = |A_c| = 4$.)

The iteratively reweighted least squares procedure is defined to converge when the maximum change in the estimates of $\psi$ and $\lambda$ is smaller than $10^{-5}$. In the simulation, if the change in the estimates exceeds $10^{-5}$ after 10 iterations, the estimates obtained in the 10th step are used. In the simulation to represent simple random sampling, with $\psi = 0.02$, the iteration converged before the 10th step in all 5000 samples. In the simple random sampling simulation with $\psi = 0.003$, 24 of the 5000 samples failed to converge before the 10th step, but the maximum change in the estimators at the last step is $9 \times 10^{-5}$. In the simulation for a two stage design with $\psi = 0.02$, one of the 5000 samples leads to an estimator that does not converge before the last step, and the final change in the estimators for this sample is $1.1 \times 10^{-5}$. In the two stage simulation when $\psi = 0.003$, seven of the samples lead to estimators that do not converge before 10 steps, and the maximum change in the estimators is $0.0004$.

The bootstrap sample size in the simulation is $B = 200$. The iteratively reweighted estimation procedure is run for a maximum of 3 steps in the bootstrap to reduce the computing time required. For the simple random sampling simulation, the procedure used to generate the bootstrap version of the direct estimator of the sampling covariance matrix is changed as follows. In the SRS simulation, the direct estimators of the sampling variances are the functions of the direct estimators of the totals and proportions defined in (6.3) and (6.4), respectively. The bootstrap versions of the direct estimators of the sampling variances are defined to be the corresponding functions of the bootstrap direct estimators of the totals and proportions. The bootstrap versions of the direct estimators of the variances of the direct estimators of the province two digit totals are also defined differently for the simulation. For both simulation models, the direct estimator of the variance of the direct estimator of two digit total in province $k$ is $\hat{\sigma}_k^2 \hat{n}_k^{-1}$. The bootstrap version of the direct estimator of the variance is defined to be $(\hat{M}_k^*)^2 \hat{n}_k^{-1}$. (In step 3 of Section 5 and in the LFS application, the bootstrap variance of the province two digit total is $(\hat{M}_k^*)^2 \hat{n}_k^{-1} c_k$.)

The bootstrap data generating procedure requires computing the quantile function of a
gamma distribution. When the generated $M_{ik}^*$ is sufficiently small, the R function “qgamma” encounters numerical problems and returns a missing value. To avoid generating missing values in the bootstrap, $\hat{M}_{ik}^*$ is set equal to $M_{ik}^*$ when the R “qgamma” function returns a missing value. In the simulations with $\psi = 0.003$, the modification for small $\alpha_{ik}^*$ is not needed in any of the MC trials. In the SRS simulation with $\psi = 0.02$, the modification is used in at least one bootstrap sample in 91 of the 2000 MC samples. Table 6.7 gives more detailed information on the number of bootstrap samples in which the modification is used for the SRS simulation with $\psi = 0.02$. For example, the modification is used in exactly one bootstrap sample in 58 of the MC samples, and the modification is used in two bootstrap samples in 16 of the MC samples. In the two stage simulation with $\psi = 0.01$, $\alpha_{ik}^*$ is too small to use in one bootstrap sample for one generated data set. Based on informal investigations, we conjecture that adjustment for small $\alpha_{ik}^*$ is needed most often in three digit code $i = 4$ of Quebec, where $p_{T,ik} = 0.09$.

<table>
<thead>
<tr>
<th>Number of Bootstrap Samples</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of MC samples</td>
<td>58</td>
<td>16</td>
<td>8</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 6.7  Number of MC samples in which the modification for small $\alpha_{ik}^*$ is used in each of 1-7 bootstrap samples

In the definition of the bootstrap MSE estimator in (5.36), a multiplicative bias correction is used if the additive bias correction produces a negative estimate. The multiplicative bias correction is only needed in the SRS simulation with $\psi = 0.02$. Table 6.8 summarizes the number of MC samples in which the multiplicative bias correction is used for each cell of the two-way table for the SRS simulation model with $\psi = 0.02$. The additive bias correction leads to positive MSE estimates for the other simulation models.
<table>
<thead>
<tr>
<th>i = 1</th>
<th>PE</th>
<th>NF</th>
<th>NS</th>
<th>NB</th>
<th>SK</th>
<th>MB</th>
<th>BC</th>
<th>AB</th>
<th>QC</th>
<th>ON</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>i = 2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>i = 3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>i = 4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>10</td>
<td>4</td>
<td>20</td>
<td>22</td>
</tr>
</tbody>
</table>

Table 6.8 Number of MC samples (out of 2000) in which the multiplicative bias correction is used. SRS simulation, $\psi = 0.02$.

### 6.3 Efficiencies of Predictors

The primary objective of the small area procedure is to obtain predictors of the proportions and totals with smaller MSE’s than the direct estimators. To compare the MSE’s of the predictors to the MSE’s of the direct estimators, we compute the ratios of the Monte Carlo (MC) MSE’s of the predictors to the MC MSE’s of the direct estimators.

Tables 6.9 and 6.10 show the averages of the MSE ratios in each province for the SRS and two stage and SRS sampling error models, respectively. The first column of each table indicates the parameter (proportion or total) and the value of $\psi$ (0.003 or 0.02). The provinces are listed in decreasing order with respect to the coefficients of variation of the direct estimators of the province two digit totals. Monte Carlo standard errors are in parentheses. Several patterns are evident in Tables 6.9 and 6.10.

- For all models, the averages of the MSE ratios are less than 1. The MSE ratios for the individual cells (not shown) show that the MSE’s of the predictors ($\tilde{p}_{ik,B}$ and $\tilde{M}_{ik,B}$) are uniformly smaller than the MSE’s of the direct estimators.

- The MSE ratios increase as the coefficients of variation of the province two digit totals decrease. In the simulations, the coefficient of variation for province $k$ is $\sqrt{n_k^{-1}}$, where $n_k$ represents an expected sample size for province $k$. The weight assigned to the direct
estimator in the initial predictor is

\[ \hat{\gamma}_{k,B} = \frac{\hat{\psi}(\hat{\psi} + \hat{c}_k \hat{n}_k^{-1}) + \hat{V}(\hat{\psi})}{(\hat{\psi} + \hat{c}_k \hat{n}_k^{-1})^2 + \hat{V}(\hat{\psi}) + \hat{n}_k^{-2}\hat{V}(\hat{c}_k)} \]

where \( E[\hat{n}_k] = n_k \). Consequently, as \( n_k \) increases, the predictor approaches the direct estimator, and the reduction in the MSE due to the small area procedure decreases.

- The MSE ratios at \( \psi = 0.003 \) are smaller than the corresponding MSE ratios at \( \psi = 0.02 \). The leading term in the MSE of the initial predictor of a proportion is \( \gamma_k^2 \sigma^2_{\epsilon,ik} + (1 - \gamma_k)^2 \psi p_{T,ik}(1 - p_{T,ik}) \). As \( \psi \) decreases, the leading term approaches zero. The MC MSE’s of the predictors of the proportions decrease at a faster rate as \( \psi \) decreases in provinces with large coefficients of variation than in provinces with small coefficients of variation. For example, for the simple random sampling simulation model, the average of the MSE ratios for the predictors of the proportions for Prince Edward Island when \( \psi = 0.003 \) is 21% of the corresponding average of MSE ratios when \( \psi = 0.02 \). In contrast, the average of the MSE ratios for the predictors of the proportions for Ontario when \( \psi = 0.003 \) for 72% of the corresponding average when \( \psi = 0.02 \).

- The averages of the MSE ratios for totals are larger than the corresponding average MSE ratios for the proportions. The reduction in the MSE due to the prediction procedure is smaller for totals than for proportions because of variability in the direct estimator of the province two digit total. If the province two digit totals were fixed, then the MSE ratios would be the same for both proportions and totals. Also, the reduction in the MSE that occurs when \( \psi \) decreases from 0.02 to 0.003 is smaller for totals than for proportions, which suggests that variability in the direct estimators of the province two digit totals limits the possible reduction in the MSE for totals.
### Table 6.9 Averages of ratios of MC MSE's of predictors to MC MSE’s of direct estimators for SRS sampling error model

<table>
<thead>
<tr>
<th></th>
<th>PE</th>
<th>NF</th>
<th>NB</th>
<th>NS</th>
<th>SK</th>
<th>MB</th>
<th>BC</th>
<th>AB</th>
<th>QC</th>
<th>ON</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_{ik}, 0.003$</td>
<td>0.0679</td>
<td>0.1051</td>
<td>0.1361</td>
<td>0.2020</td>
<td>0.2149</td>
<td>0.2733</td>
<td>0.3309</td>
<td>0.3369</td>
<td>0.4946</td>
<td>0.6339</td>
</tr>
<tr>
<td></td>
<td>(0.0012)</td>
<td>(0.0018)</td>
<td>(0.0022)</td>
<td>(0.0029)</td>
<td>(0.0033)</td>
<td>(0.0040)</td>
<td>(0.0043)</td>
<td>(0.0043)</td>
<td>(0.0051)</td>
<td>(0.0049)</td>
</tr>
<tr>
<td>$p_{ik}, 0.02$</td>
<td>0.2419</td>
<td>0.3272</td>
<td>0.4195</td>
<td>0.4990</td>
<td>0.549</td>
<td>0.592</td>
<td>0.673</td>
<td>0.689</td>
<td>0.7926</td>
<td>0.8846</td>
</tr>
<tr>
<td></td>
<td>(0.0059)</td>
<td>(0.0075)</td>
<td>(0.0089)</td>
<td>(0.0096)</td>
<td>(0.010)</td>
<td>(0.010)</td>
<td>(0.010)</td>
<td>(0.010)</td>
<td>(0.0090)</td>
<td>(0.0072)</td>
</tr>
<tr>
<td>$M_{ik}, 0.003$</td>
<td>0.3077</td>
<td>0.3290</td>
<td>0.3578</td>
<td>0.4090</td>
<td>0.4136</td>
<td>0.4544</td>
<td>0.5095</td>
<td>0.5013</td>
<td>0.6307</td>
<td>0.7278</td>
</tr>
<tr>
<td></td>
<td>(0.0043)</td>
<td>(0.0043)</td>
<td>(0.0044)</td>
<td>(0.0046)</td>
<td>(0.0045)</td>
<td>(0.0049)</td>
<td>(0.0048)</td>
<td>(0.0048)</td>
<td>(0.0048)</td>
<td>(0.0041)</td>
</tr>
<tr>
<td>$M_{ik}, 0.02$</td>
<td>0.4371</td>
<td>0.5047</td>
<td>0.5760</td>
<td>0.6477</td>
<td>0.6736</td>
<td>0.6926</td>
<td>0.7634</td>
<td>0.7679</td>
<td>0.8540</td>
<td>0.9124</td>
</tr>
<tr>
<td></td>
<td>(0.0073)</td>
<td>(0.0081)</td>
<td>(0.0085)</td>
<td>(0.0088)</td>
<td>(0.0091)</td>
<td>(0.0087)</td>
<td>(0.0083)</td>
<td>(0.0082)</td>
<td>(0.0071)</td>
<td>(0.0055)</td>
</tr>
</tbody>
</table>

### Table 6.10 Averages of ratios of MC MSE’s of predictors to MC MSE’s of direct estimators for 2-stage sampling error model

<table>
<thead>
<tr>
<th></th>
<th>PE</th>
<th>NF</th>
<th>NB</th>
<th>NS</th>
<th>SK</th>
<th>MB</th>
<th>BC</th>
<th>AB</th>
<th>QC</th>
<th>ON</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_{ik}, 0.003$</td>
<td>0.0678</td>
<td>0.1038</td>
<td>0.1468</td>
<td>0.2204</td>
<td>0.2256</td>
<td>0.2930</td>
<td>0.2822</td>
<td>0.2968</td>
<td>0.4260</td>
<td>0.5783</td>
</tr>
<tr>
<td></td>
<td>(0.0012)</td>
<td>(0.0017)</td>
<td>(0.0023)</td>
<td>(0.0034)</td>
<td>(0.0035)</td>
<td>(0.0044)</td>
<td>(0.0038)</td>
<td>(0.0040)</td>
<td>(0.0046)</td>
<td>(0.0048)</td>
</tr>
<tr>
<td>$p_{ik}, 0.02$</td>
<td>0.2242</td>
<td>0.3031</td>
<td>0.4138</td>
<td>0.4969</td>
<td>0.5588</td>
<td>0.6277</td>
<td>0.6186</td>
<td>0.6281</td>
<td>0.7097</td>
<td>0.8317</td>
</tr>
<tr>
<td></td>
<td>(0.0037)</td>
<td>(0.0046)</td>
<td>(0.0059)</td>
<td>(0.0065)</td>
<td>(0.0071)</td>
<td>(0.0077)</td>
<td>(0.0069)</td>
<td>(0.0067)</td>
<td>(0.0061)</td>
<td>(0.0051)</td>
</tr>
<tr>
<td>$M_{ik}, 0.003$</td>
<td>0.2872</td>
<td>0.3100</td>
<td>0.3523</td>
<td>0.4184</td>
<td>0.4159</td>
<td>0.4618</td>
<td>0.4250</td>
<td>0.4337</td>
<td>0.5407</td>
<td>0.6578</td>
</tr>
<tr>
<td></td>
<td>(0.0040)</td>
<td>(0.0041)</td>
<td>(0.0043)</td>
<td>(0.0048)</td>
<td>(0.0047)</td>
<td>(0.0050)</td>
<td>(0.0046)</td>
<td>(0.0046)</td>
<td>(0.0048)</td>
<td>(0.0043)</td>
</tr>
<tr>
<td>$M_{ik}, 0.02$</td>
<td>0.4144</td>
<td>0.4784</td>
<td>0.5685</td>
<td>0.6250</td>
<td>0.6702</td>
<td>0.7166</td>
<td>0.7000</td>
<td>0.7050</td>
<td>0.7700</td>
<td>0.8628</td>
</tr>
<tr>
<td></td>
<td>(0.0047)</td>
<td>(0.0050)</td>
<td>(0.0057)</td>
<td>(0.0058)</td>
<td>(0.0061)</td>
<td>(0.0063)</td>
<td>(0.0060)</td>
<td>(0.0058)</td>
<td>(0.0052)</td>
<td>(0.0043)</td>
</tr>
</tbody>
</table>

To analyze differences across the categories in each province, the MSE ratios for the separate categories based on the SRS simulation model with $\psi = 0.003$ and $\psi = 0.02$ are
plotted in Figures 6.1 and 6.2, respectively. Each province has four categories, leading to four points in each province. The triangles in the plot are MSE ratios for totals, and the circles are MSE ratios for proportions. The corresponding plots based on the two stage simulation model are similar and therefore omitted. For both values of \( \psi \), the MSE ratios for proportions are roughly constant within each province. The MSE ratios for totals vary more within a particular province than the MSE ratios for the proportions. In particular, for each province, if \( p_{T,ik} < p_{T,jk} \), then the MSE ratio for the total in cell \((i,k)\) is smaller than the MSE ratio for the total in cell \((j,k)\). This pattern is more pronounced when \( \psi = 0.003 \) (Figure 6.1) than when \( \psi = 0.02 \) (Figure 6.2). The pattern arises because the variance of the synthetic estimator of the total is highly correlated with the synthetic estimator of the total. Table 6.11 has ratios of MC MSE’s of predictors calculated with the true parameters to MC sampling variances for the totals under the SRS model with \( \psi = 0.003 \). The ratios of the leading terms in the MSE’s for the totals to the sampling variances in Table 6.11 do not exhibit the pattern in Figure 6.1. In each province, the MSE ratios based on the leading terms (Table 6.11) are smaller in categories with larger expected values. The differences among the MSE ratios for categories in the same province based on the leading terms alone (Table 6.11) are smaller than the differences in Figure 6.1.
Figure 6.1 Ratios of MC MSE’s of predictors of proportions and totals to MC sampling variances. Proportion $\hat{p}_{ik,B}$: ⋆, Total $\tilde{M}_{ik,B}$: △. SRS sampling error model, $\psi = 0.003$. 
Figure 6.2  Ratios of MC MSE’s of predictors of proportions and totals to MC sampling variances. Proportion $\bar{p}_{ik,B}$: •, Total $\bar{M}_{ik,B}$: △. SRS sampling error model, $\psi = 0.02$. 
Table 6.11 Ratios of MC MSE’s of predictors of totals calculated with true parameters to sampling variances of totals, SRS sampling error model, $\psi = 0.003$

<table>
<thead>
<tr>
<th></th>
<th>PE</th>
<th>NF</th>
<th>NB</th>
<th>NS</th>
<th>SK</th>
<th>MB</th>
<th>BC</th>
<th>AB</th>
<th>QC</th>
<th>ON</th>
</tr>
</thead>
<tbody>
<tr>
<td>$i = 1$</td>
<td>0.02406</td>
<td>0.03261</td>
<td>0.0524</td>
<td>0.0570</td>
<td>0.0903</td>
<td>0.1074</td>
<td>0.1421</td>
<td>0.1384</td>
<td>0.1852</td>
<td>0.2999</td>
</tr>
<tr>
<td></td>
<td>(0.00068)</td>
<td>(0.00093)</td>
<td>(0.0014)</td>
<td>(0.0015)</td>
<td>(0.0023)</td>
<td>(0.0027)</td>
<td>(0.0033)</td>
<td>(0.0033)</td>
<td>(0.0043)</td>
<td>(0.0058)</td>
</tr>
<tr>
<td>$i = 2$</td>
<td>0.03051</td>
<td>0.0441</td>
<td>0.0675</td>
<td>0.0873</td>
<td>0.1130</td>
<td>0.1321</td>
<td>0.1674</td>
<td>0.1686</td>
<td>0.2217</td>
<td>0.3520</td>
</tr>
<tr>
<td></td>
<td>(0.00090)</td>
<td>(0.0013)</td>
<td>(0.0018)</td>
<td>(0.0025)</td>
<td>(0.0030)</td>
<td>(0.0035)</td>
<td>(0.0043)</td>
<td>(0.0043)</td>
<td>(0.0054)</td>
<td>(0.0078)</td>
</tr>
<tr>
<td>$i = 3$</td>
<td>0.02450</td>
<td>0.03504</td>
<td>0.0524</td>
<td>0.0623</td>
<td>0.0896</td>
<td>0.1048</td>
<td>0.1419</td>
<td>0.1343</td>
<td>0.1947</td>
<td>0.3069</td>
</tr>
<tr>
<td></td>
<td>(0.00069)</td>
<td>(0.00095)</td>
<td>(0.0014)</td>
<td>(0.0017)</td>
<td>(0.0023)</td>
<td>(0.0027)</td>
<td>(0.0034)</td>
<td>(0.0033)</td>
<td>(0.0044)</td>
<td>(0.0063)</td>
</tr>
<tr>
<td>$i = 4$</td>
<td>0.0351</td>
<td>0.0413</td>
<td>0.0700</td>
<td>0.0814</td>
<td>0.1154</td>
<td>0.1410</td>
<td>0.1674</td>
<td>0.1752</td>
<td>0.2278</td>
<td>0.3620</td>
</tr>
<tr>
<td></td>
<td>(0.0011)</td>
<td>(0.0012)</td>
<td>(0.0019)</td>
<td>(0.0022)</td>
<td>(0.0033)</td>
<td>(0.0037)</td>
<td>(0.0042)</td>
<td>(0.0044)</td>
<td>(0.0057)</td>
<td>(0.0078)</td>
</tr>
</tbody>
</table>

6.3.1 Effect of Beale Ratio Estimator of $\gamma_{ik}$ on MSE of Predictor

The Beale ratio estimator of $\gamma_k$ is used to construct the predictors. The difference between the MSE of a predictor calculated with the standard estimator of $\gamma_k$ and the MSE of a predictor calculated with the Beale estimator is of smaller order than $K^{-1}$. For convenience, we refer to the predictors based on the Beale estimator of $\gamma_k$ as the “Beale predictors” and the predictors based on the standard estimators of $\gamma_k$ as the “standard predictors.” We examine the differences between the MSE’s of the Beale predictors and the standard predictors.

Tables 6.12 and 6.13 show the averages of the ratios of the MC MSE’s of the benchmarked predictors calculated with the Beale estimator, $\hat{\gamma}_{ik,B}$, to the MC MSE’s of the benchmarked predictors calculated with the standard estimator $\hat{\gamma}_k$. The SRS sampling error model is used for Table 6.12, and the two stage sampling error model is used for Table 6.13. The Beale predictors of the proportions in the small provinces have consistently smaller MSE’s than the standard predictors; however, the reduction in the MSE due to use of the Beale ratio estimator is never larger than 3% and is usually less than 1%. The smallest MSE ratios arise in the
simulation from the 2-stage sampling error model with $\psi = 0.003$. In the larger provinces, the standard predictors of the proportions have smaller MSE’s than the Beale predictors. In the smaller provinces, the differences between the MSE’s of the Beale predictors and the MSE’s of the standard predictors of the totals are negligible. The standard predictors of the totals are more efficient than the Beale predictors in the larger provinces. The differences between proportions and totals are more pronounced when $\psi = 0.003$ than when $\psi = 0.02$.

<table>
<thead>
<tr>
<th></th>
<th>PE</th>
<th>NF</th>
<th>NB</th>
<th>NS</th>
<th>SK</th>
<th>MB</th>
<th>BC</th>
<th>AB</th>
<th>QC</th>
<th>ON</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_{ik,0.003}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.9931</td>
<td>0.9938</td>
<td>0.9924</td>
<td>0.9934</td>
<td>0.9898</td>
<td>0.9895</td>
<td>0.9936</td>
<td>0.9933</td>
<td>0.9986</td>
<td>1.0127</td>
</tr>
<tr>
<td></td>
<td>(0.0012)</td>
<td>(0.0018)</td>
<td>(0.0022)</td>
<td>(0.0029)</td>
<td>(0.0033)</td>
<td>(0.0040)</td>
<td>(0.0043)</td>
<td>(0.0043)</td>
<td>(0.0051)</td>
<td>(0.0049)</td>
</tr>
<tr>
<td>$p_{ik,0.02}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.9994</td>
<td>0.9996</td>
<td>0.9992</td>
<td>0.9988</td>
<td>0.9981</td>
<td>0.9980</td>
<td>0.9984</td>
<td>0.9989</td>
<td>1.0008</td>
<td>1.0011</td>
</tr>
<tr>
<td></td>
<td>(0.0037)</td>
<td>(0.0047)</td>
<td>(0.0055)</td>
<td>(0.0061)</td>
<td>(0.0066)</td>
<td>(0.0066)</td>
<td>(0.0065)</td>
<td>(0.0065)</td>
<td>(0.0055)</td>
<td>(0.0045)</td>
</tr>
<tr>
<td>$M_{ik,0.003}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.9985</td>
<td>0.9982</td>
<td>0.9976</td>
<td>0.9978</td>
<td>0.9960</td>
<td>0.9953</td>
<td>0.9973</td>
<td>0.9970</td>
<td>0.9993</td>
<td>1.0086</td>
</tr>
<tr>
<td></td>
<td>(0.0043)</td>
<td>(0.0043)</td>
<td>(0.0044)</td>
<td>(0.0046)</td>
<td>(0.0045)</td>
<td>(0.0049)</td>
<td>(0.0048)</td>
<td>(0.0048)</td>
<td>(0.0048)</td>
<td>(0.0041)</td>
</tr>
<tr>
<td>$M_{ik,0.02}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.0000</td>
<td>1.0001</td>
<td>1.0001</td>
<td>0.9997</td>
<td>0.9992</td>
<td>0.9992</td>
<td>0.9991</td>
<td>0.9996</td>
<td>1.0007</td>
<td>1.0008</td>
</tr>
<tr>
<td></td>
<td>(0.0047)</td>
<td>(0.0051)</td>
<td>(0.0053)</td>
<td>(0.0056)</td>
<td>(0.0058)</td>
<td>(0.0057)</td>
<td>(0.0054)</td>
<td>(0.0052)</td>
<td>(0.0044)</td>
<td>(0.0035)</td>
</tr>
</tbody>
</table>

Table 6.12 Averages of ratios of MC MSE’s of Beale predictors to MC MSE’s of standard predictors, SRS sampling error model.
Table 6.13  Averages of ratios of MC MSE’s of Beale predictors to MC MSE’s of standard predictors, 2-stage sampling error model.

<table>
<thead>
<tr>
<th></th>
<th>PE</th>
<th>NF</th>
<th>NB</th>
<th>NS</th>
<th>SK</th>
<th>MB</th>
<th>BC</th>
<th>AB</th>
<th>QC</th>
<th>ON</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_{ik,0.003}$</td>
<td>0.988</td>
<td>0.988</td>
<td>0.986</td>
<td>0.987</td>
<td>0.980</td>
<td>0.981</td>
<td>0.989</td>
<td>0.990</td>
<td>1.000</td>
<td>1.020</td>
</tr>
<tr>
<td>$p_{ik,0.02}$</td>
<td>0.999</td>
<td>0.999</td>
<td>0.997</td>
<td>0.997</td>
<td>0.994</td>
<td>0.994</td>
<td>0.997</td>
<td>0.997</td>
<td>0.999</td>
<td>1.001</td>
</tr>
<tr>
<td>$M_{ik,0.003}$</td>
<td>0.997</td>
<td>0.996</td>
<td>0.995</td>
<td>0.995</td>
<td>0.991</td>
<td>0.991</td>
<td>0.994</td>
<td>0.994</td>
<td>1.000</td>
<td>1.015</td>
</tr>
<tr>
<td>$M_{ik,0.02}$</td>
<td>1.000</td>
<td>1.000</td>
<td>0.999</td>
<td>0.999</td>
<td>0.997</td>
<td>0.996</td>
<td>0.998</td>
<td>0.998</td>
<td>1.000</td>
<td>1.001</td>
</tr>
</tbody>
</table>

6.3.2 Effect of Raking on MSE of Predictor

Tables 6.14 and 6.15 contain the averages of the ratios of the MC MSE’s of the benchmarked predictors to the MC MSE’s of the initial predictors in each province for the SRS and two-stage sampling error models, respectively. The MSE’s of the benchmarked predictors are larger than the MSE’s of the corresponding initial predictors. The relative differences between the MSE’s of the benchmarked predictors and the MSE’s of the initial predictors increase as the coefficients of variation of the direct estimators of the province two digit totals decrease.

In the simulation to represent a simple random sample, the MSE of the initial predictor is an estimator of the optimal predictor, so we expect the MSE of the raked predictor to exceed the MSE of the initial predictor. Because the working model for the sampling variances does not hold for the model that represents a two stage design, the initial predictors are not estimates of the optimal predictors for two stage designs. Regardless, the effect of raking on the MSE’s of the predictors is greater under the model that represents a two stage design than under the model that represents simple random sampling. A possible reason for why the effect of raking is greater under the two stage simulation model than under the SRS simulation model is that the covariance matrix underlying raking is a multinomial covariance matrix. Because the covariance matrix of the sampling errors in the SRS simulation is a multinomial covariance matrix.
matrix, we expect the raking operation to be a more efficient benchmarking procedure for the SRS simulation model than for the two stage simulation model.

The effect of raking is also greater when $\psi = 0.003$ than when $\psi = 0.02$. An explanation for why the effect of raking increases as $\psi$ decreases is as follows. The initial predictor has the form $\hat{\gamma}_{k,B} \hat{p}_{ik} + (1 - \hat{\gamma}_{k,B}) \hat{p}_{T,ik}$. As $\hat{\psi}$ increases, $\hat{\gamma}_{k,B}$ approaches 1, and the predictors approach the direct estimators, which satisfy the benchmarking restriction. In contrast, as $\hat{\psi}$ decreases, $\hat{\gamma}_{k,B}$ approaches zero, and the predictors approach the synthetic estimators, which do not satisfy the benchmarking restriction.

<table>
<thead>
<tr>
<th></th>
<th>PE</th>
<th>NF</th>
<th>NB</th>
<th>NS</th>
<th>SK</th>
<th>MB</th>
<th>BC</th>
<th>AB</th>
<th>QC</th>
<th>ON</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_{ik}$, 0.003</td>
<td>1.0260</td>
<td>1.0247</td>
<td>1.0323</td>
<td>1.0306</td>
<td>1.0322</td>
<td>1.0323</td>
<td>1.0420</td>
<td>1.0474</td>
<td>1.0557</td>
<td>1.0882</td>
</tr>
<tr>
<td></td>
<td>(0.0031)</td>
<td>(0.0030)</td>
<td>(0.0035)</td>
<td>(0.0032)</td>
<td>(0.0037)</td>
<td>(0.0036)</td>
<td>(0.0037)</td>
<td>(0.0037)</td>
<td>(0.0037)</td>
<td>(0.0046)</td>
</tr>
<tr>
<td>$p_{ik}$, 0.02</td>
<td>1.0055</td>
<td>1.0069</td>
<td>1.0080</td>
<td>1.0084</td>
<td>1.0096</td>
<td>1.0077</td>
<td>1.0141</td>
<td>1.0115</td>
<td>1.0131</td>
<td>1.0324</td>
</tr>
<tr>
<td></td>
<td>(0.0013)</td>
<td>(0.0013)</td>
<td>(0.0015)</td>
<td>(0.0015)</td>
<td>(0.0017)</td>
<td>(0.0018)</td>
<td>(0.0020)</td>
<td>(0.0019)</td>
<td>(0.0022)</td>
<td>(0.0028)</td>
</tr>
<tr>
<td>$M_{ik}$, 0.003</td>
<td>1.00587</td>
<td>1.00740</td>
<td>1.0101</td>
<td>1.0112</td>
<td>1.0119</td>
<td>1.0150</td>
<td>1.0212</td>
<td>1.0241</td>
<td>1.0308</td>
<td>1.0552</td>
</tr>
<tr>
<td></td>
<td>(0.00092)</td>
<td>(0.00098)</td>
<td>(0.0013)</td>
<td>(0.0012)</td>
<td>(0.0016)</td>
<td>(0.0017)</td>
<td>(0.0019)</td>
<td>(0.0019)</td>
<td>(0.0020)</td>
<td>(0.0029)</td>
</tr>
<tr>
<td>$M_{ik}$, 0.02</td>
<td>1.00121</td>
<td>1.00342</td>
<td>1.00332</td>
<td>1.00488</td>
<td>1.0049</td>
<td>1.0044</td>
<td>1.0091</td>
<td>1.0084</td>
<td>1.0093</td>
<td>1.0235</td>
</tr>
<tr>
<td></td>
<td>(0.00065)</td>
<td>(0.00075)</td>
<td>(0.00088)</td>
<td>(0.00091)</td>
<td>(0.0011)</td>
<td>(0.0011)</td>
<td>(0.0013)</td>
<td>(0.0013)</td>
<td>(0.0015)</td>
<td>(0.0020)</td>
</tr>
</tbody>
</table>

Table 6.14 Averages of ratios of MC MSE’s of raked predictors to MC MSE’s of initial predictors, SRS sampling error model
Table 6.15 Averages of ratios of MC MSE's of raked predictors to MC MSE's of initial predictors, 2-stage sampling error model

<table>
<thead>
<tr>
<th></th>
<th>PE</th>
<th>NF</th>
<th>NB</th>
<th>NS</th>
<th>SK</th>
<th>MB</th>
<th>BC</th>
<th>AB</th>
<th>QC</th>
<th>ON</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_{ik}$, 0.003</td>
<td>1.0565</td>
<td>1.0633</td>
<td>1.0669</td>
<td>1.0624</td>
<td>1.0786</td>
<td>1.0804</td>
<td>1.0881</td>
<td>1.0807</td>
<td>1.0941</td>
<td>1.1500</td>
</tr>
<tr>
<td></td>
<td>(0.0045)</td>
<td>(0.0044)</td>
<td>(0.0051)</td>
<td>(0.0047)</td>
<td>(0.0054)</td>
<td>(0.0052)</td>
<td>(0.0054)</td>
<td>(0.0053)</td>
<td>(0.0052)</td>
<td>(0.0063)</td>
</tr>
<tr>
<td>$p_{ik}$, 0.02</td>
<td>1.0170</td>
<td>1.0154</td>
<td>1.0195</td>
<td>1.0213</td>
<td>1.0286</td>
<td>1.0316</td>
<td>1.0238</td>
<td>1.0225</td>
<td>1.0238</td>
<td>1.0465</td>
</tr>
<tr>
<td></td>
<td>(0.0020)</td>
<td>(0.0021)</td>
<td>(0.0024)</td>
<td>(0.0024)</td>
<td>(0.0027)</td>
<td>(0.0029)</td>
<td>(0.0028)</td>
<td>(0.0028)</td>
<td>(0.0031)</td>
<td>(0.0038)</td>
</tr>
<tr>
<td>$\tilde{M}_{ik}$, 0.003</td>
<td>1.0140</td>
<td>1.0188</td>
<td>1.0226</td>
<td>1.0244</td>
<td>1.0318</td>
<td>1.0383</td>
<td>1.0477</td>
<td>1.0450</td>
<td>1.0561</td>
<td>1.1049</td>
</tr>
<tr>
<td></td>
<td>(0.0045)</td>
<td>(0.0044)</td>
<td>(0.0051)</td>
<td>(0.0047)</td>
<td>(0.0054)</td>
<td>(0.0052)</td>
<td>(0.0054)</td>
<td>(0.0053)</td>
<td>(0.0052)</td>
<td>(0.0063)</td>
</tr>
<tr>
<td>$\tilde{M}_{ik}$, 0.02</td>
<td>1.0062</td>
<td>1.0075</td>
<td>1.0101</td>
<td>1.0120</td>
<td>1.0165</td>
<td>1.0191</td>
<td>1.0170</td>
<td>1.0165</td>
<td>1.0183</td>
<td>1.0362</td>
</tr>
<tr>
<td></td>
<td>(0.0010)</td>
<td>(0.0012)</td>
<td>(0.0014)</td>
<td>(0.0015)</td>
<td>(0.0018)</td>
<td>(0.0019)</td>
<td>(0.0020)</td>
<td>(0.0020)</td>
<td>(0.0022)</td>
<td>(0.0029)</td>
</tr>
</tbody>
</table>

6.4 Comparison of Predictors Calculated with Direct Estimators of Sampling Variances to Predictors Calculated with Model Estimators of Sampling Variances

In this section, the empirical MSE’s of predictors constructed under the working model are compared to the empirical MSE’s of predictors calculated with the direct estimators of the sampling variances. The weight $\tilde{\gamma}_{ik,B}$ assigned to the direct estimator in the predictor $\tilde{P}_{pred,ik,B}$ of (4.33) is constructed under the working model for the sampling variances. An alternative is to treat the direct estimators of the sampling variances as the true sampling variances and define the initial predictor

$$\tilde{P}_{pred,ik,B,dir} = \tilde{\gamma}_{ik,B,dir} \tilde{P}_{ik} + (1 - \tilde{\gamma}_{ik,B,dir}) \tilde{P}_{T,ik}, \quad (6.14)$$

where

$$\tilde{\gamma}_{ik,B,dir} = \frac{\psi \tilde{P}_{T,ik}(1 - \tilde{P}_{T,ik})(\psi \tilde{P}_{T,ik}(1 - \tilde{P}_{T,ik}) + \tilde{\sigma}^2 e,ik) + \tilde{V} \{\psi\} \tilde{P}_{T,ik}(1 - \tilde{P}_{T,ik})^2}{(\psi \tilde{P}_{T,ik}(1 - \tilde{P}_{T,ik}) + \tilde{\sigma}^2 e,ik)^2 + \tilde{V} \{\psi\} \tilde{P}_{T,ik}(1 - \tilde{P}_{T,ik})^2}. \quad (6.15)$$
Let $\hat{p}_{ik,B,dir}$ denote the predictor obtained from benchmarking $\hat{P}_{pred,ik,B,dir}$ to the direct estimators of the marginal totals. If a direct estimator of a proportion is zero, then the corresponding direct estimator of the sampling variance is also zero, and the initial predictor $\hat{p}_{pred,ik,B,dir}$ is equal to $\hat{p}_{ik}$. A way to avoid zero estimates of sampling variances is to use the modified estimator of the sampling variance $\hat{\sigma}^2_{e,ik,md}$ defined in (4.7) instead of the direct estimator of the sampling variance. Let $\hat{p}_{pred,ik,B,md}$ denote the predictor obtained by replacing $\hat{\sigma}^2_{e,ik}$ in $\hat{\gamma}_{ik,B,dir}$ with the modified direct estimator, and let $\hat{p}_{ik,B,md}$ denote the corresponding benchmarked predictor.

Table 6.16 contains the ratios of the MC MSE’s of the predictors calculated with the direct estimators of the sampling variances (no modification for zeros) to the MC MSE’s of the predictors calculated with the model estimates of the sampling variances for the two stage simulation model with $\psi = 0.003$. The MC MSE’s of the predictors constructed with the direct estimates of the sampling variances are uniformly larger than the MC MSE’s of the predictors constructed under the working model for the sampling variances. The MSE ratios are larger in categories $i = 2$ and $i = 4$ than in categories $i = 1$ and $i = 3$ in all of the provinces except Ontario. The proportions in categories $i = 2$ and $i = 4$ are smaller than the corresponding proportions in categories $i = 1$ and $i = 3$. In the largest province, Ontario, the MSE ratios are close to one because the predictors are close to the direct estimators, regardless of which estimator of the sampling variance is used.
Table 6.16 Ratios of MC MSE’s of predictors calculated with direct estimator (no modification for zeros) to MC MSE’s of predictors calculated with model estimator of sampling covariance matrix.

2-stage simulation model, $\psi = 0.003$.

The averages of the ratios of the MC MSE’s of the predictors $\{\tilde{p}_{ik,B,dir} : i = 1, \ldots, C; k = 1, \ldots, K\}$ to the MC MSE’s of the predictors $\{\tilde{p}_{ik,B} : i = 1, \ldots, C; k = 1, \ldots, K\}$ are given in Table 6.17. The averages of the MSE ratios are larger than one in all of the provinces. The average MSE ratios approach 1 as the province sizes increase because the predictors are close to the direct estimators in the larger provinces regardless of how the direct estimator of the covariance matrix is calculated. For both the SRS and 2-stage simulation models, the gain in efficiency due to use of the model estimator of the sampling variance is larger when $\psi = 0.003$ than when $\psi = 0.02$. For both values of $\psi$, the average MSE ratios are larger for the two stage simulation model than for the SRS simulation model in most provinces. We conjecture that the average MSE ratios are larger for the two stage simulation model because the ratios of the MC variances of the direct estimators of the sampling variances to the MC variances of the model based estimators of the sampling variances are larger for the two stage simulation model than for the SRS simulation model. (See Table 6.18.)
two stage simulation model is close to a simple random sample because each province has only
two distinct design effects, and the design effects are between 1 and 2 in all of the provinces.
(See tables 6.3 and 6.4.)

<table>
<thead>
<tr>
<th></th>
<th>PE</th>
<th>NF</th>
<th>NB</th>
<th>NS</th>
<th>SK</th>
<th>MB</th>
<th>BC</th>
<th>AB</th>
<th>QC</th>
<th>ON</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRS,0.003</td>
<td>4.78</td>
<td>2.355</td>
<td>1.354</td>
<td>1.142</td>
<td>1.0786</td>
<td>1.0400</td>
<td>1.0181</td>
<td>1.0236</td>
<td>1.0152</td>
<td>1.00525</td>
</tr>
<tr>
<td></td>
<td>(0.13)</td>
<td>(0.057)</td>
<td>(0.026)</td>
<td>(0.012)</td>
<td>(0.0091)</td>
<td>(0.0047)</td>
<td>(0.0019)</td>
<td>(0.0024)</td>
<td>(0.0013)</td>
<td>(0.00081)</td>
</tr>
<tr>
<td>SRS,0.02</td>
<td>1.936</td>
<td>1.395</td>
<td>1.1428</td>
<td>1.1056</td>
<td>1.0609</td>
<td>1.0429</td>
<td>1.0236</td>
<td>1.0246</td>
<td>1.0186</td>
<td>1.0072</td>
</tr>
<tr>
<td></td>
<td>(0.031)</td>
<td>(0.016)</td>
<td>(0.0075)</td>
<td>(0.0058)</td>
<td>(0.0040)</td>
<td>(0.0034)</td>
<td>(0.0022)</td>
<td>(0.0023)</td>
<td>(0.0015)</td>
<td>(0.0013)</td>
</tr>
<tr>
<td>2-stage,0.003</td>
<td>5.62</td>
<td>2.752</td>
<td>1.355</td>
<td>1.189</td>
<td>1.0880</td>
<td>1.0552</td>
<td>1.0505</td>
<td>1.0522</td>
<td>1.0431</td>
<td>1.0113</td>
</tr>
<tr>
<td></td>
<td>(0.14)</td>
<td>(0.065)</td>
<td>(0.024)</td>
<td>(0.013)</td>
<td>(0.0098)</td>
<td>(0.0063)</td>
<td>(0.0052)</td>
<td>(0.0045)</td>
<td>(0.0026)</td>
<td>(0.0013)</td>
</tr>
<tr>
<td>2-stage,0.02</td>
<td>2.169</td>
<td>1.546</td>
<td>1.1530</td>
<td>1.1065</td>
<td>1.0559</td>
<td>1.0425</td>
<td>1.0446</td>
<td>1.0367</td>
<td>1.0333</td>
<td>1.0096</td>
</tr>
<tr>
<td></td>
<td>(0.037)</td>
<td>(0.021)</td>
<td>(0.0087)</td>
<td>(0.0061)</td>
<td>(0.0047)</td>
<td>(0.0040)</td>
<td>(0.0035)</td>
<td>(0.0032)</td>
<td>(0.0023)</td>
<td>(0.0017)</td>
</tr>
</tbody>
</table>

Table 6.17 Averages of ratios of MC MSE’s of predictors calculated with
direct estimator (no modification for zeros) to MC MSE’s of
predictors calculated with model estimator of sampling covariance matrix.

<table>
<thead>
<tr>
<th></th>
<th>PE</th>
<th>NF</th>
<th>NB</th>
<th>NS</th>
<th>SK</th>
<th>MB</th>
<th>BC</th>
<th>AB</th>
<th>QC</th>
<th>ON</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRS</td>
<td>2.60</td>
<td>2.61</td>
<td>2.58</td>
<td>2.01</td>
<td>2.58</td>
<td>2.29</td>
<td>2.23</td>
<td>2.34</td>
<td>1.73</td>
<td>1.90</td>
</tr>
</tbody>
</table>

Table 6.18 Averages of ratios of MC variances of direct estimators of sam-
pling variances (no modification for zeros) to MC variances of
model based estimators of sampling variances, $\psi = 0.003$

Table 6.19 shows the averages of the MSE ratios when the direct estimator of the
sampling covariance matrix is modified to eliminate zero estimates. In the largest five provinces,
where zero estimates are relatively rare, the averages of the MSE ratios in Table 6.19 do not differ significantly from the averages of the MSE ratios in Table 6.17. The averages of the MSE ratios in Table 6.19 are closest to 1 in the smallest two provinces because the probabilities of zeros are largest in the smallest two provinces, and the modified direct estimator is equal to the model based estimator when a direct estimator of a proportion is zero. In the SRS simulation model when $\psi = 0.003$, the ratios of the MSE’s of predictors based on the modified direct estimators of the sampling variances to the MSE’s of predictors based on the model estimators of the sampling variances in categories $i = 2$ and $i = 4$ in the province labeled Prince Edward Island are 0.9984 (MC SE, 0.0025) and 0.9902 (MC SE, 0.0027), respectively. Similarly, in the simulation from the two stage model with $\psi = 0.003$, the MSE ratios are 0.9965 (MC SE, 0.0039) and 0.9971 (MC SE, 0.0039) in category $i = 2$ for Prince Edward Island and New Foundland, respectively. The probabilities of zeros are greatest in categories $i = 2$ and $i = 4$ in Prince Edward Island and New Foundland. For example, in the 2-stage simulation model with $\psi = 0.003$, the empirical probability that a direct estimate for category $i = 2$ in Prince Edward Island is zero is 0.1596, and the empirical probability that a direct estimate for category $i = 2$ in New Foundland is zero is 0.1028. For comparison, the next largest proportion of zero direct estimates is 0.0276, which occurs category $i = 4$ in the province labeled New Brunswick.
<table>
<thead>
<tr>
<th></th>
<th>PE</th>
<th>NF</th>
<th>NB</th>
<th>NS</th>
<th>SK</th>
<th>MB</th>
<th>BC</th>
<th>AB</th>
<th>QC</th>
<th>ON</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRS,0.003</td>
<td>1.0018</td>
<td>1.0119</td>
<td>1.0292</td>
<td>1.0277</td>
<td>1.0347</td>
<td>1.0330</td>
<td>1.0187</td>
<td>1.0243</td>
<td>1.0144</td>
<td>1.00465</td>
</tr>
<tr>
<td></td>
<td>(0.0015)</td>
<td>(0.0020)</td>
<td>(0.0031)</td>
<td>(0.0023)</td>
<td>(0.0033)</td>
<td>(0.0030)</td>
<td>(0.0019)</td>
<td>(0.0024)</td>
<td>(0.0012)</td>
<td>(0.00072)</td>
</tr>
<tr>
<td>SRS,0.02</td>
<td>1.0236</td>
<td>1.0322</td>
<td>1.0403</td>
<td>1.0378</td>
<td>1.0441</td>
<td>1.0324</td>
<td>1.0230</td>
<td>1.0229</td>
<td>1.0170</td>
<td>1.0056</td>
</tr>
<tr>
<td></td>
<td>(0.0029)</td>
<td>(0.0029)</td>
<td>(0.0033)</td>
<td>(0.0034)</td>
<td>(0.0030)</td>
<td>(0.0029)</td>
<td>(0.0021)</td>
<td>(0.0022)</td>
<td>(0.0015)</td>
<td>(0.0012)</td>
</tr>
<tr>
<td>2-stage,0.003</td>
<td>1.0118</td>
<td>1.0202</td>
<td>1.0297</td>
<td>1.0320</td>
<td>1.0343</td>
<td>1.0350</td>
<td>1.0422</td>
<td>1.0494</td>
<td>1.0387</td>
<td>1.0106</td>
</tr>
<tr>
<td></td>
<td>(0.0052)</td>
<td>(0.0030)</td>
<td>(0.0035)</td>
<td>(0.0031)</td>
<td>(0.0039)</td>
<td>(0.0036)</td>
<td>(0.0035)</td>
<td>(0.0039)</td>
<td>(0.0024)</td>
<td>(0.0013)</td>
</tr>
<tr>
<td>2-stage,0.02</td>
<td>1.0090</td>
<td>1.0365</td>
<td>1.0310</td>
<td>1.0349</td>
<td>1.0316</td>
<td>1.0261</td>
<td>1.0411</td>
<td>1.0343</td>
<td>1.0276</td>
<td>1.0079</td>
</tr>
<tr>
<td></td>
<td>(0.0035)</td>
<td>(0.0037)</td>
<td>(0.0041)</td>
<td>(0.0035)</td>
<td>(0.0035)</td>
<td>(0.0033)</td>
<td>(0.0033)</td>
<td>(0.0031)</td>
<td>(0.0022)</td>
<td>(0.0016)</td>
</tr>
</tbody>
</table>

Table 6.19  Averages of ratios of MC MSE’s of predictors calculated with modified direct estimator (modified for zeros) to MC MSE’s of predictors calculated with model estimator of sampling covariance matrix.

Use of the working model for the sampling variances leads to more efficient predictors under the 2-stage simulation model even though the estimator of the sampling covariance matrix based on the working model is biased for the true sampling covariance matrix. Table 6.20 shows standardized MC biases of the model estimators of the sampling variances for the 2-stage simulation model with $\psi = 0.003$. The MC biases for the 2-stage simulation model with $\psi = 0.02$ are similar and omitted. The standardized bias for category $i$ in province $k$ is

$$
\tilde{b}_{ik,mc, std} = \frac{\text{bias}_{ik,mc}}{\sqrt{\hat{V}_{ik,b,mc}}},
$$

(6.16)

where $\text{bias}_{ik,mc} = M^{-1} \sum_{j=1}^{M} [\hat{\sigma}_{ik}^{2(j)} - (\hat{p}_{ik}^{(j)} - p_{ik})^2]$, $\hat{\sigma}_{ik}^{2(j)}$ denotes an estimator of a sampling variance (direct or working model) in simulation trial $j$. 

$$
\hat{V}_{ik,b,mc} = [M(M-1)]^{-1} \sum_{j=1}^{M} [\hat{\sigma}_{ik}^{2(j)} - (\hat{p}_{ik}^{(j)} - p_{ik})^2 - \text{bias}_{ik,mc}]^2,
$$

(6.17)
The biases of the model estimators of the sampling variances in Table 6.20 are consistent with the properties of the sampling variance model described above in conjunction with Table 6.4. In the provinces other than those labeled British Columbia and Alberta, the true \( c_k p_{T,ik} (1 - p_{T,ik}) n_k^{-1} \) is larger than the empirical sampling variance in categories \( i = 1, 2, \) and 3 and smaller than the empirical sampling variance in category \( i = 4 \). As a consequence, the MC biases of the model estimators of the sampling variances are positive in categories \( i = 1, 2, \) and 3 but negative in category \( i = 4 \) in provinces other than those labeled British Columbia and Alberta. In the provinces labeled British Columbia and Alberta, the design is a cluster sample, so the working model for the sampling variances is true, and the model estimators of the sampling variances are unbiased for the true sampling variances.

<table>
<thead>
<tr>
<th>PE</th>
<th>NF</th>
<th>NB</th>
<th>NS</th>
<th>SK</th>
<th>MB</th>
<th>BC</th>
<th>AB</th>
<th>QC</th>
<th>ON</th>
</tr>
</thead>
<tbody>
<tr>
<td>( i = 1 )</td>
<td>5.40</td>
<td>5.20</td>
<td>4.61</td>
<td>4.01</td>
<td>3.73</td>
<td>3.66</td>
<td>0.54</td>
<td>-0.02</td>
<td>2.19</td>
</tr>
<tr>
<td>( i = 2 )</td>
<td>4.03</td>
<td>1.82</td>
<td>1.97</td>
<td>3.75</td>
<td>4.31</td>
<td>6.76</td>
<td>0.11</td>
<td>1.12</td>
<td>3.18</td>
</tr>
<tr>
<td>( i = 3 )</td>
<td>3.71</td>
<td>4.39</td>
<td>3.18</td>
<td>4.53</td>
<td>2.85</td>
<td>7.42</td>
<td>-1.40</td>
<td>0.77</td>
<td>2.07</td>
</tr>
<tr>
<td>( i = 4 )</td>
<td>-7.22</td>
<td>-7.53</td>
<td>-4.38</td>
<td>-2.57</td>
<td>-4.58</td>
<td>-9.29</td>
<td>-0.82</td>
<td>-0.94</td>
<td>-5.53</td>
</tr>
</tbody>
</table>

Table 6.20  Standardized biases of model estimators of sampling variances, 2-stage sampling error model, \( \psi = 0.003 \)

Two reasons for the gain in efficiency due to using the biased model estimator of the covariance matrix instead of the unbiased direct estimator are that the model estimators of the sampling variances have smaller variances and are less correlated with the direct estimators of the proportions than the direct estimators of the sampling variances. Table 6.21 shows the ratios of the MC MSE’s of the working model estimators of the sampling variances to the MC MSE’s of the direct estimators of the sampling variances, where the direct estimators are not modified for zeros. The MC MSE is \( [\text{bias}_{ik,mc}]^2 + M\tilde{\text{V}}_{ik,b,mc} \), where the MC bias and variance are defined in (6.16) and (6.17), respectively. Although the bias of the model estimator of the covariance matrix is significant in many cases, the increase in the bias is negligible relative to
the reduction in the variance due to using the model instead of the direct estimator. As a consequence, the MSE’s of the model estimators of the sampling variances are smaller than the MSE’s of the corresponding direct estimators. The reduction in the MSE due to use of the working covariance structure is greatest in categories $i = 2$ and $i = 4$ in the smallest two provinces, where the empirical probabilities of zeros are largest. Tables 6.22 and 6.23 show the empirical correlations between the weight assigned to the direct estimator used to form the initial predictor and the residual, $\hat{p}_{ik} - \hat{p}_{T,ik}$ for the 2-stage simulation model with $\psi = 0.003$. Because $\hat{\sigma}_{e,ik}^2$ is positively correlated with the residual, the weight based on the direct estimator of the sampling variance defined in (6.15) is negatively correlated with $\hat{p}_{ik} - \hat{p}_{T,ik}$. Modifying the direct estimators of the sampling variances to eliminate zero estimates reduces the absolute values of the correlations in categories $i = 2$ and $i = 4$ in the smallest four provinces but has little or no effect on the empirical correlations in the larger provinces. Use of the working model for the sampling variances substantially reduces the correlations between the weight assigned to the direct estimator in the initial predictor and the residual, $\hat{p}_{ik} - \hat{p}_{T,ik}$ (Table 6.23). The patterns in Tables 6.21 - 6.23 for $\psi = 0.003$ also hold for $\psi = 0.02$.

<table>
<thead>
<tr>
<th></th>
<th>PE</th>
<th>NF</th>
<th>NB</th>
<th>NS</th>
<th>SK</th>
<th>MB</th>
<th>BC</th>
<th>AB</th>
<th>QC</th>
<th>ON</th>
</tr>
</thead>
<tbody>
<tr>
<td>$i = 1$</td>
<td>0.297</td>
<td>0.357</td>
<td>0.457</td>
<td>0.494</td>
<td>0.524</td>
<td>0.644</td>
<td>0.334</td>
<td>0.336</td>
<td>0.467</td>
<td>0.521</td>
</tr>
<tr>
<td>$i = 2$</td>
<td>0.205</td>
<td>0.210</td>
<td>0.227</td>
<td>0.299</td>
<td>0.284</td>
<td>0.451</td>
<td>0.181</td>
<td>0.167</td>
<td>0.280</td>
<td>0.364</td>
</tr>
<tr>
<td>$i = 3$</td>
<td>0.238</td>
<td>0.387</td>
<td>0.428</td>
<td>0.480</td>
<td>0.481</td>
<td>0.871</td>
<td>0.346</td>
<td>0.370</td>
<td>0.484</td>
<td>0.536</td>
</tr>
<tr>
<td>$i = 4$</td>
<td>0.094</td>
<td>0.165</td>
<td>0.166</td>
<td>0.234</td>
<td>0.183</td>
<td>0.348</td>
<td>0.170</td>
<td>0.164</td>
<td>0.308</td>
<td>0.336</td>
</tr>
</tbody>
</table>

Table 6.21  Ratios of MC MSE’s of model estimators of sampling variances to MC MSE’s of direct estimators of sampling variances, 2-stage simulation, $\psi = 0.003$
The correlation between the weight $\hat{\gamma}_{ik,B,\text{dir}}$ and the residual $\hat{p}_{ik} - \hat{p}_{T,ik}$ leads to a bias in the predictor. Tables 6.24, 6.25, and 6.26 show MC standardized biases of the predictors constructed with the direct estimators, the modified direct estimators, and the model estimators of the sampling variances. The standardized bias for category $i$ in province $k$ is $b_{ik,p,st} = b_{ik,p}(\sqrt{V_{b,ik,p}})^{-1}$, where

$$b_{ik,p} = M^{-1} \sum_{j=1}^{M} (p_{ik,pred}^{(j)} - p_{ik}^{(j)}),$$

$$V_{b,ik,p} = [M(M-1)]^{-1} \sum_{j=1}^{M} (p_{ik,pred}^{(j)} - p_{i,k}^{(j)} - b_{ik,p})^2,$$

$p_{ik,pred}^{(j)}$ denotes a predictor obtained in simulation trial $j$, $p_{ik}^{(j)}$ is the true value obtained in simulation trial $j$, and $M = 5000$ is the MC sample size. The absolute values of the standard-
ized biases are largest in the smallest two provinces when the direct estimator of the sampling variance with no modification for zeros is used to compute the predictors. Modifying the direct estimator in cases of zero estimates substantially reduces the absolute values of the standardized biases in the smallest two provinces. The predictors based on the model estimators of the sampling variances have a negative biases in category $i = 4$ in Prince Edward Island. Otherwise, the biases of the predictors calculated with the model estimators of the sampling variances are negligible in the smallest two provinces. The differences between the standardized biases of predictors calculated with the direct estimators of the sampling variances and the standardized biases of predictors calculated with the modified direct estimators of the sampling variances tend to decrease as the province sizes increase because the probability of a zero estimated variance decreases. For the two stage simulation model, the predictors calculated with the model estimator of the sampling variances have the smallest absolute standardized biases in 22 of the cells when $\psi = 0.003$, and the model based predictors have the smallest absolute standardized biases in 26 cells when $\psi = 0.02$.

<table>
<thead>
<tr>
<th></th>
<th>PE</th>
<th>NF</th>
<th>NB</th>
<th>NS</th>
<th>SK</th>
<th>MB</th>
<th>BC</th>
<th>AB</th>
<th>QC</th>
<th>ON</th>
</tr>
</thead>
<tbody>
<tr>
<td>$i = 1$</td>
<td>26.14</td>
<td>18.45</td>
<td>3.13</td>
<td>0.44</td>
<td>3.87</td>
<td>2.68</td>
<td>0.49</td>
<td>-0.73</td>
<td>-0.74</td>
<td>-3.79</td>
</tr>
<tr>
<td>$i = 2$</td>
<td>-15.46</td>
<td>-14.32</td>
<td>-4.01</td>
<td>-3.79</td>
<td>-1.37</td>
<td>-2.12</td>
<td>0.64</td>
<td>-3.06</td>
<td>1.13</td>
<td>5.99</td>
</tr>
<tr>
<td>$i = 3$</td>
<td>15.75</td>
<td>11.35</td>
<td>5.96</td>
<td>2.06</td>
<td>2.96</td>
<td>1.42</td>
<td>0.39</td>
<td>4.73</td>
<td>0.36</td>
<td>0.41</td>
</tr>
<tr>
<td>$i = 4$</td>
<td>-34.23</td>
<td>-14.42</td>
<td>-6.14</td>
<td>0.97</td>
<td>-7.73</td>
<td>-3.29</td>
<td>-1.78</td>
<td>-2.67</td>
<td>-0.56</td>
<td>-1.96</td>
</tr>
</tbody>
</table>

Table 6.24 Standardized biases of benchmarked predictors calculated with direct estimators of sampling variances, 2-stage, $\psi = 0.003$
<table>
<thead>
<tr>
<th>PE</th>
<th>NF</th>
<th>NB</th>
<th>NS</th>
<th>SK</th>
<th>MB</th>
<th>BC</th>
<th>AB</th>
<th>QC</th>
<th>ON</th>
</tr>
</thead>
<tbody>
<tr>
<td>$i = 1$</td>
<td>0.19</td>
<td>-0.05</td>
<td>-0.66</td>
<td>-2.12</td>
<td>3.57</td>
<td>2.85</td>
<td>0.88</td>
<td>-0.32</td>
<td>-0.37</td>
</tr>
<tr>
<td>$i = 2$</td>
<td>3.39</td>
<td>0.23</td>
<td>-1.03</td>
<td>0.60</td>
<td>-1.52</td>
<td>-2.36</td>
<td>0.32</td>
<td>-3.36</td>
<td>0.62</td>
</tr>
<tr>
<td>$i = 3$</td>
<td>-1.99</td>
<td>-2.76</td>
<td>2.15</td>
<td>-0.59</td>
<td>2.57</td>
<td>1.50</td>
<td>0.68</td>
<td>5.08</td>
<td>0.65</td>
</tr>
<tr>
<td>$i = 4$</td>
<td>-0.74</td>
<td>3.18</td>
<td>-0.91</td>
<td>2.09</td>
<td>-7.02</td>
<td>-3.49</td>
<td>-2.39</td>
<td>-3.45</td>
<td>-0.90</td>
</tr>
</tbody>
</table>

Table 6.25  Standardized biases of benchmarked predictors calculated with modified direct estimators of sampling variances, 2-stage, $\psi = 0.003$

<table>
<thead>
<tr>
<th>PE</th>
<th>NF</th>
<th>NB</th>
<th>NS</th>
<th>SK</th>
<th>MB</th>
<th>BC</th>
<th>AB</th>
<th>QC</th>
<th>ON</th>
</tr>
</thead>
<tbody>
<tr>
<td>$i = 1$</td>
<td>1.73</td>
<td>0.76</td>
<td>-0.95</td>
<td>-2.61</td>
<td>2.20</td>
<td>1.73</td>
<td>0.14</td>
<td>-1.02</td>
<td>-0.87</td>
</tr>
<tr>
<td>$i = 2$</td>
<td>1.57</td>
<td>-0.42</td>
<td>0.32</td>
<td>2.40</td>
<td>-0.14</td>
<td>-0.61</td>
<td>1.83</td>
<td>-0.92</td>
<td>0.67</td>
</tr>
<tr>
<td>$i = 3$</td>
<td>-0.22</td>
<td>-1.03</td>
<td>1.68</td>
<td>-0.49</td>
<td>1.60</td>
<td>0.77</td>
<td>0.12</td>
<td>3.72</td>
<td>0.38</td>
</tr>
<tr>
<td>$i = 4$</td>
<td>-4.29</td>
<td>0.73</td>
<td>-1.21</td>
<td>1.05</td>
<td>-5.26</td>
<td>-2.81</td>
<td>-2.16</td>
<td>-3.07</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Table 6.26  Standardized biases of benchmarked predictors calculated with model estimators of sampling variances, 2-stage, $\psi = 0.003$

6.5 Comparison of Bootstrap and Taylor MSE Estimators

Two types of estimators of the MSE’s of the predictors are proposed in Section 5. In Section 5.1.1 and Section 5.1.2, closed form approximations for the MSE’s of the initial (not benchmarked) predictors are derived using Taylor linearizations. In Section 5.2 a bootstrap MSE estimator is proposed. In this section, the empirical relative biases of the Taylor MSE estimators of Sections 5.1.1 and 5.1.2 are compared to the empirical relative biases of the bootstrap MSE estimators of Section 5.2.

The MC relative bias of an MSE estimator, $\widehat{MSE}_{ik}$, for the MSE of the benchmarked
predictor of the proportion is defined,

\[ MCRB = 100 \frac{E_{MC}\left[\hat{MSE}_{ik}\right] - MSE_{MC}(\hat{p}_{ik,B})}{MSE_{MC}(\hat{p}_{ik,B})}, \]  

(6.18)

where

\[ E_{MC}\left[\hat{MSE}_{ik}\right] = M^{-1} \sum_{j=1}^{M} \hat{MSE}_{ik}^{(j)}, \]

\[ MSE_{MC}(\hat{p}_{ik}) = M^{-1} \sum_{j=1}^{M} (\hat{p}_{ik,B}^{(j)} - \hat{p}_{ik}^{(j)})^2, \]

\( j \) denotes the simulation run, and \( \hat{MSE}_{ik}^{(j)}, \hat{p}_{ik,B}^{(j)}, \) and \( \hat{p}_{ik}^{(j)} \) denote the MSE estimate, the predictor, and the true proportion for cell \( ik \) in simulation run \( j \). The MC relative bias of the Taylor estimator of the MSE of the predictor of the proportion is obtained by setting \( \hat{MSE}_{ik} = \hat{MSE}_{ik,2,B} \), where \( \hat{MSE}_{ik,2,B} \) is defined in (5.14). The MC relative bias of the bootstrap estimator of the MSE of the predictor of the proportion is obtained by setting \( \hat{MSE}_{ik} = \hat{MSE}_{bs}^{ik} \), where \( \hat{MSE}_{bs}^{ik} \) is defined in (5.36). The MC relative biases of the Taylor and bootstrap MSE estimators for the totals are defined analogously. The Taylor estimator of the MSE of the predictor of the total is denoted \( \hat{MSE}_{2,ik,T,B} \) and is defined in (5.24). The bootstrap MSE estimator for the total, denoted \( \hat{MSE}_{bs}^{ik,T} \), is defined following (5.36).

Figures 6.3 - 6.10 show the MC relative biases of the MSE estimators for the proportions and totals for the four simulation models. The provinces are listed on the horizontal axis of each plot in decreasing order with respect to the coefficient of variation of the direct estimator of the province two digit total. Each province has four points for the MC relative biases of the Taylor MSE estimators and four points for the MC relative biases of the bootstrap MSE estimators. The MC relative biases of the Taylor MSE estimators are labeled “T”, and the MC relative biases of the bootstrap MSE estimators are labeled “B”. The MC relative biases of the Taylor or bootstrap MSE estimates that exceed two MC standard deviations in absolute value are marked with a “*” or a “#”, respectively.

The MSE of the benchmarked predictor can be written as a sum of three terms:

\[ MSE(\hat{p}_{ik,B}) = LT_{ik} + Parest_{ik} + Rake_{ik}, \]
where $LT_{ik} = E[(p_{pred, ik}(p_{T, ik}, \gamma_{k}) - p_{ik})^2]$, $p_{pred, ik}(p_{T, ik}, \gamma_{k}) = \gamma_{k}\hat{p}_{ik} + (1 - \gamma_{k})p_{T, ik}$, $Parest_{ik} = E[(\hat{p}_{pred, ik, B} - p_{ik})^2] - LT_{ik}$, $\hat{p}_{pred, ik, B} = \hat{\gamma}_{k,B}\hat{p}_{ik} + (1 - \hat{\gamma}_{k,B})\hat{p}_{T, ik}$, and $Rake_{ik} = E[(\hat{p}_{ik, B} - p_{ik})^2] - E[(\hat{p}_{pred, ik, B} - p_{ik})^2]$. The leading term, $LT_{ik}$, is the MSE of a predictor calculated with the true parameters. The second term, $Parest_{ik}$ is the difference between the MSE of the initial (not benchmarked) predictor and the MSE of the predictor constructed with the true parameters. The third term, $Rake_{ik}$ is the difference between the MSE of the benchmarked predictor and the MSE of the initial convex combination of $\hat{p}_{ik}$ and $\hat{p}_{T, ik}$. The MSE of a predictor of a total has an analogous decomposition. The Monte Carlo estimates of the three components of the MSE’s are obtained by substituting expectations for appropriate averages of squared differences. We use the notation $LT_{ik}$, $Parest_{ik}$, and $Rake_{ik}$ to denote both the true values and the MC estimates. For simplicity, we refer to $Parest_{ik}$ and $Rake_{ik}$ as the “effect of parameter estimation” and the “raking effect,” respectively.

The MSE estimators can be expressed as sums of the estimators of the three components of the MSE defined above. A bias corrected estimator of the leading term based on the approach of Rao (2003, pg. 104) is

$$\hat{L}T_{ik,T} = \hat{MSE}_{11,ik} + \hat{g}_{4,ik},$$

where $\hat{MSE}_{11,ik}$ is defined in (5.11), and $-\hat{g}_{4,ik}$, defined in (5.13), is an estimator of the bias of $\hat{MSE}_{11,ik}$ for the leading term. To distinguish between $\hat{MSE}_{11,ik}$ and $\hat{L}_{ik,T}$, we refer to $\hat{MSE}_{11,ik}$ as the “Taylor estimator of the leading term,” and we call $\hat{L}_{ik,T}$ the bias corrected Taylor estimator of the leading term. The bias corrected bootstrap estimator of the leading term is

$$\hat{L}T_{ik,bs} = \hat{MSE}_{11,ik}^{bs} - \hat{b}_{ik}^{bs},$$

where

$$\hat{MSE}_{11,ik}^{bs} = B^{-1}\sum_{b=1}^{B} (\hat{\gamma}_{k,B}\hat{p}_{ik, B} + (1 - \hat{\gamma}_{k,B})\hat{p}_{T, ik} - p_{ik}^{*(b)})^2,$$

$b$ denotes the bootstrap sample, and $\hat{b}_{ik}^{bs}$, defined in (5.36) of Section 5.5, is a bootstrap estimate of the bias of $\hat{MSE}_{11,ik}$ for the leading term. To distinguish between $\hat{MSE}_{11,ik}^{bs}$ and $\hat{L}T_{ik,bs}$,
we refer to \( \hat{MSE}_{11,ik}^{bs} \) as the bootstrap estimator of the leading term, and we refer to \( \hat{LT}_{ik,bs} \) as the bias corrected bootstrap estimator of the leading term. The Taylor estimator of \( Parest_{ik} \), the effect of estimating the parameters on the MSE, is the sum of \( \hat{g}_{2,ik} \) and \( \hat{g}_{3,ik} \), defined in Section 5.1, where \( \hat{g}_{2,ik} \) is an estimate of \( E[(1 - \gamma_k)^2(\hat{p}_{T,ik} - p_{T,ik})^2] \), and \( \hat{g}_{3,ik} \) estimates \( E[(\hat{\gamma}_{k,B} - \gamma_k)^2(\hat{p}_{ik} - p_{T,ik})^2] \). The bootstrap estimator of the effect of parameter estimation is

\[
\hat{Parest}_{ik,bs} = B^{-1} \sum_{b=1}^{B} (\hat{p}_{\text{pred},ik,B}^{(b)} - \hat{p}_{ik}^{(b)})^2 - \hat{MSE}_{11,ik}^{bs},
\]

where \( \hat{p}_{\text{pred},ik,B}^{(b)} = \hat{\gamma}_{k,B}^{(b)} p_{ik}^{(b)} - (1 - \hat{\gamma}_{k,B}^{(b)}) \hat{p}_{T,ik}^{(b)} \). The Taylor MSE estimator is an estimator of the MSE of \( \hat{p}_{\text{pred},ik,B} \), so the Taylor estimator of the effect of raking is zero. The bootstrap estimator of the raking effect, the difference between the MSE of the benchmarked predictor and the MSE of the initial predictor, is

\[
\hat{Rake}_{ik} = B^{-1} \sum_{b=1}^{B} (\hat{p}_{\text{pred},ik,B}^{(b)} - \hat{p}_{ik}^{(b)})^2 - B^{-1} \sum_{b=1}^{B} (\hat{p}_{\text{pred},ik,B}^{(b)} - \hat{p}_{ik}^{(b)})^2.
\]

To better understand the nature of the biases of the MSE estimators, we study the empirical properties of three components of the MSE estimators defined above. The empirical MSE’s of the predictors and the MC means of the corresponding Taylor and bootstrap estimators are decomposed into the components in Tables 6.27 - 6.34. The organization of Tables 6.27 - 6.34 is as follows. The cell in the first row and column of each Table contains the percent of the MC MSE of \( \hat{p}_{ik,B} \) explained by the leading term, \( 100(\hat{LT}_{ik})(MSE_{MC}(\hat{p}_{ik,B}))^{-1} \). The second two rows of the first column contain the ratios of the MC means of the Taylor and bootstrap estimators of the leading term to the empirical MSE, \( 100(E_{MC}[\hat{MSE}_{11,ik}^{bs}](MSE_{MC}(\hat{p}_{ik,B}))^{-1}) \) and \( 100(E_{MC}[\hat{MSE}_{11,ik}])(MSE_{MC}(\hat{p}_{ik,B}))^{-1} \), respectively. The first row of the second column contains the ratio of the empirical bias of the Taylor estimator of the leading term to the MC MSE of \( \hat{p}_{ik,B} \) (multiplied by 100): \( 100(E_{MC}[\hat{MSE}_{11,ik}] - \hat{LT}_{ik})(MSE_{MC}(\hat{p}_{ik,B}))^{-1} \). The second two rows of the second column contain the MC means of the Taylor and bootstrap estimators of the bias of the Taylor estimator of the leading term as fractions of the MC MSE: \( 100(E_{MC}[\hat{g}_{4,ik}](MSE_{MC}(\hat{p}_{ik,B}))^{-1} \) and \( 100(E_{MC}[\hat{g}_{ik}^{bs}](MSE_{MC}(\hat{p}_{ik,B}))^{-1} \). The third column gives the percent of the MC MSE explained by the effect of estimating parameters along with the Taylor and bootstrap estimates: \( 100(Parest_{ik})(MSE_{MC}(\hat{p}_{ik,B}))^{-1} \) (row 1),
100(E_{MC}[\hat{P}_{reest,ik,T}](MSE_{MC}(\hat{p}_{ik,B})))^{-1} \text{ (row 2)}, \text{ and } 100(E_{MC}[\hat{P}_{reest,ik,bs}](MSE_{MC}(\hat{p}_{ik,B})))^{-1} \text{ (row 3). The percent of the MC MSE explained by raking and the MC mean of the bootstrap estimate of the raking effect are given in the first and third rows, respectively, of the third column. The Taylor estimator of the raking effect is, by definition, zero.}

6.5.1 SRS, 0.003

Figure 6.3 shows the MC relative biases of the MSE estimators of the predictors of the proportions for the SRS sampling error model with $\psi = 0.003$. For most of the cells, the MC relative biases of both the Taylor and bootstrap MSE estimators are negligible relative to the Monte Carlo error. The Taylor MSE estimators typically have positive MC biases. In the five cells labeled with asterisks, the MC relative biases of the Taylor MSE estimators exceed two MC standard errors. In contrast, the bootstrap MSE estimators typically have a negative MC bias. The magnitudes of the biases of the bootstrap MSE estimators is greater than two MC standard errors in the three cells labeled with a # sign.

The difference between the bootstrap and Taylor MSE estimators in Figure 6.3 is largely explained by the effect of the lower bound on the estimator of $\psi$. For the SRS simulation model with $\psi = 0.003$, 16% of the estimates of $\psi$ are equal to the lower bound, and the Monte Carlo mean of $\hat{\psi}$ is 0.0034. The positive bias of the estimator of $\psi$ contributes to the patterns observed in Figure 6.3 through the estimator of the leading term, and the estimator of the effect of parameter estimation.
Figure 6.3 MC relative biases of MSE estimators for proportions.

B=Bootstrap, T=Taylor. SRS sampling error model, $\psi = 0.003$.

Table 6.27 is split into two sub-tables. The top sub-table contains the MSE decomposition for three digit code $i = 4$ of Prince Edward Island, and the bottom sub-table contains the MSE decomposition for three digit code $i = 1$ of Ontario. Each sub-table is organized as described above.

The top half of Table 6.27 illustrates why the MC means of the Taylor estimators of the MSE’s are larger than MC means of the the bootstrap MSE estimators in the small provinces. Both the bootstrap and Taylor estimators of the leading leading term have positive MC biases. We suspect that the positive MC biases for the estimators of the leading term are due to the positive bias of the estimator of $\psi$ that results from use of the lower bound.
The bootstrap estimate of the bias of the initial estimator of the leading term accounts for the lower bound effect, while the Taylor estimator of the bias does not. As a consequence, the MC mean of the bootstrap estimator of the bias of $\hat{MSE}_{11,ik}$ is closer to the MC bias of the Taylor estimator of the leading term than the MC mean of the Taylor estimator of the bias (column 2 of Table 6.27). The difference between the bootstrap and Taylor estimators of the bias of the estimator of the leading term is the main reason why the MC mean of the Taylor MSE estimators is larger than the MC mean of the bootstrap MSE estimator in category $i = 4$ in Prince Edward Island. The MC mean of the bootstrap estimator of the leading term is somewhat smaller than MC mean of the corresponding Taylor estimator of the leading term. While the MC means of the two estimators of the leading term are similar, we conjecture that the difference between the MC means of the bootstrap and Taylor estimators of the leading term occurs because the bootstrap data generating procedure does not produce unbiased estimators of the variances of the proportions. (See Appendix 4.) The patterns noted for category $i = 4$ of Prince Edward Island are representative of the properties of the MSE estimators in the other categories of Prince Edward Island and in the other small provinces.

The decomposition of the MSE and the MSE estimators for three digit code $i = 1$ of Ontario is given in the bottom subtable of Table 6.27. Both the bootstrap and Taylor estimators of the leading terms are approximately unbiased for the MC leading term. One reason that the MC mean of the Taylor MSE estimator is larger than the MC mean of the bootstrap MSE estimator is that the MC mean of the Taylor estimator of the bias of $\hat{MSE}_{11,ik}$ is smaller than the bootstrap estimator of the bias. A second reason that the MC mean of the Taylor estimator of the MSE is larger than the MC mean of the bootstrap MSE estimator in category $i = 3$ in Ontario is that the Taylor estimator, $\hat{g}_{3,ik}$, of $E[(\hat{\gamma}_k - \gamma_k)^2(u_{ik} + e_{ik})^2]$ has a positive MC bias when $\psi = 0.003$. The positive bias of $\hat{g}_{3,ik}$ when $\psi$ and $\sigma_{e,ik}^2$ are both small is consistent with the results of Wang and Fuller (2003) and explains why the MC average of the Taylor estimators of the effect of parameter estimation is larger than the MC effect of parameter estimation (column 3 of Table 6.27). In three digit code $i = 3$ of Ontario, the raking effect is 8.60% of the MC MSE, and the MC mean of the bootstrap estimator of the raking
The additional variability due to the raking operation reduces the MC bias of the Taylor estimator of the MSE. The MC relative bias of the Taylor MSE estimator of the MSE of the initial (non-benchmarked predictor) is 19, while the MC relative bias of the Taylor MSE estimator of the MSE of the benchmarked predictor, \( \hat{p}_{ik,B} \), is 8.8 in three digit code \( i = 1 \) of Ontario. The behavior of the MSE estimators for three digit code \( i = 3 \) in Ontario is typical of the patterns observed in the other categories in Ontario and in the other large provinces.

<table>
<thead>
<tr>
<th>Leading Term</th>
<th>Bias of Taylor Est. Param. Est.</th>
<th>Raking</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prince Edward Island, ( i = 4 ), ( MSE(\hat{p}_{ik,B}) = 0.14% )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Empirical</td>
<td>60.15</td>
<td>6.26</td>
</tr>
<tr>
<td>Taylor</td>
<td>66.41</td>
<td>-1.29</td>
</tr>
<tr>
<td>Bootstrap</td>
<td>65.63</td>
<td>5.98</td>
</tr>
<tr>
<td>Ontario, ( i = 1 ), ( MSE(\hat{p}_{ik,B}) = 0.070% )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Empirical</td>
<td>58.94</td>
<td>-0.45</td>
</tr>
<tr>
<td>Taylor</td>
<td>58.48</td>
<td>-9.66</td>
</tr>
<tr>
<td>Bootstrap</td>
<td>58.39</td>
<td>1.17</td>
</tr>
</tbody>
</table>

Table 6.27 Percent of MC MSE of \( \hat{p}_{ik,B} \) explained by leading term (column 1), effect of parameter estimation (column 3), and the raking effect (column 4). Column 2 has ratios of empirical biases (row 1) and MC means of the Taylor (row 2) and bootstrap (row 3) estimators of biases of the estimator of the leading term to the MC MSE of \( \hat{p}_{ik,B} \) (multiplied by 100). SRS, \( \psi = 0.003 \)

Figure 6.4 shows the MC relative biases for the totals for the SRS sampling error model with \( \psi = 0.003 \). As for the proportions, MC relative biases of both the Taylor and bootstrap MSE estimators in most of the cells are less than two MC standard errors in absolute value. The differences between the MC relative biases of the bootstrap and Taylor MSE estimators
are smaller for the totals than for the proportions. The Taylor and bootstrap MSE estimators typically have positive MC biases in the small provinces and negative MC biases in the large provinces.

Figure 6.4 MC relative biases of MSE estimators for totals. B=Bootstrap, T=Taylor. SRS sampling error model, $\psi = 0.003$.

Table 6.28 shows the percentages of the MC MSE of $\hat{M}_{ik,B}$ explained by the different parts of the MSE and the corresponding estimators. The organization of Table 6.28 is analogous to the organization of Table 6.27. As for the proportions, the bootstrap and Taylor estimators of the leading terms have positive MC biases. The MC mean of the bootstrap estimator of the bias of the estimator of the leading term is closer to the MC bias than the MC mean of the Taylor estimator of the bias. As a result, the MC mean of the bootstrap MSE estimator is smaller than that of the Taylor MSE estimator in three digit code $i = 4$ of Prince Edward
Island. The effect of parameter estimation explains a larger fraction of the MSE for the totals than for the proportions. For example, in three digit code \( i = 4 \) in Prince Edward Island, 80% of the MSE of the predictor of the total is due to parameter estimation, while 39% of the MSE of the predictor of the proportion is due to the effect of estimating the unknown parameters. The fractions of the MSE due to parameter estimation are larger in the categories with larger means. For example, 95% of the MC MSE of the predictor of the total in category \( i = 1 \) of Prince Edward Island is attributable to the effect of estimating parameters. Both the bootstrap and Taylor estimators of the effect of parameter estimation have positive MC biases. The patterns in category \( i = 4 \) of Prince Island are typical of the other categories in the small provinces.

The bottom sub-table of Table 6.28 gives the decomposition of the empirical MSE and the MSE estimators for the total in three digit code \( i = 1 \) of Ontario. Both the Taylor and bootstrap estimators of the leading term have a positive bias. The MC bias of the Taylor estimator of the leading term is 2.54% of the empirical MSE of the predictor of the total, and the MC bias of the bootstrap estimator of the leading term is 2.44% of the empirical MSE of the predictor of the total. The bootstrap MSE estimator is smaller than the Taylor MSE estimator because the bootstrap estimator of the bias of the estimator of the leading term is larger than the Taylor estimator of the bias of the estimator of the leading term. The negative biases of the Taylor and bootstrap MSE estimators for the total are primarily due to the negative biases of the estimators of the effect of parameter estimation. The empirical effect of estimating the parameters explains 57% of the MSE. The MC means of the Taylor and bootstrap estimators of the effect of parameter estimation are 48% and 49% of the MC MSE of the predictor of the total in three digit code \( i = 1 \) of Ontario. We do not have an explanation for the negative bias.
Table 6.28: Percent of MC MSE of $\tilde{M}_{ik,B}$ explained by leading term (column 1), effect of parameter estimation (column 3), and the raking effect (column 4). Column 2 has ratios of empirical biases (row 1) and MC means of the Taylor (row 2) and bootstrap (row 3) estimators of biases of the estimator of the leading term to the MC MSE of $\tilde{M}_{ik,B}$ (multiplied by 100). SRS, $\psi = 0.003$

<table>
<thead>
<tr>
<th>Leading Term Bias of Taylor Est.</th>
<th>Param. Est.</th>
<th>Raking of Leading Term</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prince Edward Island, $i = 4$, $MSE(\tilde{M}_{ik,B}) = 1009$</td>
<td>Empirical</td>
<td>Taylor</td>
</tr>
<tr>
<td>18.70</td>
<td>3.78</td>
<td>80.28</td>
</tr>
<tr>
<td>22.48</td>
<td>1.11</td>
<td>84.40</td>
</tr>
<tr>
<td>22.56</td>
<td>3.98</td>
<td>84.77</td>
</tr>
<tr>
<td>Ontario, $i = 1$, $MSE(\tilde{M}_{ik,B}) = 26685493$</td>
<td>Empirical</td>
<td>Taylor</td>
</tr>
<tr>
<td>38.23</td>
<td>2.54</td>
<td>57.02</td>
</tr>
<tr>
<td>40.77</td>
<td>-5.48</td>
<td>47.81</td>
</tr>
<tr>
<td>40.67</td>
<td>1.48</td>
<td>49.40</td>
</tr>
</tbody>
</table>

6.5.2 SRS, $\psi = 0.02$

Figure 6.5 shows the Monte Carlo relative biases of the bootstrap and Taylor estimators of the MSE’s of the predictors of the proportions for the SRS simulation model with $\psi = 0.02$. The MC relative biases of the Taylor MSE estimators are all smaller than 10 in absolute value. Two of the bootstrap MSE estimates have negative MC relative biases that are smaller than 10. Five of the MC relative biases exceed two MC standard errors in absolute value for both the bootstrap and Taylor procedures. (The Taylor MSE estimators were computed for 5000 generated data sets. All of the MC relative biases based on the 5000 generated data sets are smaller than 5 in absolute value. The MC relative biases based on the 5000 generated data sets in three digit codes $i = 3$ and $i = 4$ in Ontario are -3.57 and -3.08, respectively.) While the MC
relative biases are small relative to the MC standard errors, the bootstrap and Taylor MSE estimators typically have negative MC biases. The MC means of the bootstrap MSE estimators are larger than the MC means of the Taylor MSE estimators in the large provinces, and the MC means of the bootstrap MSE estimators are typically smaller than the corresponding MC means of the Taylor MSE estimators in the smaller provinces.

Figure 6.5 MC relative biases of MSE estimators for proportions.

B=Bootstrap, T=Taylor. SRS sampling error model, $\psi = 0.02$.

Table 6.29 contains the components of the MSE’s and MSE estimators as fractions of the MC MSE of the predictor of the proportion in three digit code $i = 3$ of Prince Edward Island and three digit code $i = 2$ of Ontario. The leading term explains a larger fraction of the MSE when $\psi = 0.02$ than when $\psi = 0.003$. For example, the leading term accounts for 75% of the MSE of the MC MSE three digit code $i = 3$ of Prince Edward Island and 91%
of the MC MSE in three digit code $i = 2$ of Ontario. The negative MC biases of both the bootstrap and Taylor MSE estimators arise mainly because the estimators of the leading terms have negative biases. The MC means of the bootstrap MSE estimators are smaller than the MC means of the Taylor MSE estimators in the small provinces because the MC means of the bootstrap estimators of the leading terms are typically smaller than the MC means of the Taylor estimators of the leading terms in the small provinces. As explained above, we conjecture that the difference between the MC means of the bootstrap and Taylor estimators of the leading terms occurs because the bootstrap data generating procedure does not preserve the variances of the proportions.

In three digit code $i = 2$ of Ontario, the MC mean of the bootstrap MSE estimator is larger than the MC mean of the Taylor MSE estimator. The relative increase in the MSE due to raking is greater in the large provinces than in the small provinces. The raking effect explains why the MC means of the bootstrap MSE estimators are larger than the MC means of the Taylor MSE estimators in many of the categories in the large provinces (such as three digit code $i = 2$ of Ontario).
Table 6.29 Percent of MC MSE of $\hat{p}_{ik,B}$ explained by leading term (column 1), effect of parameter estimation (column 3), and the raking effect (column 4). Column 2 has ratios of empirical biases (row 1) and MC means of the Taylor (row 2) and bootstrap (row 3) estimators of biases of the estimator of the leading term to the MC MSE of $\hat{p}_{ik,B}$ (multiplied by 100). SRS, $\psi = 0.02$

Figure 6.6 shows the MC relative biases of the Taylor and bootstrap estimators of the MSE’s of the predictors of the totals under the SRS simulation model with $\psi = 0.02$. The MC relative biases of the Taylor MSE estimators are smaller than 10 in absolute value, and the MC relative biases of the bootstrap MSE estimators are smaller than 10 in all but two cells. The Taylor MSE estimators have negative MC relative biases that exceed two MC standard errors in absolute value in three of the cells. The bootstrap MSE estimators have negative MC relative biases that exceed two MC standard errors in absolute value in four of the cells. The bootstrap MC relative bias for three digit code $i = 1$ of New Foundland exceeds two MC standard errors.
Figure 6.6 MC relative biases of MSE estimators for totals. B=Bootstrap, T=Taylor. SRS sampling error model, $\psi = 0.02$.

The MC relative biases of the Taylor and bootstrap MSE estimators are more similar for the totals than for the proportions. The MC means of the bootstrap and Taylor estimators of the leading terms are not significantly different in any of the provinces. The differences between the bootstrap and Taylor MSE estimators in Figure 6.6 reflect differences between the bootstrap and Taylor estimators of the other three components of the MSE. The bootstrap estimator of the bias of the estimator of the leading term is larger than the Taylor estimator of the bias in all but five cells. The bootstrap estimator of the effect of parameter estimation is also greater than the Taylor estimator in all but two of the cells. The fraction of the MSE due to raking increases as the province sizes increase, which partly explains the negative MC bias of the Taylor MSE estimators in the larger provinces.
Table 6.30 illustrates the sources of the differences between the bootstrap and Taylor MSE estimators for three digit code $i = 3$ in Prince Edward Island and three digit code $i = 2$ in Ontario. The MC means of the bootstrap and Taylor estimators of the leading terms are nearly identical. The MC mean of the bootstrap estimator of the bias of $\hat{MSE}_{11,ik}$ is larger than the MC mean of the Taylor estimator of the bias. The difference between the MC MSE of the benchmarked predictor and the MC MSE of the initial predictor is 3% of the MC MSE of the benchmarked predictor, and the MC mean of the bootstrap estimator of the difference in MSE's is 2% of the MC MSE of the benchmarked predictor.

<table>
<thead>
<tr>
<th>Leading Term of Leading Term</th>
<th>Bias of Taylor Est.</th>
<th>Param. Est.</th>
<th>Raking</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prince Edward Island, $i = 3$, $MSE(\tilde{M}_{ik,B}) = 8046$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Empirical</td>
<td>29.30</td>
<td>1.76</td>
<td>71.03</td>
</tr>
<tr>
<td>Taylor</td>
<td>31.06</td>
<td>1.43</td>
<td>71.57</td>
</tr>
<tr>
<td>Bootstrap</td>
<td>31.06</td>
<td>3.57</td>
<td>72.93</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ontario, $i = 2$, $MSE(\tilde{M}_{ik,B}) = 14160423$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Empirical</td>
<td>85.39</td>
<td>-1.49</td>
<td>11.60</td>
</tr>
<tr>
<td>Taylor</td>
<td>83.90</td>
<td>-1.12</td>
<td>10.15</td>
</tr>
<tr>
<td>Bootstrap</td>
<td>83.86</td>
<td>1.26</td>
<td>11.06</td>
</tr>
</tbody>
</table>

Table 6.30 Percent of MC MSE of $\tilde{M}_{ik,B}$ explained by leading term (column 1), effect of parameter estimation (column 3), and the raking effect (column 4). Column 2 has ratios of empirical biases (row 1) and MC means of the Taylor (row 2) and bootstrap (row 3) estimators of biases of the estimator of the leading term to the MC MSE of $\tilde{M}_{ik,B}$ (multiplied by 100). SRS, $\psi = 0.02$. 
6.5.3 Two Stage, 0.003

Figure 6.7 shows the MC relative biases of the Taylor and bootstrap estimators of the MSE’s of the predictors of the proportions under the two stage simulation model with $\psi = 0.003$. The Taylor and bootstrap MSE estimators have positive MC biases in most of the cells. The MC bias of the Taylor MSE estimator exceeds two MC standard errors in ten of the cells, and the MC relative bias of the bootstrap MSE estimator exceeds two MC standard errors in five of the cells. The positive MC biases are partly due to the effect of the lower bound on the estimator of $\psi$. Under the two stage simulation model with $\psi = 0.003$, 29% of the estimates of $\psi$ are equal to the lower bound, and the MC mean of $\hat{\psi}$ is 0.0036.

![Two Stage, 0.003 Proportions](image)

Figure 6.7 MC relative biases of MSE estimators for proportions. B=Bootstrap, T=Taylor. Two stage sampling error model, $\psi = 0.003$. 

The components of the empirical MSE’s and the MSE estimates for the proportions in category \( i = 1 \) of New Foundland and category \( i = 4 \) of Quebec are given in Table 6.31. The lower bound on the estimator of \( \psi \) causes both the bootstrap and Taylor estimators of the leading term to have positive MC biases. Because the bootstrap estimator of the bias of the estimator of the leading term accounts for the lower bound effect, the MC mean of the bootstrap estimator of the bias is positive. In the two stage simulation, the bootstrap version of the direct estimator of the sampling covariance matrix has a Wishart distribution and is independent of the direct estimators of the proportions. In the simulation, however, the direct estimator of the sampling covariance matrix is a sample variance of vectors with multinomial-Dirichlet mixture distributions, and the elements of the direct estimator of the sampling covariance matrix are positively correlated with the elements of diag(\( \hat{p}_k \)) - \( \hat{p}_k \hat{p}_k' \). The bootstrap version of the sampling covariance matrix fails to reproduce both the variance of the direct estimator of the covariance matrix and the correlation between the direct estimator of the sampling covariance matrix and the direct estimators of the proportions. As a consequence, the bootstrap gives a poor approximation for the bias of the leading term. The magnitude of the MC mean of the Taylor estimator of the bias is approximately equal to the difference between the MC MSE of the benchmarked predictor and the MC MSE of the initial predictor. Without the raking operation, the positive MC bias of the Taylor MSE estimator would be larger.

The properties of the MSE estimators in the three digit code \( i = 2 \) of Quebec are typical of the properties of the MSE estimators in the large provinces. The lower bound on the estimator of \( \psi \) causes the Taylor estimator of the leading term and the Taylor estimator of \( E[(\hat{\gamma}_k - \gamma_k)^2(u_{ik} + e_{ik})^2] \) to have positive MC biases. The bootstrap estimator of the leading term also has a positive MC bias, but the MC bias of the bootstrap estimator of the effect of parameter estimation is negligible. Both the bootstrap and the Taylor estimators of the biases of the estimators of the leading terms are poor approximations for the empirical biases. The MC mean of the bootstrap estimator of the leading term is smaller than the MC mean of the Taylor estimator of the leading term, which may arise because the bootstrap data generating
procedure does not produce an unbiased estimator of the variance of the proportions.

<table>
<thead>
<tr>
<th>Leading Term</th>
<th>Bias of Taylor Est.</th>
<th>Param. Est.</th>
<th>Raking of Leading Term</th>
</tr>
</thead>
<tbody>
<tr>
<td>New Foundland, ( i = 1 ), ( MSE(\hat{p}_{ik,B}) = 0.14% )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Empirical</td>
<td>46.31</td>
<td>10.88</td>
<td>46.87</td>
</tr>
<tr>
<td>Taylor</td>
<td>57.19</td>
<td>-2.48</td>
<td>48.48</td>
</tr>
<tr>
<td>Bootstrap</td>
<td>56.70</td>
<td>2.83</td>
<td>47.52</td>
</tr>
<tr>
<td>Quebec, ( i = 2 ), ( MSE(\hat{p}_{ik,B}) = 0.067% )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Empirical</td>
<td>48.81</td>
<td>4.46</td>
<td>43.07</td>
</tr>
<tr>
<td>Taylor</td>
<td>53.26</td>
<td>-8.32</td>
<td>50.21</td>
</tr>
<tr>
<td>Bootstrap</td>
<td>52.97</td>
<td>0.94</td>
<td>44.89</td>
</tr>
</tbody>
</table>

Table 6.31 Percent of MC MSE of \( \hat{p}_{ik,B} \) explained by leading term (column 1), effect of parameter estimation (column 3), and the raking effect (column 4). Column 2 has ratios of empirical biases (row 1) and MC means of the Taylor (row 2) and bootstrap (row 3) estimators of biases of the estimator of the leading term to the MC MSE of \( \hat{p}_{ik,B} \) (multiplied by 100). Two stage, \( \psi = 0.003 \)

Figure 6.8 shows the MC relative biases of the bootstrap and Taylor estimators of the MSE’s of the predictors of the totals for the two stage simulation model with \( \psi = 0.003 \). The bootstrap MSE estimators have positive MC relative biases in the smallest two provinces that exceed two MC standard errors. The Taylor and bootstrap MC relative biases are smaller in magnitude than two MC standard errors in most of the other cells. The MC relative biases of the bootstrap MSE estimators are larger than the corresponding MC relative biases of the Taylor MSE estimators.
Two Stage, 0.003
Totals

Figure 6.8  MC relative biases of MSE estimators for totals. B=Bootstrap,
T=Taylor. Two stage sampling error model, $\psi = 0.003$.

Table 6.32 shows the components of the MSE’s and the MSE estimators for three digit code $i = 1$ in New Foundland and three digit code $i = 2$ in Quebec. A reason that both the bootstrap and Taylor estimates of the MSE’s have positive biases in the small provinces is that the bootstrap and Taylor estimators of the leading terms have positive MC biases. The positive MC biases result primarily from overestimation of the sampling variances of the totals at small sample sizes. In the two stage simulation model, the direct estimator of the sampling covariance matrix of the errors in the proportions is an unbiased estimator, while the direct estimator of the sampling covariance matrix of the totals is obtained though a Taylor linearization. The linear approximation improves as the sample sizes increase, so the relative bias of the direct estimator of the sampling covariance matrix of the totals decreases as
the sample size increases. The positive MC bias of the bootstrap MSE estimator exceeds the positive MC bias of the Taylor MSE estimator primarily because the MC mean of the bootstrap estimator of the effect of estimating parameters exceeds the MC mean of the Taylor estimator of the effect of estimating the parameters. In three digit code $i = 2$ of New Foundland, the MC mean of the bootstrap estimator of the raking effect exceeds the empirical raking effect. The MC means of the bootstrap estimators of the raking effects are not uniformly larger than the empirical raking effects for the other categories in the two smallest provinces. The raking effect is larger in the large provinces than in the small provinces, and the main reason that the MC means of the bootstrap MSE estimators exceed the MC means of the Taylor MSE estimators in the large provinces is that the raking operation increases the MC mean of the bootstrap MSE estimator but leaves the Taylor MSE estimator unchanged. In three digit code $i = 2$ of Quebec, the MC mean of the Taylor estimator of the bias of the estimator of the leading term has roughly the same magnitude as the empirical raking effect.
Leading Term Bias of Taylor Est. Param. Est. Raking of Leading Term

<table>
<thead>
<tr>
<th></th>
<th>New Foundland, $i = 1$, $MSE(\tilde{M}_{ik,B}) = 44167$</th>
<th>Quebec, $i = 2$, $MSE(\tilde{M}_{ik,B}) = 4804653$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Empirical</td>
<td>7.25, 2.33, 92.57, 0.18</td>
<td>37.24, 3.78, 57.69, 5.08</td>
</tr>
<tr>
<td>Taylor</td>
<td>9.58, 0.25, 98.90, 0.00</td>
<td>41.02, -5.66, 55.28, 0.00</td>
</tr>
<tr>
<td>Bootstrap</td>
<td>9.59, 1.13, 101.71, 2.14</td>
<td>41.13, 1.20, 58.22, 7.60</td>
</tr>
</tbody>
</table>

Table 6.32 Percent of MC MSE of $\tilde{M}_{ik,B}$ explained by leading term (column 1), effect of parameter estimation (column 3), and the raking effect (column 4). Column 2 has ratios of empirical biases (row 1) and MC means of the Taylor (row 2) and bootstrap (row 3) estimators of biases of the estimator of the leading term to the MC MSE of $\tilde{M}_{ik,B}$ (multiplied by 100). Two stage, $\psi = 0.003$

6.5.4 Two Stage, 0.02

Figure 6.9 shows the MC relative biases of the bootstrap and Taylor estimators of the MSE’s of the predictors of the proportions under the two stage simulation model with $\psi = 0.02$. The MC relative biases of the bootstrap and Taylor MSE estimators are smaller than two MC standard errors in absolute value in most of the cells. The MC means of the bootstrap MSE estimators are typically larger than the corresponding MC means of the Taylor MSE estimators primarily because of the raking effect.
Table 6.33 contains the components of the empirical MSE’s and the components of the MC means of the MSE estimators for the three digit code $i = 1$ in New Foundland and three digit code $i = 2$ in Quebec. The MC means of the bootstrap estimators of the leading terms are smaller than the MC means of the Taylor estimates, which may occur because the bootstrap data generating procedure distorts the covariance matrix of the proportions. The MC means of the bootstrap estimators of the effects of estimating parameters exceed the MC means of the Taylor estimators. The main reason that the bootstrap MSE estimators exceed the Taylor MSE estimators is the raking effect. Also, the MC mean of the bootstrap estimator of the bias of the estimator of $\hat{MSE}_{11,ik}$ exceeds the MC mean of the Taylor estimator of the
bias in magnitude.

<table>
<thead>
<tr>
<th>Leading Term</th>
<th>Bias of Taylor Est.</th>
<th>Param. Est.</th>
<th>Raking of Leading Term</th>
</tr>
</thead>
<tbody>
<tr>
<td>New Foundland, $i = 1$, $MSE(\hat{p}_{ik,B}) = 0.45%$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Empirical</td>
<td>79.79</td>
<td>-2.81</td>
<td>17.51</td>
</tr>
<tr>
<td>Taylor</td>
<td>76.99</td>
<td>-0.70</td>
<td>17.35</td>
</tr>
<tr>
<td>Bootstrap</td>
<td>74.40</td>
<td>-2.94</td>
<td>17.81</td>
</tr>
<tr>
<td>Quebec, $i = 2$, $MSE(\hat{p}_{ik,B}) = 0.12%$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Empirical</td>
<td>83.74</td>
<td>1.07</td>
<td>13.96</td>
</tr>
<tr>
<td>Taylor</td>
<td>84.81</td>
<td>-2.12</td>
<td>14.13</td>
</tr>
<tr>
<td>Bootstrap</td>
<td>83.95</td>
<td>-2.52</td>
<td>14.68</td>
</tr>
</tbody>
</table>

Table 6.33 Percent of MC MSE of $\hat{p}_{ik,B}$ explained by leading term (column 1), effect of parameter estimation (column 3), and the raking effect (column 4). Column 2 has ratios of empirical biases (row 1) and MC means of the Taylor (row 2) and bootstrap (row 3) estimators of biases of the estimator of the leading term to the MC MSE of $\hat{p}_{ik,B}$ (multiplied by 100). Two stage, $\psi = 0.02$

Figure 6.10 shows the MC relative biases of the bootstrap and Taylor estimators of the MSE’s of the predictors of the totals under the two stage simulation model with $\psi = 0.02$. In the smallest two provinces, the positive MC bias of the bootstrap MSE estimator exceeds two MC standard errors. The Taylor MSE estimators also have positive MC biases in the smallest two provinces. As discussed in the context of the two stage simulation model with $\psi = 0.003$, the direct estimator of the sampling covariance matrix for the proportions is unbiased in the two stage simulation model, and the linear approximation used for the direct estimator of the sampling covariance matrix of the totals improves as the sample sizes increase. The MC means of the bootstrap MSE estimators are uniformly larger than the MC means of the Taylor MSE
estimators. The magnitudes of the negative MC biases of the Taylor MSE estimators exceed two MC standard errors in eight of the cells.

![Two Stage, 0.02 Totals](image)

Figure 6.10 MC relative biases of MSE estimators for totals. B=Bootstrap, T=Taylor. Two stage sampling error model, $\psi = 0.02$.

Table 6.34 shows the components of the empirical MSE’s and the MC means of the estimators of the MSE’s for the three digit code $i = 1$ in New Foundland and three digit code $i = 2$ in Quebec. Both the bootstrap and Taylor estimators of the leading terms for three digit code $i = 1$ in New Foundland have positive MC biases. The positive MC biases of the estimators of the leading terms occur because the direct estimator of the variance of $\hat{M}_{ik} - M_{ik}$ has a positive MC bias in the two stage simulation model. In both of the provinces, the main reason that the MC mean of the bootstrap MSE estimator exceeds the MC mean of the Taylor MSE estimator is that the MC mean of the bootstrap MSE estimator of the effect
of estimating parameters exceeds the MC mean of the Taylor MSE estimator of the effect of estimating parameters. The raking effect also causes the MC means of the bootstrap MSE estimators to exceed the MC means of the Taylor MSE estimators.

<table>
<thead>
<tr>
<th>Leading Term of Leading Term</th>
<th>Bias of Taylor Est.</th>
<th>Param. Est.</th>
<th>Raking</th>
</tr>
</thead>
<tbody>
<tr>
<td>New Foundland, $i = 1$, $MSE(\tilde{M}_{1k,B}) = 60630$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Empirical</td>
<td>31.66</td>
<td>2.03</td>
<td>67.16</td>
</tr>
<tr>
<td>Taylor</td>
<td>33.69</td>
<td>1.26</td>
<td>69.79</td>
</tr>
<tr>
<td>Bootstrap</td>
<td>33.64</td>
<td>0.19</td>
<td>72.46</td>
</tr>
<tr>
<td>Quebec, $i = 2$, $MSE(\tilde{M}_{2k,B}) = 7895580$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Empirical</td>
<td>73.81</td>
<td>1.83</td>
<td>24.01</td>
</tr>
<tr>
<td>Taylor</td>
<td>75.64</td>
<td>-1.63</td>
<td>19.82</td>
</tr>
<tr>
<td>Bootstrap</td>
<td>75.43</td>
<td>-2.00</td>
<td>22.42</td>
</tr>
</tbody>
</table>

Table 6.34 Percent of MC MSE of $\tilde{M}_{ik,B}$ explained by leading term (column 1), effect of parameter estimation (column 3), and the raking effect (column 4). Column 2 has ratios of empirical biases (row 1) and MC means of the Taylor (row 2) and bootstrap (row 3) estimators of biases of the estimator of the leading term to the MC MSE of $\tilde{M}_{ik,B}$ (multiplied by 100). Two stage, $\psi = 0.02$

6.5.5 Estimators of Leading Terms

In Figures 6.11 and 6.12, the empirical relative biases of the Taylor and bootstrap estimators of the leading terms in the MSE’s of the predictors of the proportions are plotted along with the Taylor and bootstrap estimates of the biases. Figure 6.11 is for the SRS simulation model, and Figure 6.12 is for the two stage simulation model. For both figures, $\psi = 0.003$. The points labeled “B” are MC relative biases of the bootstrap estimates of the
leading terms and are defined,

\[ \frac{100}{E_{MC}[\hat{MSE}^{bs}_{11,ik}]} - \frac{L_{ik}}{L_{ik}}. \]

The points labeled “T” are the corresponding MC relative biases of the Taylor MSE estimators. The points labeled “b” are the ratios of the MC means of the bootstrap estimators of the biases of \( \hat{MSE}_{11,ik} \) to the empirical leading terms. The points labeled “t” are the ratios of the MC means of the Taylor estimators of the biases of \( \hat{MSE}_{11,ik} \) to the empirical leading terms.

**Figure 6.11** Ratios of MC biases of bootstrap (B) and Taylor (T) estimators of leading terms to empirical leading terms, and ratios of MC means of bootstrap (b) and Taylor (t) estimators of biases of estimators of leading terms to empirical leading terms (multiplied by 100). SRS, \( \psi = 0.003 \), proportions.
Figure 6.12 Ratios of MC biases of bootstrap (B) and Taylor (T) estimators of leading terms to empirical leading terms, and ratios of MC means of bootstrap (b) and Taylor (t) estimators of biases of estimators of leading terms to empirical leading terms (multiplied by 100). Two stage, $\psi = 0.003$, proportions.

The lower bound on the estimator of $\psi$ leads to a positive MC bias for both the two stage and SRS simulation models. The MC relative biases approach zero as the province sample sizes increase. Because the bootstrap accounts for the lower bound effect, the bootstrap estimators of the bias are larger than the Taylor estimators of the biases for both simulation models. For the SRS simulation model, the ratios of the MC means of the bootstrap and Taylor estimates of the biases to the empirical leading terms also approach zero as the sample sizes increase. In the SRS simulation, the model used to generate a bootstrap version of the direct estimator of
the sampling covariance matrix corresponds to the underlying data generating procedure. In contrast, the properties of the bootstrap version of the direct estimator of the covariance matrix used in the two stage simulation do not represent the underlying data generating method. As a consequence, the quality of the bootstrap estimator of the bias of the estimator of the leading term is worse in the two stage simulation than in the SRS simulation.

The structure of Figures 6.13 and 6.14 are analogous to the structure of Figure 6.11. The output in Figures 6.13 and 6.14 is from the two stage simulation with $\psi = 0.02$. Figure 6.13 is for proportions, and Figure 6.14 is for totals. For the proportions, the MC means of the bootstrap estimators of the leading terms are smaller than the MC means of the Taylor MSE estimators, especially in the small provinces. In the two-stage simulations, the MC means of the bootstrap estimators of the leading terms for the proportions are also smaller than the MC means of the corresponding Taylor estimators. The differences between the MC means of the bootstrap and Taylor estimators of the leading terms are greater for $\psi = 0.02$ than for $\psi = 0.003$. The MC relative biases of the bootstrap and Taylor estimators of the leading terms in the MSE’s are more similar for the totals than for the proportions. The bootstrap data generating procedure preserves the covariance matrix of the totals but distorts the covariance matrix for the proportions. We conjecture that the distortion of the covariance matrix for the proportions explains the differences between the MC means of the bootstrap and Taylor estimators of the leading terms for the proportions.
Figure 6.13  Ratios of MC biases of bootstrap (B) and Taylor (T) estimators of leading terms to empirical leading terms, and ratios of MC means of bootstrap (b) and Taylor (t) estimators of biases of estimators of leading terms to empirical leading terms (multiplied by 100). Two stage, $\psi = 0.02$, proportions.
Figure 6.14 Ratios of MC biases of bootstrap (B) and Taylor (T) estimators of leading terms to empirical leading terms, and ratios of MC means of bootstrap (b) and Taylor (t) estimators of biases of estimators of leading terms to empirical leading terms (multiplied by 100). Two stage, $\psi = 0.02$, totals.

6.5.6 Effect of Estimating Parameters

The MC relative biases of the bootstrap and Taylor estimators of the effect of estimating parameters in the scale of totals are plotted in Figure 6.15 for the two stage simulation with $\psi = 0.003$. The MC relative bias of the bootstrap estimator of the effect of parameter estimation is

$$
100 \frac{E_{MC}[\hat{P}_{arest_{ik},bs}] - \hat{P}_{arest_{ik}}}{\hat{P}_{arest_{ik}}}
$$
and the MC relative bias of the Taylor MSE estimator of the effect of parameter estimation is obtained by replacing $\hat{P}_{\text{aret}_{ik,bs}}$ with $P_{\text{aret}_{ik}}$. The MC relative biases of the bootstrap estimators are uniformly larger than the MC relative biases of the Taylor estimators. When $\psi = 0.003$, the effect of estimating parameters is the largest fraction of the empirical MSE’s of the predictors of the totals. The differences between the bootstrap and Taylor estimators of the effect of parameter estimation largely explains why the bootstrap MC relative biases are larger than the Taylor MC relative biases for the totals when $\psi = 0.003$. We do not understand why MC means of the bootstrap estimators of the effect of parameter estimation exceed the corresponding MC means of the Taylor MSE estimators.

**Two Stage, 0.003**

![Figure 6.15](image_url)

Figure 6.15  MC relative biases of bootstrap (B) and Taylor (T) estimators of the contribution of estimation of the unknown parameters to the MSE. Two stage, $\psi = 0.003$, totals.
6.5.7 Raking Effect

Figure 6.16 shows the empirical raking effects along with the MC means of the bootstrap estimators of the raking effects as fractions of the MC MSE’s of the predictors of the proportions for the two stage simulation model with $\psi = 0.003$. The points labeled “E” are the ratios of the empirical raking effects to the MC MSE’s of the predictors of the proportions. The points labeled “B” are the ratios of the MC means of the bootstrap estimators of the differences between the MSE’s of the benchmarked predictors and the MC means of the initial predictors of the proportions to the MC MSE’s of the benchmarked predictors of the proportions (multiplied by 100):

$$100\frac{E_{MC}[\hat{Rake}_{ik}]}{MSE_{MC}(\hat{p}_{ik,B})}.$$ 

An alternative estimator of the raking effect, suggested by Dick (1995), is to multiply an estimator of the MSE of an initial (non-benchmarked) predictor by the squared ratio of the benchmarked predictor to the initial predictor. Each point labeled “D” in Figure 6.16 is

$$\frac{E_{MC}[\hat{MSE}_{ik,2,B}\hat{p}_{ik,B}^2 - \hat{p}_{pred,ik,B}^2]}{MSE_{MC}(\hat{p}_{ik,B})}.$$ 

The bootstrap estimates of the raking effects (labeled “B”) are positively correlated with the empirical raking effects (points labeled “E”). The average of the differences,

$$100\frac{E_{MC}[\hat{Rake}_{ik} - Rake_{ik}]}{MSE_{MC}(\hat{p}_{ik,B})},$$

across the forty cells is 0.94, and the maximum of the differences is 4.10. (The median of the MC standard errors for the differences is 0.92). The estimator suggested by Dick performs poorly in our simulations. The points labeled “D” oscillate in Figure 6.16 because the MC means of $\{\hat{p}_{ik,B}\hat{p}_{pred,ik,B}^{-1} : i = 1, \ldots, 4; k = 1, \ldots, 10\}$ are larger than one in categories $i = 2$ and $i = 4$ and are smaller than one in the categories $i = 1$ and $i = 3$. 
Figure 6.16  Ratios of MC means of bootstrap estimates of the raking effects to the MC MSE’s of the predictors of the proportions (B). Ratios of empirical raking effects to MC MSE’s of the predictors of the proportions (E). Two stage, $\psi = 0.003$.

In the simulations discussed in the previous sections, the MC MSE’s of the benchmarked predictors seldom exceed the MC MSE’s of the initial predictors by more than 10%. A separate simulation study was conducted to evaluate the properties of the bootstrap MSE estimator in a situation where the raking operation has a greater effect on the MC MSE’s of the predictors. The parameters used in the separate simulation are the same as the parameters used in the SRS simulation model except for the province sample sizes, where the province sample sizes used are given in Table 6.35. In the simulations discussed above, the province sample sizes tend to increase as the province margins increase. In Table 6.35, the largest sample sizes are
assigned to provinces with intermediate two digit totals. Two simplifications to the procedure are implemented to reduce the computing time for the separate simulation. One, the estimator of $p_{T,i,k}$ is the initial estimator based on the multinomial log likelihood. Two, instead of combining estimators of $\psi$ and $c_k$ across four two-way tables, a single four by ten table is used to estimate the variance parameters.

<table>
<thead>
<tr>
<th>PE</th>
<th>NF</th>
<th>NB</th>
<th>NS</th>
<th>SK</th>
<th>MB</th>
<th>BC</th>
<th>AB</th>
<th>QC</th>
<th>ON</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_k$</td>
<td>770</td>
<td>2243</td>
<td>4538</td>
<td>6091</td>
<td>5622</td>
<td>9220</td>
<td>35052</td>
<td>31323</td>
<td>72926</td>
</tr>
<tr>
<td>$n_k$</td>
<td>37</td>
<td>29</td>
<td>57</td>
<td>53</td>
<td>114</td>
<td>246</td>
<td>92</td>
<td>83</td>
<td>66</td>
</tr>
</tbody>
</table>

Table 6.35 Expected province two digit totals and province sample sizes

Table 6.36 shows the ratios of the MC MSE’s of the benchmarked predictors to the MC MSE’s of the initial predictors in the simulation with the sample sizes in Table 6.35. The MSE’s of the benchmarked predictors exceed the MSE’s of the initial predictors by more than 10% in all of the cells.

<table>
<thead>
<tr>
<th>PE</th>
<th>NF</th>
<th>NB</th>
<th>NS</th>
<th>SK</th>
<th>MB</th>
<th>BC</th>
<th>AB</th>
<th>QC</th>
<th>ON</th>
</tr>
</thead>
<tbody>
<tr>
<td>$i = 1$</td>
<td>1.150</td>
<td>1.172</td>
<td>1.157</td>
<td>1.169</td>
<td>1.214</td>
<td>1.315</td>
<td>1.191</td>
<td>1.180</td>
<td>1.166</td>
</tr>
<tr>
<td>$i = 2$</td>
<td>1.173</td>
<td>1.133</td>
<td>1.165</td>
<td>1.120</td>
<td>1.198</td>
<td>1.185</td>
<td>1.155</td>
<td>1.143</td>
<td>1.174</td>
</tr>
<tr>
<td>$i = 3$</td>
<td>1.140</td>
<td>1.146</td>
<td>1.161</td>
<td>1.173</td>
<td>1.182</td>
<td>1.302</td>
<td>1.176</td>
<td>1.182</td>
<td>1.158</td>
</tr>
<tr>
<td>$i = 4$</td>
<td>1.161</td>
<td>1.207</td>
<td>1.185</td>
<td>1.244</td>
<td>1.207</td>
<td>1.324</td>
<td>1.232</td>
<td>1.204</td>
<td>1.105</td>
</tr>
</tbody>
</table>

Table 6.36 Ratios of MC MSE’s of benchmarked predictors of proportions to MC MSE’s of initial predictors of proportions in the simulation with sample sizes given in Table 6.35

Figure 6.17 shows the relationship between the empirical raking effects and the MC means of the bootstrap estimators of the raking effects as fractions of the total MSE’s. The
empirical raking effects, on the horizontal axis, are defined,

$$\frac{MSE_{MC}(\hat{p}_{ik,B}) - MSE_{MC}(\hat{p}_{ik,B})}{MSE_{MC}(\hat{p}_{ik,B})}.$$ 

The ratios of the differences between the MC means of the bootstrap estimators of the MSE's of the benchmarked predictors and the MC means of the bootstrap estimators of the MSE's of the initial predictors to the MC MSE's of the benchmarked predictors are plotted on the vertical axis. The positive correlation in Figure 6.17 suggests that the bootstrap procedure is moderately successful in capturing the additional variability due to the raking operation.
6.5.8 Summary

Many of the properties of the Taylor and bootstrap estimators of the MSE’s reflect properties of the estimators of the leading terms in the MSE’s. For the two-stage simulation model, the MC means of the bootstrap estimators of the leading terms for the proportions are smaller than the corresponding MC means of the Taylor MSE estimators. The relative differences between the bootstrap and Taylor estimators of the leading terms are greater for
the two-stage simulation model than for the SRS simulation model, and the relative differences are greater when $\psi = 0.02$ than when $\psi = 0.003$. We suspect that the differences between the MC means of the bootstrap and Taylor estimators of the leading terms for the proportions arise because the bootstrap data generating procedure does not produce correct estimators of the variances of the proportions. (See Appendix 4.) The relative differences between the MC means of the bootstrap and Taylor estimators of the leading terms for the totals are smaller than for the proportions. The bootstrap data generating procedure preserves the direct estimators of the variances of the totals.

The effect of the lower bound on the estimator of $\psi$ explains several of the differences between the bootstrap and Taylor MSE estimators when $\psi = 0.003$ for both the SRS and two-stage simulation models. When $\psi = 0.003$, the lower bound causes both the bootstrap and Taylor estimators of the leading terms to have positive MC biases. The Taylor estimator of the bias of the estimator of the leading term does not include a term for the positive bias in the estimator of $\psi$ that results from using the lower bound. As a consequence, the Taylor estimator of the bias is too small for sufficiently small values of $\psi$. Because the bootstrap estimator of the bias of the estimator of the leading term accounts for the lower bound effect, the bootstrap estimator of the bias is larger than the Taylor estimator of the bias when $\psi = 0.003$. The relatively large MC mean of the bootstrap bias estimator partly explains why the MC means of the bootstrap MSE estimators are smaller than the MC means of the Taylor MSE estimators in the SRS simulation with $\psi = 0.003$.

The increase in the MSE due to raking is greater for $\psi = 0.003$ than for $\psi = 0.02$. The raking effect increases the bootstrap MSE estimator relative to the Taylor MSE estimator. However, the MC means of the bootstrap MSE estimators are usually smaller than the MC means of the Taylor MSE estimators at $\psi = 0.003$ because of the estimators of the biases of the estimators of the leading terms. As discussed above, the MC mean of the bootstrap estimator of the bias of the estimator of the leading term is larger than the MC mean of the Taylor estimator of the bias. Because the MC mean of the Taylor estimator of the bias is negative, the bias correction in the Taylor MSE estimator accounts for part of the raking effect.
In the simulation models considered, the MC MSE’s of the benchmarked predictors exceed the MC MSE’s of the initial predictors. The relative differences between the MSE’s of the benchmarked predictors and the MC MSE’s of the initial predictors increase as $\psi$ decreases and the province sizes increase. The negative biases of the Taylor MSE estimators in the large provinces are partly explained by the effect of the raking operation. The MC means of the bootstrap estimators of the raking effect relative to the MC MSE of the benchmarked predictor are positively correlated with the corresponding empirical quantities. This correlation provides limited justification for using the proposed bootstrap procedure to estimate the contribution of benchmarking to the MSE of the predictor.

The bootstrap and Taylor estimators of the bias of the estimators of the leading terms are poor approximations for the empirical biases. The assumptions underlying the bootstrap estimator of the bias does not hold in the two-stage simulation. The assumptions underlying the Taylor estimators of the bias do not hold for either simulation model. Improving the bias estimators is an area for future study.

The properties of the bootstrap and Taylor MSE estimators with respect to the variance due to parameter estimation differ for totals and proportions. In the two stage simulation model, the MC means of the bootstrap estimators of the contribution of parameter estimation to the MSE’s of the predictors are larger than the corresponding MC means of the Taylor estimators. We do not have an explanation for this difference between the bootstrap and Taylor MSE estimators. The bootstrap sample size used in the simulation is $B = 200$. A larger bootstrap sample size may improve the bootstrap MSE estimator or reduced some of the differences between the Taylor and bootstrap procedures.

The properties of the empirical relative biases of the bootstrap and Taylor MSE estimators depend on the value of $\psi$ (0.003 or 0.02), the sampling error model (SRS or two stage) and the parameter being estimated (proportion or totals). The analysis of the components of the MSE reveals flaws in both the Taylor and bootstrap MSE estimators. Nonetheless, the absolute values of the MC relative biases of the MSE estimators are typically smaller than two MC standard errors. Both MSE estimators provide reasonable approximations for the MC
MSE’s.

6.6 Empirical Properties of Normal Theory Prediction Intervals

The empirical properties of normal theory nominal 95% confidence intervals are investigated in this section. Normal theory confidence intervals for the predictors of the proportions based on the Taylor MSE estimator are defined by $[L_{T,ik}, U_{T,ik}]$, where

$$L_{T,ik} = \max\{0, \hat{p}_{ik,B} - t_{df}(0.975)\sqrt{\hat{MSE}_{2,ik,B}}\},$$

$$U_{T,ik} = \min\{1, \hat{p}_{ik,B} + t_{df}(0.975)\sqrt{\hat{MSE}_{2,ik,B}}\},$$

t

$t_{df}(0.975)$ is the 97.5 percentile of a $t$ distribution with $df = 104 = 4((m - 1)(K - 1) - 1)$ degrees of freedom. Normal theory confidence intervals based on the bootstrap MSE estimator are defined by substituting $\hat{MSE}_{2,ik,B}$ with $\hat{MSE}_{bs,ik}$. Normal theory confidence intervals for the totals are obtained by using $\hat{M}_{ik,B}$ instead of $\hat{p}_{ik,B}$ and by replacing the MSE estimators for the proportions with MSE estimators for the totals.

6.6.1 Empirical Coverages of Normal Theory 95% Confidence Intervals

Tables 6.37 - 6.40 contain the averages of the empirical coverages of the Taylor and bootstrap prediction intervals for the proportions and totals under the four simulation models.

The top half of each table is for proportions, and the bottom half is for totals. The intervals are based on a MC sample size of 2000, so an approximate MC standard error for a coverage for a single cell $(i, k)$ is 0.5%. The properties of the empirical coverages are summarized for the four simulation models separately.

SRS, 0.003

The average empirical coverages for the proportions (top half of Table 6.37) based on the Taylor MSE estimator are between 93% and 94% in the smallest three provinces and are between 94% and 96% in the larger provinces. The average empirical coverages for the proportions based on the bootstrap MSE estimators are smaller than 95% in all of the provinces. That the average empirical coverages for the proportions based on the bootstrap are smaller
than the average empirical coverages based on the Taylor MSE estimator is consistent with the observation from the previous section that the MC means of the bootstrap MSE estimators are usually smaller than the MC means of the Taylor MSE estimators for the proportions in the SRS simulation model with $\psi = 0.003$.

As for proportions, the average empirical coverages of the Taylor intervals for the totals (bottom half of Table 6.37) are between 93% and 94% in the smallest three provinces and tend to increase as the province sample sizes increase. The averages of the empirical coverages of the intervals for the totals based on the bootstrap MSE estimators are smaller than 95% in all of the provinces. The increasing trend in the coverages is smaller for the bootstrap intervals than for the Taylor intervals.

<table>
<thead>
<tr>
<th></th>
<th>PE</th>
<th>NF</th>
<th>NB</th>
<th>NS</th>
<th>SK</th>
<th>MB</th>
<th>BC</th>
<th>AB</th>
<th>QC</th>
<th>ON</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Proportions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>93.4</td>
<td>93.8</td>
<td>93.8</td>
<td>94.3</td>
<td>94.3</td>
<td>94.5</td>
<td>94.5</td>
<td>94.8</td>
<td>95.1</td>
<td>95.6</td>
</tr>
<tr>
<td>B</td>
<td>92.1</td>
<td>92.6</td>
<td>92.0</td>
<td>93.0</td>
<td>92.4</td>
<td>92.8</td>
<td>92.8</td>
<td>93.1</td>
<td>93.5</td>
<td>94.0</td>
</tr>
<tr>
<td></td>
<td>Totals</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>93.9</td>
<td>93.2</td>
<td>93.7</td>
<td>94.5</td>
<td>94.8</td>
<td>94.1</td>
<td>94.4</td>
<td>94.5</td>
<td>95.0</td>
<td>94.7</td>
</tr>
<tr>
<td>B</td>
<td>93.8</td>
<td>93.1</td>
<td>93.2</td>
<td>94.3</td>
<td>94.3</td>
<td>93.7</td>
<td>93.6</td>
<td>94.0</td>
<td>94.4</td>
<td>93.9</td>
</tr>
</tbody>
</table>

Table 6.37  Averages of empirical coverages of nominal 95% confidence intervals for proportions and totals. T=Taylor, B=bootstrap. SRS, $\psi = 0.003$.

**SRS, 0.02**

For the SRS simulation model with $\psi = 0.02$, the averages of the empirical coverages for the Taylor intervals are between 94.4% and 95.1% in all of the provinces and do not exhibit as much of an increasing trend as they do when $\psi = 0.003$. The averages of the empirical coverages of the bootstrap intervals are smaller than 94.3% in all of the provinces.

The averages of the empirical coverages of the Taylor intervals for the totals are closer
to 95% than the averages of the empirical coverages of the bootstrap intervals. The averages of the empirical coverages of the Taylor intervals for the totals tend to increase as the province sample sizes increase. The averages of the empirical coverages of the bootstrap intervals for the totals are between 92.4% and 93% in all of the provinces. The averages of the coverages of the bootstrap intervals for the totals do not increase as the province sizes increase.

<table>
<thead>
<tr>
<th></th>
<th>PE</th>
<th>NF</th>
<th>NB</th>
<th>NS</th>
<th>SK</th>
<th>MB</th>
<th>BC</th>
<th>AB</th>
<th>QC</th>
<th>ON</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>94.6</td>
<td>94.4</td>
<td>95.1</td>
<td>94.4</td>
<td>94.8</td>
<td>94.9</td>
<td>94.8</td>
<td>94.5</td>
<td>94.4</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>94.0</td>
<td>93.6</td>
<td>94.2</td>
<td>93.8</td>
<td>93.5</td>
<td>93.7</td>
<td>93.6</td>
<td>93.7</td>
<td>93.6</td>
<td></td>
</tr>
</tbody>
</table>

Table 6.38  Averages of empirical coverages of nominal 95% confidence intervals for proportions and totals. T=Taylor, B=bootstrap.

SRS, $\psi = 0.02$.

**Two Stage, 0.003**

Table 6.39 contains the average coverages under the two stage simulation model with $\psi = 0.003$. The properties of the average coverages of the Taylor intervals for the two stage simulation model with $\psi = 0.003$ are similar to the properties of the Taylor intervals for the SRS simulation with $\psi = 0.003$. In both cases, the average coverages of the Taylor intervals are smaller than 95% in the small provinces and tend to increase as the province sizes increase. The average coverages of the bootstrap intervals do not exceed 93.4% in the smallest five provinces and are between 93.4% and 94.3% in the largest five provinces. The averages of the empirical coverages for the totals based on both methods are smaller than 95%. The properties of the average empirical coverages of the intervals for the totals based on the bootstrap and Taylor MSE estimators are similar.
Two Stage, $\psi = 0.003$.

The averages of the empirical coverages of the bootstrap intervals for the proportions are larger than the corresponding averages of the empirical coverages of the Taylor intervals in all of the provinces. The averages of the empirical coverages of the intervals for the proportions based on the bootstrap MSE estimators are at least 94% in all of the provinces. The averages of the empirical coverages of the Taylor intervals for the proportions are at least 93.9% in all provinces. The averages of the empirical coverages arising from both the Taylor and bootstrap MSE estimators tend to increase as the province sample sizes increase. The averages of the empirical coverages of both the Taylor and bootstrap intervals for the totals are smaller than 94.6% in all of the provinces. The empirical coverages are closest to 95% in Ontario.
<table>
<thead>
<tr>
<th></th>
<th>PE</th>
<th>NF</th>
<th>NB</th>
<th>NS</th>
<th>SK</th>
<th>MB</th>
<th>BC</th>
<th>AB</th>
<th>QC</th>
<th>ON</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Proportions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>94.5</td>
<td>93.9</td>
<td>94.2</td>
<td>94.8</td>
<td>94.3</td>
<td>94.1</td>
<td>94.2</td>
<td>94.3</td>
<td>94.5</td>
<td>94.8</td>
</tr>
<tr>
<td>B</td>
<td>94.7</td>
<td>94.0</td>
<td>94.4</td>
<td>94.8</td>
<td>94.7</td>
<td>94.7</td>
<td>94.6</td>
<td>94.7</td>
<td>94.7</td>
<td>95.2</td>
</tr>
<tr>
<td><strong>Totals</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>93.6</td>
<td>93.1</td>
<td>94.0</td>
<td>93.4</td>
<td>93.6</td>
<td>93.8</td>
<td>94.0</td>
<td>93.9</td>
<td>94.3</td>
<td>94.4</td>
</tr>
<tr>
<td>B</td>
<td>94.1</td>
<td>93.5</td>
<td>94.2</td>
<td>93.6</td>
<td>93.9</td>
<td>94.1</td>
<td>94.1</td>
<td>94.2</td>
<td>94.1</td>
<td>94.5</td>
</tr>
</tbody>
</table>

Table 6.40 Averages of empirical coverages of nominal 95% confidence intervals for proportions and totals. T=Taylor, B=bootstrap.

Two stage, $\psi = 0.02$.

### 6.6.2 Interval Widths

Because the confidence intervals are normal theory intervals, the widths are determined by the sizes of the square roots of the MSE estimators. In the previous section, we observed that the relative magnitudes of the bootstrap and Taylor MSE estimators depend on the value of $\psi$ and the parameter being estimated (proportion or total). The bootstrap and Taylor MSE estimators are more similar for totals than for proportions.

Tables 6.41 and 6.42 contain the MC averages of the prediction interval widths for the proportions based on the Taylor and Bootstrap MSE estimators, respectively, for the two stage simulation model with $\psi = 0.003$. As we expect, the average interval widths tend to decrease as the province sizes increase. In each province, the average interval widths are smaller for categories $i = 2$ and $i = 4$ than for categories $i = 1$ and $i = 3$ because the expected values of the proportions are smaller for categories $i = 2$ and $i = 4$ than for categories $i = 1$ and $i = 3$. The average widths of the intervals based on the bootstrap MSE estimators are smaller than the corresponding widths based on the Taylor MSE estimator for all of the cells except category $i = 1$ in Prince Edward Island and New Foundland. However, the empirical coverages of prediction intervals based on the bootstrap MSE estimator are smaller than 95% for the two
stage simulation model with $\psi = 0.003$.

<table>
<thead>
<tr>
<th></th>
<th>PE</th>
<th>NF</th>
<th>NB</th>
<th>NS</th>
<th>SK</th>
<th>MB</th>
<th>BC</th>
<th>AB</th>
<th>QC</th>
<th>ON</th>
</tr>
</thead>
<tbody>
<tr>
<td>$i = 1$</td>
<td>0.167</td>
<td>0.149</td>
<td>0.141</td>
<td>0.141</td>
<td>0.138</td>
<td>0.144</td>
<td>0.136</td>
<td>0.139</td>
<td>0.138</td>
<td>0.127</td>
</tr>
<tr>
<td>$i = 2$</td>
<td>0.110</td>
<td>0.105</td>
<td>0.104</td>
<td>0.120</td>
<td>0.101</td>
<td>0.104</td>
<td>0.103</td>
<td>0.100</td>
<td>0.108</td>
<td>0.101</td>
</tr>
<tr>
<td>$i = 3$</td>
<td>0.166</td>
<td>0.171</td>
<td>0.141</td>
<td>0.141</td>
<td>0.138</td>
<td>0.137</td>
<td>0.139</td>
<td>0.145</td>
<td>0.137</td>
<td>0.128</td>
</tr>
<tr>
<td>$i = 4$</td>
<td>0.102</td>
<td>0.144</td>
<td>0.113</td>
<td>0.145</td>
<td>0.098</td>
<td>0.106</td>
<td>0.102</td>
<td>0.100</td>
<td>0.111</td>
<td>0.090</td>
</tr>
</tbody>
</table>

Table 6.41 Average widths of nominal 95% prediction intervals for proportions constructed with Taylor MSE estimator. 2-stage sampling error model, $\psi = 0.003$

<table>
<thead>
<tr>
<th></th>
<th>PE</th>
<th>NF</th>
<th>NB</th>
<th>NS</th>
<th>SK</th>
<th>MB</th>
<th>BC</th>
<th>AB</th>
<th>QC</th>
<th>ON</th>
</tr>
</thead>
<tbody>
<tr>
<td>$i = 1$</td>
<td>0.168</td>
<td>0.151</td>
<td>0.140</td>
<td>0.140</td>
<td>0.136</td>
<td>0.142</td>
<td>0.134</td>
<td>0.136</td>
<td>0.135</td>
<td>0.122</td>
</tr>
<tr>
<td>$i = 2$</td>
<td>0.109</td>
<td>0.103</td>
<td>0.102</td>
<td>0.118</td>
<td>0.097</td>
<td>0.100</td>
<td>0.099</td>
<td>0.097</td>
<td>0.104</td>
<td>0.096</td>
</tr>
<tr>
<td>$i = 3$</td>
<td>0.165</td>
<td>0.169</td>
<td>0.140</td>
<td>0.140</td>
<td>0.135</td>
<td>0.134</td>
<td>0.136</td>
<td>0.142</td>
<td>0.134</td>
<td>0.123</td>
</tr>
<tr>
<td>$i = 4$</td>
<td>0.100</td>
<td>0.141</td>
<td>0.111</td>
<td>0.142</td>
<td>0.095</td>
<td>0.102</td>
<td>0.099</td>
<td>0.097</td>
<td>0.109</td>
<td>0.087</td>
</tr>
</tbody>
</table>

Table 6.42 Average widths of nominal 95% prediction intervals for proportions constructed with bootstrap MSE estimator. 2-stage sampling error model, $\psi = 0.003$

Tables 6.43 and 6.44 contain the average widths of the nominal 95% prediction intervals for the proportions and totals for the four simulation models. The average widths in Table 6.43 are based on the bootstrap MSE estimator, and the average widths in Table 6.44 are based on the Taylor MSE estimator. The averages are taken across the 2000 MC samples and the 40 cells of the two-way table. The relative magnitudes of the average interval widths are consistent with the MC means of the MSE estimators. For the proportions, the average widths based on the bootstrap MSE estimators are smaller than the corresponding average widths based on the Taylor MSE estimator for three out of the four simulation models. In the two stage simulation
with $\psi = 0.02$, the average of the widths based on the bootstrap MSE estimator is larger than the average of the widths based on the Taylor MSE estimators. For the totals, the average widths based on the Taylor MSE estimator are larger than the average widths based on the bootstrap MSE estimator for the SRS simulation, and the opposite pattern holds for the two stage simulation.

<table>
<thead>
<tr>
<th>Proportions</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.003 0.02</td>
<td>0.003 0.02</td>
</tr>
<tr>
<td>SRS 0.1088 0.1682</td>
<td>4195 4799</td>
</tr>
<tr>
<td>2-stage 0.1232 0.1848</td>
<td>4634 5490</td>
</tr>
</tbody>
</table>

Table 6.43 Average widths of nominal 95% prediction intervals for proportions and totals constructed with bootstrap MSE estimator.

<table>
<thead>
<tr>
<th>Proportions</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.003 0.02</td>
<td>0.003 0.02</td>
</tr>
<tr>
<td>SRS 0.1128 0.1704</td>
<td>4221 4877</td>
</tr>
<tr>
<td>2-stage 0.1257 0.1822</td>
<td>4578 5379</td>
</tr>
</tbody>
</table>

Table 6.44 Average widths of nominal 95% prediction intervals for proportions and totals constructed with Taylor MSE estimator.

### 6.6.3 Summary

The empirical coverages of normal theory 95% prediction intervals based on both the bootstrap and Taylor MSE estimators are typically between 93% and 96%. For both the bootstrap and Taylor intervals, changing $\psi$ has more of an effect on the properties of the average empirical coverages than changing the sampling error model. For example, the properties of the average empirical coverages of the Taylor intervals are similar at $\psi = 0.003$ for both
the two stage and SRS sampling error models. When $\psi = 0.003$, the empirical coverages for the proportions based on either method tend to increase as the province sizes increase. Both methods lead to average empirical coverages closer to 95% in the small provinces when $\psi = 0.02$ than when $\psi = 0.003$. The average empirical coverages of the bootstrap intervals are typically smaller than the corresponding average empirical coverages of the Taylor intervals for the proportions, which may occur because the bootstrap estimators of the leading terms in the MSE’s of the predictors of the proportions are typically smaller than the corresponding Taylor estimators. The average empirical coverages of the bootstrap and Taylor intervals are more similar for totals than for proportions, which is consistent with the observation noted in the previous subsection that the MC means of the bootstrap and Taylor estimators of the leading terms in the MSE’s are more similar for totals than for proportions. The differences between average empirical coverages of the bootstrap and Taylor intervals are smaller at $\psi = 0.02$ than at $\psi = 0.003$. As discussed in the Section 6.5, the lower bound on the estimator of $\psi$ leads to many of the differences between the bootstrap and Taylor MSE estimators. Because the lower bound effect is negligible when $\psi = 0.02$, it is not surprising that the properties of the empirical coverages of intervals based on the bootstrap and Taylor MSE estimators are more similar at $\psi = 0.02$ than at $\psi = 0.003$. Despite the flaws of the MSE estimators both the bootstrap and Taylor estimators lead to confidence intervals with coverages between 93% and 96%.

### 6.7 Augmented Model Predictors

The efficiencies of the augmented model predictors are compared to the efficiencies of the raked predictors in this section. The model for the simulations in this section is a reduced version of the model used for the previous simulations in which the category effects for the current time point are proportional to the category effects from the Census, and the coefficient on the category effects is equal to the coefficient on the Census interactions. Formally, the model for the mean is

$$\log[p_{T,ik}p_{T,1k}^{-1}] = \lambda \log[p_{c,ik}p_{c,1k}^{-1}].$$
In the notation of Section 4.2.3, the covariate $x_{ik} = \log[p_{c,ik}p_{c,1k}^{-1}]$, and $\lambda$ is the parameter to estimate. In the model for the simulation, $\lambda = 1$, so $p_{c,ik} = p_{T,ik}$. The expected values of the totals in the current time point differ from the Census totals because the expected values of the province 2-digit totals differ from the Census 2-digit totals. The Census proportions in Table 6.1 and the province 2-digit totals in Table 6.2 are used to define $p_{T,ik}$ and $T_{ik}$ for this simulation. Three different configurations of expected sample sizes, $\{n_k : k = 1, \ldots, K\}$ are used. In the first, $n_k = [0.5T_k^{0.5}]$, where $[a]$ denotes the integer closest to $a$. In the second configuration, the $\{n_k : k = 1, \ldots, K\}$ from the first are permuted so that the largest sample size is not associated with the largest $T_k$. In the third simulation, $n_k = 70$ for all $k$. In all simulations, $\psi = 0.02$, and the sampling errors are generated from the SRS model. For all of the predictors, $\hat{\gamma}_k = (\hat{c}_k n_k^{-1} + \hat{\psi})^{-1}\hat{\psi}$; the Beale estimator of $\gamma_k$ is not used.

The averages of the ratios of the MC MSE’s of the augmented model predictors, $\hat{p}_{pred,aug1}$ and $\hat{p}_{pred,aug2}$ defined in (4.40) and (4.42), respectively, to the MC MSE’s of the raked predictors $\tilde{p}_{ik}$ are in Tables 6.45- 6.47. In the tables, aug1/rak and aug2/rak are used to denote the averages of the ratios of the MSE’s of the augmented model predictors $\hat{p}_{pred,aug1}$ and $\hat{p}_{pred,aug2}$, respectively, to the MSE’s of the raked predictors. The notation aug1/aug2 denotes the averages of the ratios of the MSE’s of $\hat{p}_{pred,aug1}$ to to the MSE’s of $\hat{p}_{pred,aug2}$.

Four main patterns are clear from the tables. The predictors $\{\hat{p}_{pred,ik,aug2} : i = 1, \ldots, C; k = 1, \ldots, K\}$, arising from the second augmented model method, have smaller MC MSE’s than the predictors $\{\hat{p}_{pred,ik,aug1} : i = 1, \ldots, C; k = 1, \ldots, K\}$, in most of the provinces. In the provinces labeled British Columbia, Alberta and Quebec, the averages of the MSE ratios comparing the two augmented model predictors do not differ from one by more than two MC standard errors when the sample sizes increase with $T_k$. When the sample sizes are not proportional to $\sqrt{T_k}$, the average ratios of the MSE’s of the predictors based on augmented model method 1 to the MSE’s of the predictors based on augmented model method 2 do not differ significantly from 1 in several provinces. Regardless of the configuration of the expected sample sizes, the predictors based on the second augmented model method ($\hat{p}_{pred,ik,aug2}$) are at least as efficient as the raked predictors in all provinces except the province with the largest two
digit total, where raking is the most efficient method. With the exception of largest province, raking is inefficient when the province sample sizes do not increase as the 2-digit totals increase. For each method, the relative efficiencies of the predictors of the proportions are similar to the relative efficiencies for the totals. (The MSE ratios for the totals are omitted.)

<table>
<thead>
<tr>
<th></th>
<th>PE</th>
<th>NF</th>
<th>NB</th>
<th>NS</th>
<th>SK</th>
<th>MB</th>
<th>BC</th>
<th>AB</th>
<th>QC</th>
<th>ON</th>
</tr>
</thead>
<tbody>
<tr>
<td>aug2/rak</td>
<td>0.9829</td>
<td>0.9781</td>
<td>0.9781</td>
<td>0.9837</td>
<td>0.9785</td>
<td>0.9840</td>
<td>0.9890</td>
<td>0.9893</td>
<td>1.0006</td>
<td>1.0165</td>
</tr>
<tr>
<td>SE</td>
<td>(0.0028)</td>
<td>(0.0027)</td>
<td>(0.0022)</td>
<td>(0.0020)</td>
<td>(0.0018)</td>
<td>(0.0013)</td>
<td>(0.0014)</td>
<td>(0.0014)</td>
<td>(0.0012)</td>
<td></td>
</tr>
<tr>
<td>aug1/rak</td>
<td>1.0629</td>
<td>1.0143</td>
<td>0.9895</td>
<td>1.0049</td>
<td>0.9944</td>
<td>1.0097</td>
<td>0.9887</td>
<td>0.9879</td>
<td>0.9989</td>
<td>1.0396</td>
</tr>
<tr>
<td>SE</td>
<td>(0.0093)</td>
<td>(0.0065)</td>
<td>(0.0051)</td>
<td>(0.0054)</td>
<td>(0.0055)</td>
<td>(0.0054)</td>
<td>(0.0018)</td>
<td>(0.0025)</td>
<td>(0.0021)</td>
<td>(0.0026)</td>
</tr>
<tr>
<td>aug1/aug2</td>
<td>1.0814</td>
<td>1.0370</td>
<td>1.0116</td>
<td>1.0214</td>
<td>1.0162</td>
<td>1.0261</td>
<td>0.9998</td>
<td>0.9986</td>
<td>0.9983</td>
<td>1.0227</td>
</tr>
<tr>
<td>SE</td>
<td>(0.0082)</td>
<td>(0.0055)</td>
<td>(0.0039)</td>
<td>(0.0048)</td>
<td>(0.0047)</td>
<td>(0.0048)</td>
<td>(0.0011)</td>
<td>(0.0019)</td>
<td>(0.0016)</td>
<td>(0.0022)</td>
</tr>
<tr>
<td>nk</td>
<td>14</td>
<td>24</td>
<td>34</td>
<td>37</td>
<td>39</td>
<td>48</td>
<td>88</td>
<td>94</td>
<td>135</td>
<td>189</td>
</tr>
</tbody>
</table>

Monte Carlo standard error of average of MSE ratio

Table 6.45 aug2/rak: Average of ratios of MC MSE’s of \( \hat{p}_{\text{pred,ik,}\text{aug2}} \) to MC MSE’s of \( \hat{p}_{\text{ik}} \); aug1/rak: Average of ratios of MC MSE’s of \( \hat{p}_{\text{pred,ik,}\text{aug1}} \) to MC MSE’s of \( \hat{p}_{\text{ik}} \); aug1/aug2: Average of ratios of MC MSE’s of \( \hat{p}_{\text{pred,ik,}\text{aug1}} \) to MC MSE’s of \( \hat{p}_{\text{pred,ik,}\text{aug2}} \).
<table>
<thead>
<tr>
<th></th>
<th>PE</th>
<th>NF</th>
<th>NB</th>
<th>NS</th>
<th>SK</th>
<th>MB</th>
<th>BC</th>
<th>AB</th>
<th>QC</th>
<th>ON</th>
</tr>
</thead>
<tbody>
<tr>
<td>aug2/rak</td>
<td>0.7157</td>
<td>0.657</td>
<td>0.6236</td>
<td>0.4512</td>
<td>0.3959</td>
<td>0.5231</td>
<td>0.5158</td>
<td>0.6473</td>
<td>0.7987</td>
<td>1.727</td>
</tr>
<tr>
<td>SE</td>
<td>(0.0099)</td>
<td>(0.010)</td>
<td>(0.0090)</td>
<td>(0.0085)</td>
<td>(0.0086)</td>
<td>(0.0094)</td>
<td>(0.0086)</td>
<td>(0.0084)</td>
<td>(0.0053)</td>
<td>(0.015)</td>
</tr>
<tr>
<td>aug1/rak</td>
<td>0.722</td>
<td>0.661</td>
<td>0.6243</td>
<td>0.4513</td>
<td>0.3966</td>
<td>0.5247</td>
<td>0.5129</td>
<td>0.6429</td>
<td>0.7873</td>
<td>1.780</td>
</tr>
<tr>
<td>SE</td>
<td>(0.011 )</td>
<td>(0.011)</td>
<td>(0.0095)</td>
<td>(0.0087)</td>
<td>(0.0088)</td>
<td>(0.0099)</td>
<td>(0.0088)</td>
<td>(0.0089)</td>
<td>(0.0058)</td>
<td>(0.016)</td>
</tr>
<tr>
<td>aug1/aug2</td>
<td>0.0085</td>
<td>1.0064</td>
<td>1.0009</td>
<td>1.0000</td>
<td>1.0016</td>
<td>0.9945</td>
<td>0.9929</td>
<td>0.9861</td>
<td>1.0305</td>
<td></td>
</tr>
<tr>
<td>SE</td>
<td>(0.0029)</td>
<td>(0.0022)</td>
<td>(0.0019)</td>
<td>(0.0017)</td>
<td>(0.0016)</td>
<td>(0.0022)</td>
<td>(0.0015)</td>
<td>(0.0018)</td>
<td>(0.0015)</td>
<td>(0.0010)</td>
</tr>
<tr>
<td>n_k</td>
<td>14</td>
<td>34</td>
<td>39</td>
<td>135</td>
<td>189</td>
<td>88</td>
<td>48</td>
<td>37</td>
<td>24</td>
<td></td>
</tr>
</tbody>
</table>

Table 6.46  aug2/rak: Average of ratios of MC MSE’s of $\hat{p}_{pred,ik,aug2}$ to MC MSE’s of $\hat{p}_{i,k}$; aug1/rak: Average of ratios of MC MSE’s of $\hat{p}_{pred,ik,aug1}$ to MC MSE’s of $\hat{p}_{i,k}$; aug1/aug2: Average of ratios of MC MSE’s of $\hat{p}_{pred,ik,aug1}$ to MC MSE’s of $\hat{p}_{pred,ik,aug2}$.

<table>
<thead>
<tr>
<th></th>
<th>PE</th>
<th>NF</th>
<th>NB</th>
<th>NS</th>
<th>SK</th>
<th>MB</th>
<th>BC</th>
<th>AB</th>
<th>QC</th>
<th>ON</th>
</tr>
</thead>
<tbody>
<tr>
<td>aug2/rak</td>
<td>0.830</td>
<td>0.8373</td>
<td>0.8337</td>
<td>0.8585</td>
<td>0.8321</td>
<td>0.8359</td>
<td>0.8561</td>
<td>0.8652</td>
<td>0.9589</td>
<td>1.2213</td>
</tr>
<tr>
<td>SE</td>
<td>(0.0087)</td>
<td>(0.0093)</td>
<td>(0.0088)</td>
<td>(0.0087)</td>
<td>(0.0083)</td>
<td>(0.0079)</td>
<td>(0.0060)</td>
<td>(0.0055)</td>
<td>(0.0026)</td>
<td>(0.0061)</td>
</tr>
<tr>
<td>aug1/rak</td>
<td>0.838</td>
<td>0.846</td>
<td>0.8361</td>
<td>0.8622</td>
<td>0.8362</td>
<td>0.8385</td>
<td>0.8515</td>
<td>0.8589</td>
<td>0.9527</td>
<td>1.2604</td>
</tr>
<tr>
<td>SE</td>
<td>(0.010 )</td>
<td>(0.010)</td>
<td>(0.0097)</td>
<td>(0.0097)</td>
<td>(0.0095)</td>
<td>(0.0090)</td>
<td>(0.0065)</td>
<td>(0.0061)</td>
<td>(0.0030)</td>
<td>(0.0070)</td>
</tr>
<tr>
<td>aug1/aug2</td>
<td>1.0100</td>
<td>1.0100</td>
<td>1.0029</td>
<td>1.0038</td>
<td>1.0049</td>
<td>1.0029</td>
<td>0.9946</td>
<td>0.9927</td>
<td>0.9936</td>
<td>1.0320</td>
</tr>
<tr>
<td>SE</td>
<td>(0.0033)</td>
<td>(0.0026)</td>
<td>(0.0023)</td>
<td>(0.0025)</td>
<td>(0.0027)</td>
<td>(0.0028)</td>
<td>(0.0014)</td>
<td>(0.0017)</td>
<td>(0.0013)</td>
<td>(0.0014)</td>
</tr>
<tr>
<td>n_k</td>
<td>70</td>
<td>70</td>
<td>70</td>
<td>70</td>
<td>70</td>
<td>70</td>
<td>70</td>
<td>70</td>
<td>70</td>
<td></td>
</tr>
</tbody>
</table>

Table 6.47  aug2/rak: Average of ratios of MC MSE’s of $\hat{p}_{pred,ik,aug2}$ to MC MSE’s of $\hat{p}_{i,k}$; aug1/rak: Average of ratios of MC MSE’s of $\hat{p}_{pred,ik,aug1}$ to MC MSE’s of $\hat{p}_{i,k}$; aug1/aug2: Average of ratios of MC MSE’s of $\hat{p}_{pred,ik,aug1}$ to MC MSE’s of $\hat{p}_{pred,ik,aug2}$.
CHAPTER 7. Canadian Labour Force Survey

In this Section, the model of Section 3 is fit to a subset of the LFS data, and the procedures described in Sections 4 and 5 are applied to obtain estimators of the model parameters, predictors, and prediction MSE’s. Section 7.1 describes the National Occupational Classification system and introduces notation to describe the hierarchical data structure. Relevant features of the LFS and Census data collection and estimation procedures are summarized in Section 7.2. In Section 7.3, specific modifications to the estimation and prediction procedures used for this data example are explained. Section 7.4 deals with estimation of the working model for the sampling variances. In Sections 7.5 and 7.6, the estimators of the model parameters and the predictors are presented for the two digit codes A1 (specialist managers) and E0 (judges, lawyers, psychologists, social workers, ministers of religion, and policy and program officers).

7.1 National Occupational Classification

Canada’s National Occupational Classification (NOC) organizes employment into occupations using a hierarchical system. Ten one digit codes (labeled A through J) are partitioned into a total of forty-seven two digit codes. Sixteen of the two digit codes are not further subdivided into three digit codes. One two digit code has ten three digit codes. The remaining thirty two digit codes have between two and 8 three digit codes.

Notation for the Occupational Classification System

The specification of the model (3.1) in Section 3 is for a single two digit code. The analysis of this section uses data from multiple two digit codes. We specify notation to describe the structure of the occupational classification. Let $s$ index the one digit codes, where $s =$
1, 2, ..., 10 for A through J. Let \( t \) index the two digit codes nested in the one digit codes, where \( t = 1, \ldots, D_s \), and \( D_s \) denotes the number of two digit codes with at least two three digit codes in one digit code \( s \). For example, \( \hat{p}_{ik,ts} \) denotes the direct estimator of the proportion in three digit code \( i \), province \( k \), two digit code \( t \), and one digit code \( s \). In the previous sections, \( C \) denotes the number of three digit codes in a two digit code. Because different two digit codes have different numbers of categories, we use \( C_{ts} \) to denote the number of three digit codes in two digit code \( t \) in one digit code \( s \). When dealing with the one digit codes separately, we suppress the index \( s \) for simplicity. Similarly, we omit the subscript \( ts \) when dealing with a single two digit code.

### 7.2 Census and LFS Data

The LFS uses a stratified two stage cluster sample. Strata are intersections of Economic Regions and Employment Insurance Economic Regions. Primary sampling units are clusters, each consisting of approximately 200 households. A systematic sample of households is selected from each sampled cluster, and all members of selected households are included in the sample. The total number of sampled households is approximately 53,000, leading to a sampling rate of approximately 1/240. One-sixth of the sample is replaced each month.

Direct estimators of occupation totals are calculated on a monthly basis with the LFS data. Direct estimators are weighted sums of sampled units. The original sampling weights are adjusted to account for nonresponse, control totals, and the sample overlap. The jackknife is used to obtain a direct estimator of the sampling covariance matrix of the direct estimators of the totals. The jackknife procedure involves calculating replicates of the weighted estimates after omitting each sampled cluster and recomputing the adjusted weights. Direct estimators of province two digit totals and national three digit totals are published on a monthly basis. LFS estimates for three digit codes are not published at the province level because estimated coefficients of variation exceed reliability standards. In 2008, the Human Resources and Skills Development department in Canada requested an investigation of the possibility of obtaining estimates at the three digit level by province.
Table 7.1 shows the estimated coefficients of variation for the three digit proportions in the two digit code A1. The last two rows of Table 7.1 have the realized province sample sizes and the direct estimators of the province two digit totals. The CV’s for the proportions in three digit codes A12 and A14 in small provinces often exceed 50%. Because larger sample sizes are allocated to larger provinces, the estimated CV’s tend to decrease as the province sizes increase. In each province except Nova Scotia, the estimated CV’s in three digit codes A12 and A14 are larger than the estimated CV’s in A11 and A13. As discussed further in Section 7.5, this pattern arises because the estimated proportions in A11 and A13 are larger than the estimated proportions in A12 and A14 (in all provinces except Nova Scotia), and the direct estimates of the variances are related to the direct estimates of the means. The estimate of the CV in A14 in Nova Scotia is smaller than the estimate of the CV in A11 because the direct estimate of the proportion in A11 is smaller than the direct estimate of the proportion in A14.

<table>
<thead>
<tr>
<th></th>
<th>PE</th>
<th>NF</th>
<th>NB</th>
<th>NS</th>
<th>SK</th>
<th>MB</th>
<th>BC</th>
<th>AB</th>
<th>QC</th>
<th>ON</th>
</tr>
</thead>
<tbody>
<tr>
<td>A11</td>
<td>0.34</td>
<td>0.46</td>
<td>0.28</td>
<td>0.40</td>
<td>0.28</td>
<td>0.17</td>
<td>0.15</td>
<td>0.17</td>
<td>0.12</td>
<td>0.09</td>
</tr>
<tr>
<td>A12</td>
<td>0.69</td>
<td>1.18</td>
<td>0.43</td>
<td>0.52</td>
<td>0.30</td>
<td>0.76</td>
<td>0.21</td>
<td>0.21</td>
<td>0.21</td>
<td>0.18</td>
</tr>
<tr>
<td>A13</td>
<td>0.26</td>
<td>0.25</td>
<td>0.20</td>
<td>0.16</td>
<td>0.18</td>
<td>0.16</td>
<td>0.16</td>
<td>0.17</td>
<td>0.13</td>
<td>0.08</td>
</tr>
<tr>
<td>A14</td>
<td>1.20</td>
<td>0.54</td>
<td>0.60</td>
<td>0.26</td>
<td>0.25</td>
<td>0.28</td>
<td>0.37</td>
<td>0.28</td>
<td>0.37</td>
<td>0.16</td>
</tr>
<tr>
<td>(n_k)</td>
<td>16</td>
<td>20</td>
<td>34</td>
<td>42</td>
<td>54</td>
<td>66</td>
<td>105</td>
<td>116</td>
<td>166</td>
<td>368</td>
</tr>
<tr>
<td>(\hat{M}_k)</td>
<td>770</td>
<td>2243</td>
<td>4538</td>
<td>6091</td>
<td>5622</td>
<td>9220</td>
<td>35052</td>
<td>31323</td>
<td>72926</td>
<td>142835</td>
</tr>
</tbody>
</table>

Table 7.1 Estimated CV’s for A1, province sample sizes, and direct estimates of province 2-digit totals.

The Canadian Census of Population, conducted every five years, publishes occupational counts through the three digit level of detail for each province. The Census estimates are based on a 1/5 sample, so they are subject to less sampling variation than the LFS estimators. The model and estimator that we use requires positive Census totals. When a Census total is
zero, the zero estimate is replaced by the estimate under an assumption that provinces and three digit codes are independent. This modification to the Census table is discussed in more detail in Section 7.3. The observed Census totals, after modifying zeros, are treated as fixed covariates in the model for the expected values of the direct estimators of the proportions.

7.3 Estimation and Prediction for Canadian LFS

The procedure of Sections 4 and 5 is applied to obtain predictors for the two digit codes A1 and E0. In this subsection, we explain some specific changes to the procedure that we implement for this application.

Estimates of the parameters of model (3.1) are obtained under a reduced model. The reduced model reflects the procedures implemented in Hidiroglou and Patak (2009 a), where the SPREE procedure is used to estimate the expected value, \( p_{T,ik} \). The model underlying SPREE preserves the interactions in the auxiliary table. The equivalent reduced model is the model in which the coefficients on the Census interactions, \( \{ \theta_{o,ts} : t = 1, \ldots, D_s; s = 1, \ldots, 10 \} \), are assumed to equal 1. Hidiroglou and Patak (2009 a) use data from multiple years to estimate a common value of the variance parameter \( \psi \) for each two digit code. Because only the direct estimates for May 2005 are available to us, we combine information across two digit codes instead of years. We are interested in a reduced model in which the parameters \( \{ \psi_{ts} : t = 1, \ldots, D_s \} \) are assumed to be the same for all two digit codes \( t \) in one digit code \( s \).

Use of a model for the sampling variances has the potential to lead to predictors with smaller MSE’s than predictors calculated with the direct estimators of the sampling variances. The leading terms of the proposed estimators of the MSE’s do not rely on an assumption that the working model for the sampling variances is true. The MSE’s of the predictors depend on the variance of \( \hat{c}_k \). Because we have doubts about the quality of our estimators of the variance of \( \hat{c}_k \), we desire an estimator of \( c_k \) with a small variance. An estimator of \( c_k \) based on multiple two-way tables is defined in (4.5) of Section 4.1.1. In an attempt to obtain an estimator of \( c_k \) with a small variance, all 31 two digit codes are used to construct the estimator defined in (4.5). The goodness of fit of the working model for the sampling variances in the two digit
codes A1 and E0 and the decision to combine the estimates of $c_k$ across the two digit codes are discussed in Section 7.4.

Modifications to the bootstrap procedure described in Section 5.2 are implemented to accommodate the use of multiple two digit codes to estimate the model parameters. As explained above, the ten one digit codes ($s = 1, \ldots, 10$) are partitioned into 31 two digit codes with at least two three digit codes. For each province, $k$, the data for all of the two digit codes ($t = 1, \ldots, D_s; s = 1, \ldots, 10$) are used to obtain a common estimate, $\hat{c}_k$, as shown in (4.5) of Section 4.1.1. Let $\hat{c}_k$ denote the original estimate of $c_k$ obtained from the 31 two digit codes. The estimator of $p_{T,ik,ts}$ uses an estimator of the coefficient on the Census interactions, previously denoted $\theta_{o,ts}$. The estimates of $\theta_{o,ts}$ are restricted to equal one for all of the two digit codes in the ten one digit codes ($\hat{\theta}_{ts} = 1$ for $t = 1, \ldots, D_s; s = 1, \ldots, 10$). Another assumption of the reduced model is that $\psi_{ts} = \psi_s$ for $t = 1, \ldots, D_s$, and $s = 1, \ldots, 10$. For each one digit code $s$, the two digit codes $t = 1, \ldots, D_s$ are used to obtain a common estimate $\hat{\psi}_s$ as shown in (4.21) and (4.22). Let $\hat{\psi}_s$ denote the estimate of $\psi_s$ under the assumption that $\theta_{o,ts} = 1$ and the assumption that $\psi_{ts} = \psi_s$ for $t = 1, \ldots, D_s$. The bootstrap procedure involves completing the following steps for $b = 1, \ldots, B$.

(i) Generate bootstrap versions of the direct estimators for all of the two digit codes in each of the 10 one digit codes using the distributions of Section 5.2.1 with $\hat{\theta}_{ts} = 1$ and $\hat{\psi}_{ts} = \hat{\psi}_s$ for $t = 1, \ldots, D_s; s = 1, \ldots, 10$.

(ii) Compute the SPREE estimate $\hat{p}^{(b)}_{T,ik,ts}$ using the direct estimators generated in the first step.

(iii) Obtain a bootstrap estimate, $\hat{c}_k^{(b)}$, using $\{\hat{\Sigma}_{ee,k,ts}^{(b)} : t = 1, \ldots, D_s; s = 1, \ldots, 10\}$ from step (i) and the SPREE estimates from step (ii).

(iv) Using a maximum of three steps of the iterative estimation procedure, update the estimates of $\{p_{T,ik,ts} : i = 1, \ldots, C_{ts}; k = 1, \ldots, 10; t = 1, \ldots, D_s; s = 1, 5\}$ and obtain estimates $\{\hat{\psi}_s^{(b)} : s = 1, 5\}$. (Indices $s = 1$ and 5 are for the one digit codes A and E, respectively.) The coefficients on the Census interactions are restricted to equal 1.
(v) Construct predictors of the proportions and totals for A1 and E0 using the direct estimates generated in the first step and the parameter estimates obtained in steps (iii) and (iv).

Estimates of the MSE’s are obtained as described in the bootstrap algorithm of Section 5. Bootstrap estimators of the variances of the parameter estimators are the sample variances of the B bootstrap estimates. The bootstrap sample size used is B=2000.

The bootstrap procedure described above involves generating data sets (bootstrap versions of direct estimators of the totals in the two-way table and bootstrap versions of direct estimators of sampling covariance matrices) for all 31 two digit codes from the original estimates $\hat{p}_{T,ik}, \hat{c}_k,$ and $\hat{\psi}_s,$ and $\hat{\Sigma}_{ee,k,md}.$ The 31 generated data sets are used to obtain bootstrap versions of the estimators of $c_k$ for $k = 1, \ldots, K.$ Computing the bootstrap estimate of $c_k$ requires the SPREE estimates for the 31 two digits codes in each bootstrap sample. For this illustration, we are only interested in constructing predictors for the two digit codes A1 and E0. Therefore, bootstrap estimates of $\psi_s$ are only obtained for $s = 1, 5$ (A, E). Likewise, bootstrap versions of the predictors of the totals and proportions are only obtained for A1 and E0.

In our implementation of the bootstrap data generating procedure of Section 5.2, the R function “qgamma” is used to convert correlated uniform random variables to correlated gamma random variables. When the shape parameter of the gamma distribution, denoted $\alpha_{ik}^\ast,$ is sufficiently small, the R function returns a missing value. In cases where the R function returns a missing value, the bootstrap direct estimator, $\hat{M}_{ik}^\ast,$ is set equal to the bootstrap true value, $M_{ik}^\ast.$ In our application of the bootstrap to the LFS data, the modification for small $\alpha_{ik}^\ast$ is used when generating bootstrap samples for A3, C0, I1, and J1. In 456 of the 2000 bootstrap samples, the shape parameter of the gamma distribution for A39 in Prince Edward Island was too small. In A36, the shape parameter was too small in at least eleven bootstrap samples in all provinces except New Brunswick and Nova Scotia, the only two provinces where $\hat{p}_{T,ik}$ for A36 exceeds 0.012. In A36, the modification was used in 119 samples in New Foundland, in 95 samples in Prince Edward Island, in 16 samples in Quebec, in 11 samples in Ontario, in 23 samples in Manitoba, in 71 samples in Saskatchewan, in 56 samples in Alberta, and in
24 samples in British Columbia. In New Brunswick and Nova Scotia, none of the bootstrap samples required the modification for small $\alpha_{ik}^*$ for any of the categories. The modification was also used in A32 in 17 samples in New Foundland and in 39 samples in Prince Edward Island.

The generated $\{\hat{\psi}_s^{*}\} : s = 1, \ldots, 10; b = 1, \ldots, B\}$ are used to test the null hypothesis $\psi_{ts} = \psi_s$ for $t = 1, \ldots, D_s$ using the score test statistic defined in (4.23). Let $S_{\psi(s)}^{*}(b)$ denote the score test statistic based on the bootstrap versions of the direct estimators in one digit code $s$. A p-value for testing the null hypothesis that the two digit codes in a single one digit code are generated from a common value, $\psi_s$, is the fraction of generated $\{S_{\psi(s)}^{*}(b) : b = 1, \ldots, B\}$ that exceed the observed score statistic calculated with the original data.

We also use the bootstrap to test the null hypotheses that $\psi_5 = 0$ ($s = 5$ for the one digit code E). The bootstrap test procedure consists of simulating bootstrap data sets with the value of $\hat{\psi}_{15}$ set equal to zero for the two digit codes in the one digit code E. The estimates $\{\hat{\psi}_s : s \neq 5\}$ are used to generate the bootstrap data sets for the other one digit codes. This data generating procedure is equivalent to setting $\hat{p}_{k,t5}^{*}$ equal to the original synthetic estimate, $\hat{p}_{T,k,t5}$ in each bootstrap sample. Let $\hat{\psi}_{5,0}$ be the estimator of $\psi_5$ obtained from the bootstrap data set with $\hat{p}_{k,t5}$ equal to $\hat{p}_{T,k,t5}$. A bootstrap p-value for testing the null hypothesis that $\psi_5 = 0$ is the proportion of generated $\{\hat{\psi}_{5,0}^{*}(b) : b = 1, \ldots, B\}$ that exceed the original estimate, $\hat{\psi}_5$.

The two digit codes H0 and H7 contain zeros in the Census. If a Census table has zeros, then the interactions in a saturated loglinear model fit to the Census table are not defined (Agresti, 2003, pg. 70). As a consequence, the covariate $(\alpha \beta)^{cen}_{ik}$ used in the model for $p_{T,ik}$ is not defined. A solution to this problem is to use the SPREE procedure, which converges to a solution in the presence of a limited number of zero counts. Using SPREE in the presence of zeros, however, is not completely satisfying because the linear approximation used to derive the variance of the SPREE estimators uses an assumption that the Census totals are all positive. To avoid the problems associated with zeros in the Census, we replace a zero cell in the Census with the predicted value under an assumption that rows and columns are independent in the Census table. More precisely, if $N_{ik} = 0$, then we replace $N_{ik}$ with
\( \tilde{N}_{ik} = N_1 N_{ik} N_{..}^{-1} \) and use \( \tilde{N}_{ik} \) to construct the Census interactions needed for the estimator of \( p_{T,ik} \). The marginal totals and interactions of the new Census table (with entry \( \tilde{N}_{ik} \)) differ from the marginal totals and interactions in the original Census table (with entry \( N_{ik} \)). The modification can be viewed as a redefinition of the covariate used in the model for \( p_{T,ik} \). If there exists a number \( \theta_0 \) such that the \( E[\hat{p}_{ik}] \) has the form defined by (3.3) and (3.4), with Census interactions obtained from \( \tilde{N}_{ik} \), then the modification to the Census table to eliminate zeros does not bias the predictors. Because our objective is not to estimate a relationship between the means of the direct estimators for May 2005 and the true 2001 occupational totals, we are not concerned about errors in the Census. Instead, we are interested in estimating a relationship between the means of the 2005 direct estimators and the observed covariates. The observed covariates are functions of \( \tilde{N}_{ik} \) when \( N_{ik} = 0 \).

The LFS uses the jackknife to construct a direct estimator of the sampling variance for the totals. The direct estimator of the sampling variance for the proportions is obtained from the jackknife covariance matrix using the linear approximation in Appendix 1. When direct estimates are zero, the corresponding direct estimates of the variances and covariances are also zero. When a direct estimate is zero, we replace the zero variance estimate with the estimate of the variance under the working model, as shown in equations (4.6) and (4.7) in Section 4.

In Section 5.2, Taylor and bootstrap estimates of the MSE's are developed. The Taylor MSE estimator uses an approximation for the MSE of the initial (not benchmarked) predictor. The bootstrap MSE estimator is developed as an attempt to account for the effects of raking on the MSE. In Section 7.5 and Section 7.6, the raking operation has a noticeable effect on the predictors. Therefore, we use the bootstrap MSE estimates for the LFS example.

### 7.4 Sampling Variances

The direct estimate of the sampling variance of the vector of proportions is obtained by applying the Taylor linearization in Appendix 1 to the jackknife estimate of the covariance matrix of the direct estimators of the totals. As observed in the simulation of Section 6, use of the direct estimates of the sampling variances to form the predictors can lead to a larger
MSE than use of a model estimate of the sampling variance. In the simulation, the predictor constructed with the direct estimator is less efficient than the predictor constructed under the working model because the direct estimators of the sampling variances have high variances and are correlated with the direct estimators of the proportions. Use of the model based estimate of the covariance matrix to estimate the MSE’s of the predictors leads to biased MSE estimators if the working covariance structure is wrong. In this section, we investigate properties of the direct estimates of the sampling variances. Because we present results for A1 and E0, we restrict our diagnosis of the goodness of fit of the working model to the two digit codes A1 and E0.

Under the working model for $\Sigma_{ee,k,ts}$, $\Sigma_{ee,k,ts} = c_{k,ts}n_{k,ts}^{-1}\Gamma_{uu,k,ts}$ for some constant $c_{k,ts}$. Under an assumption that $E[\hat{\Sigma}_{ee,k,ts}] = \Sigma_{ee,k}$, where $\hat{\Sigma}_{ee,k}$ is the direct estimator of the covariance matrix of $e_k$, the assumption that the working model for the variance of $e_k$ is true is equivalent to an assumption that $E[\hat{\Sigma}_{ee,k,ts}] = E[d_{k,ts}n_{k,ts}^{-1}(\text{diag}(\hat{p}_{k,ts}) - \hat{p}_{k,ts}\hat{p}_{k,ts}')\Gamma_{uu,k,ts}]$ for a constant $d_{k,ts}$. If the sample design is such that the the covariance matrix of the vector of sampling errors for each province is proportional to a multinomial covariance matrix, then an unbiased estimator of the sampling covariance matrix of the vector of proportions in two digit code $t$, one digit code $s$ and province $k$ is proportional to $[\text{diag}(\hat{p}_{k,ts}) - \hat{p}_{k,ts}\hat{p}_{k,ts}']\Gamma_{uu,k,ts}^{-1}$. (Examples of such designs include simple random samples and some cluster samples.) Because of the complexity of the LFS design and adjustments for nonresponse and population controls used in the estimation procedure, the direct estimators of the sampling variances are not proportional to $[\text{diag}(\hat{p}_{k,ts}) - \hat{p}_{k,ts}\hat{p}_{k,ts}']\Gamma_{uu,k,ts}^{-1}$.

To check for a relationship between the direct estimators of the sampling variances and the unbiased estimators under a simple random sample design, we plot the vector half of the direct estimate of the sampling covariance matrix for each two digit code against the vector half of $[\text{diag}(\hat{p}_{k,ts}) - \hat{p}_{k,ts}\hat{p}_{k,ts}']\Gamma_{uu,k,ts}^{-1}$. Figure 7.1 shows the plots for the two digit codes A1 and E0 ($ts = 11$ and $ts = 15$). The filled circles are for the two digit code A1, and the open triangles are for the two digit code E0. The values of $R^2$ from ordinary least squares regressions of the vector halves of the direct estimators of the covariance matrices on the vector halves of the
\[ \{ \text{diag}(\hat{p}_{k,ts}) - \hat{p}_{k,ts} \hat{p}_{k,ts}' \} n_{k,ts}^{-1} : t = 1, \ldots, D_s; s = 1, 2, 3 \} \] range from 0.938 (New Foundland) to 0.996 (Manitoba). The plots in Figure 7.1 and the values of \( R^2 \) suggest a linear association between the direct estimates of the sampling variances and the multinomial covariance matrix.

As discussed in Section 3.3, if \( d_{k,ts} [\text{diag}(\hat{p}_{k,ts}) - \hat{p}_{k,ts} \hat{p}_{k,ts}'] \) is conditionally unbiased for the conditional variance of \( e_k \) given \( u_k \) for a constant \( d_{k,ts} \) not depending on \( u_k \), then the unconditional variance is \( \Sigma_{ee,k,ts} = d_{k,ts}(1 - \psi) \Gamma_{uu,k,ts} \). In Figure 7.2, the vector halves of the direct estimates of the sampling variances are plotted against the vector halves of \( n_{k,ts}^{-1} \hat{\Gamma}_{uu,k,ts} \), where \( \hat{\Gamma}_{uu,k,ts} = \text{diag}(\hat{p}_{T,k,ts}) - \hat{p}_{T,k,ts} \hat{p}_{T,k,ts}' \). The elements of the direct estimators of the sampling variances and covariances vary more around \( n_{k,ts}^{-1} \hat{\Gamma}_{uu,k,ts} \) than around the simple random sampling covariance matrix evaluated at the direct estimates of the proportions. This is not surprising because the error in the estimator of \( p_{T,ik,ts} \) is less correlated with the error in the direct estimator of the sampling covariance matrix than the error in \( \hat{p}_{ik,ts} \). The values of \( R^2 \) corresponding to the plots in Figure 7.2 range from 0.779 in Nova Scotia to 0.983 in Quebec.
Figure 7.1 Direct estimates of sampling variances and covariances (y-axis); corresponding elements of \( \text{diag}(\hat{P}_{k,ts} - \hat{P}_{k,ts} \hat{P}_{k,ts}' n_{k,ts}^{-1} \) \( (t = 1, s = 1, 5) \) for A1 and E0 (x-axis). A1: closed circles; E0: open triangles
The decision of whether or not to use the working model for the covariance matrix involves a comparison of full and reduced models. We consider a single two digit code and suppress the subscript $ts$. The model underlying the direct estimate of the sampling variance is an unstructured covariance matrix with the property that rows and columns sum to zero. The number of parameters to estimate in the full model underlying the direct estimator is $0.5C(C - 1)$, where $C$ is the number of three digit codes in the two digit code. For any particular two digit code, the working model for the sampling covariance matrix is a reduced model where $\Sigma_{ee,k} = c_k n_k^{-1} \Gamma_{uu,k}$. Under an assumption that $\hat{\Sigma}_{ee,k}$ is unbiased for $\Sigma_{ee,k}$, an
equivalent way to express the null hypothesis is \( H_0 : E[\hat{\Sigma}_{ee,k}] = E[d_k n_k^{-1}(\text{diag}(\hat{p}_k) - \hat{p}_k \hat{p}_k')] \) for a constant \( d_k \). For a known \( \Gamma_{uu,k} \), the reduced model has one unknown parameter, \( c_k \).

If the direct estimator of the sampling covariance matrix has a Wishart distribution with \( \nu_k \) degrees of freedom, and if \( \Gamma_{uu,k} \) is known, then the log likelihood ratio statistic for testing the null hypothesis, \( H_0 : \Sigma_{ee,k} = \sigma_k \Gamma_{uu,k} \) for some \( \sigma_k \) (after multiplying by -2) is

\[
T_k = \nu_k \left[ p \log(\text{trace}(\hat{\Sigma}_{ee,k}^{(1)} [\Gamma_{uu,k}^{(1)}]^{-1}/p)) - \log(\text{det}(\hat{\Sigma}_{ee,k}^{(1)} [\Gamma_{uu,k}^{(1)}]^{-1})) \right],
\]

where \( p = C - 1 \), and \( \Gamma_{uu,k}^{(1)} \) is the inverse of \( (C - 1) \times (C - 1) \) submatrix of \( \Gamma_{uu,k} \) obtained by omitting the first row and column from \( \Gamma_{uu,k} \). The test statistic is invariant to which row and column are removed. Under the null hypothesis that \( \Sigma_{ee,k} \) is proportional to \( \Gamma_{uu,k} \) and the assumption that the direct estimator has a Wishart distribution, \( T_k \) has an approximate chi-squared distribution with \( 0.5C(C - 1) - 1 \) degrees of freedom. The log likelihood ratio test statistic does not penalize the full model for requiring that a relatively large number of parameters be estimated. Two criteria, also based on the Wishart model, that penalize larger models are AIC and BIC. The difference between the AIC of the multinomial model and the AIC for the full (unstructured) model is

\[
T_k + 2 - C(C - 1).
\]

(7.2)

Treating \( \nu_k + 1 \) as a sample size, the difference between the BIC of the reduced model and the BIC of the full model is

\[
T_k + \log(\nu_k + 1)(1 - 0.5C(C - 1)).
\]

(7.3)

Computation of the likelihood ratio test statistic, the AIC, and the BIC requires an estimate of \( p_{T,k} \) and a value for the degrees of freedom, \( \nu_k \). We replace the unknown \( p_{T,k} \) in \( \Gamma_{uu,k} \) with the SPREE estimator and consider three different possibilities for the degrees of freedom. One choice for the degrees of freedom is \( \nu_k = n_k - 1 \), where \( n_k - 1 \) is the sample size in the two digit code for province \( k \). The realized sample sizes for A1 and E0 are in Table 7.5. A second choice is \( \nu_k = n_k (\hat{c}_k)^{-1} - 1 \), where \( \hat{c}_k \) is the estimate of \( c_k \). The \( \{\hat{c}_k : k = 1, \ldots, 10\} \) used to compute the test statistics are given in the first row of Table 7.6. A third option,
$\nu_k = n_k^{0.5}$, is motivated by Maples and Bell (2009), in which a simulation with micro-data from the Current Population Survey conducted by the US Census Bureau indicates that the degrees of freedom associated with a replication variance estimator is proportional to the square root of the realized sample size. We expect that an appropriate value for the degrees of freedom is between $n_k^{0.5}$ and $n_k - 1$.

Many of the assumptions that justify the asymptotic chi-squared distribution of the log likelihood ratio test statistic and the validity of AIC and BIC for model selection are violated in this example. The Wishart model for the direct estimator of the sampling variance may be a poor approximation for the true distribution. Also, the fixed matrix $\Gamma_{uu,k}$ is unknown and estimated. Misspecification of the model for the mean underlying the estimator $\hat{p}_{T,k}$ affects the test statistics and could lead us to incorrectly reject the null hypothesis that $\Sigma_{ee,k} = \sigma_k \Gamma_{uu,k}$ for some constant $\sigma_k$. We use p-values based on the chi-squared distribution for guidance in identifying areas of potential lack of fit. We also compare the magnitudes of the log likelihood ratio test statistics, AIC and BIC to evaluate how the discrepancy between the multinomial model and the unstructured model underlying the direct estimator of the covariance matrix varies across the domains.

Table 7.2 contains the test statistics $\{T_k : k = 1, \ldots, 10\}$ calculated with degrees of freedom $\nu_k = n_k - 1$, $\nu_k = n_k (\hat{c}_k)^{-1} - 1$, and $\nu_k = n_k^{0.5}$ for the two digit codes A1 and E0. The two digit code A1 has four categories ($C = 4$), and the two digit code E0 has three categories ($C = 3$). The 95th percentile of the $\chi^2$ reference distribution for $C = 4$ is 11.07, and the 95th percentile for $C = 3$ is 5.99. When the degrees of freedom is $n_k - 1$ or $n_k (\hat{c}_k)^{-1} - 1$, the test statistics for A1 in Nova Scotia, Saskatchewan, Manitoba, British Columbia, and Ontario exceed 11.07. For E0, the test statistics based on degrees of freedom $n_k - 1$ exceed 5.99 in Quebec, Alberta, and Manitoba. The test statistics for E0 calculated with degrees of freedom $n_k (\hat{c}_k)^{-1} - 1$ exceed 5.99 in Quebec and Manitoba. All of the test statistics calculated with the degrees of freedom $n_k^{0.5}$ are smaller than the 95th percentiles of the respective $\chi^2$ distributions.

Table 7.3 shows the difference between the AIC for the working model and the AIC for the full model, where the difference in AIC's is calculated as in (7.2). The AIC for the
working model is smaller than the AIC for the full model if the test statistic $T_k$ is smaller than $2(0.5C(C - 1) - 1)$. If $C = 4$, then $2(0.5C(C - 1) - 1) = 10$, and if $C = 3$, then $2(0.5C(C - 1) - 1) = 4$. Because the $95^{th}$ percentiles of the chi-square distributions are 11.07 and 5.99 for $C = 4$ and $C = 3$, respectively, using AIC for model selection is similar to using the log likelihood ratio test statistic with the chi-squared reference distribution. In the two digit code A1, the AIC's for the working model exceed the AIC's for the full model in the provinces where the log likelihood ratio test statistics exceed 11.07. The same correspondence between the log likelihood ratio test statistic and the difference in AIC's also holds for most cases in E0. An exception occurs in Alberta, where the log likelihood ratio test statistic calculated with degrees of freedom $n_k\hat{c}_k - 1$ is smaller than 5.99, but the AIC for the reduced model is larger than the AIC for the full model.

The BIC criterion penalizes models for additional parameters more than the AIC criterion. Table 7.4 shows the differences between the BIC’s for the working model and the BIC’s for the full unstructured model. In A1, the BIC for the working model is larger than the BIC for the full model in British Columbia, Manitoba, and Nova Scotia when the degrees of freedom is $n_k - 1$. In E0, the BIC for the working model based on $n_k - 1$ degrees of freedom exceeds the BIC for the full model in Quebec. When the degrees of freedom is $n_k\hat{c}_k - 1$, the BIC for the reduced model exceeds the BIC for the full model in Nova Scotia for A1. The remaining BIC’s for the reduced model are smaller than the corresponding BIC’s for the full model.

The likelihood ratio criteria, AIC and BIC from the Wishart model with degrees of freedom $n_k - 1$ and $n_k\hat{c}_k - 1$ provide evidence of lack of fit of the working model for the sampling variances in many of the domains. The criteria suggest that the working model fits better in E0 than in A1. In both E0 and A1, the working model fits relatively poorly in Manitoba. The analyses of Sections 7.5 and 7.6 show that the departures between the working covariance model and the direct estimators of the sampling variances are greatest in provinces where an estimate of $\hat{p}_{T,ik}$ differs from the corresponding direct estimator of the proportion by more than two estimated standard deviations. (See Tables 7.17 and 7.29). For example, the likelihood ratio, AIC and BIC criteria as well as figures 7.1 and 7.2 reveal evidence
of lack of fit in Nova Scotia for A1 and in Manitoba for A1 and E0. In Nova Scotia, the direct estimate of the proportion in A11 is 3.12 estimated standard deviations smaller than the corresponding synthetic estimate, and the the direct estimate of the proportion in A14 is 2.07 standard deviations larger than the corresponding direct estimate of the proportion. Similarly, the direct estimate of the proportion in A12 in Manitoba is 2.18 standard deviations smaller than the corresponding synthetic estimate. In E0, the direct estimate of the proportion in Manitoba is 2.15 standard deviations larger than the corresponding synthetic estimate.

In the simulations, we observed that predictors constructed with the direct estimator of the sampling variance are often less efficient than predictors constructed under the working covariance model when $E[(\hat{\sigma}_{e,ik}^2 - \sigma_{e,ik}^2)^2] > E[(\hat{c}_k n_k^{-1} \hat{p}_{T,ik} (1 - \hat{p}_{T,ik}) - \sigma_{e,ik}^2)^2]$. Because the LFS uses complex sample selection and estimation procedures, we do not expect an estimator based on the working model to be an unbiased estimator of the covariance matrix of the sampling errors. Nonetheless, the simulation results suggest that we might achieve a gain in efficiency for the predictors by taking advantage of the relationship between the estimated means and the estimated sampling variances exhibited in Figures 7.1 and 7.2. As a consequence, we use the estimates of the sampling variances constructed under the working model to compute the predictors. To protect against bias due to misspecification of the working covariance model, we use the modified direct estimators of the sampling covariance matrix (modified to eliminate zero estimates) defined in (4.6) and (4.7) to estimate the MSE.
Table 7.2 Likelihood ratio test statistics defined in (7.1) for A1 and E0

<table>
<thead>
<tr>
<th>$\nu_k$</th>
<th>PE</th>
<th>NF</th>
<th>NB</th>
<th>NS</th>
<th>SK</th>
<th>MB</th>
<th>BC</th>
<th>AB</th>
<th>QC</th>
<th>ON</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n_k - 1$</td>
<td>6.33</td>
<td>7.35</td>
<td>2.88</td>
<td>23.03</td>
<td>14.94</td>
<td>22.42</td>
<td>25.57</td>
<td>8.64</td>
<td>3.18</td>
<td>19.81</td>
</tr>
<tr>
<td>$n_k \hat{c}_k^{-1} - 1$</td>
<td>4.73</td>
<td>5.59</td>
<td>2.45</td>
<td>20.03</td>
<td>13.37</td>
<td>18.09</td>
<td>17.74</td>
<td>5.62</td>
<td>1.90</td>
<td>13.21</td>
</tr>
<tr>
<td>$\sqrt{n_k}$</td>
<td>1.69</td>
<td>1.73</td>
<td>0.51</td>
<td>3.64</td>
<td>2.07</td>
<td>2.80</td>
<td>2.52</td>
<td>0.81</td>
<td>0.25</td>
<td>1.04</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\nu_k$</th>
<th>PE</th>
<th>NF</th>
<th>NB</th>
<th>NS</th>
<th>SK</th>
<th>MB</th>
<th>BC</th>
<th>AB</th>
<th>QC</th>
<th>ON</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n_k - 1$</td>
<td>2.70</td>
<td>0.16</td>
<td>3.46</td>
<td>2.73</td>
<td>0.18</td>
<td>7.89</td>
<td>1.83</td>
<td>6.70</td>
<td>16.44</td>
<td>3.49</td>
</tr>
<tr>
<td>$n_k \hat{c}_k^{-1} - 1$</td>
<td>2.04</td>
<td>0.13</td>
<td>2.95</td>
<td>2.38</td>
<td>0.16</td>
<td>6.38</td>
<td>1.27</td>
<td>4.36</td>
<td>9.81</td>
<td>2.33</td>
</tr>
<tr>
<td>$\sqrt{n_k}$</td>
<td>0.48</td>
<td>0.02</td>
<td>0.45</td>
<td>0.34</td>
<td>0.02</td>
<td>0.84</td>
<td>0.16</td>
<td>0.63</td>
<td>1.13</td>
<td>0.18</td>
</tr>
</tbody>
</table>

Table 7.3 Differences between AIC's for multinomial model and AIC's for full model, as defined in (7.2)
A1; C=4

<table>
<thead>
<tr>
<th>$\nu_k$</th>
<th>PE</th>
<th>NF</th>
<th>NB</th>
<th>NS</th>
<th>SK</th>
<th>MB</th>
<th>BC</th>
<th>AB</th>
<th>QC</th>
<th>ON</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n_k - 1$</td>
<td>-7.53</td>
<td>-7.63</td>
<td>-14.75</td>
<td>4.34</td>
<td>-5.01</td>
<td>1.47</td>
<td>2.30</td>
<td>-15.12</td>
<td>-22.38</td>
<td>-9.73</td>
</tr>
<tr>
<td>$n_k^{-1}c_k - 1$</td>
<td>-7.78</td>
<td>-8.10</td>
<td>-14.40</td>
<td>2.02</td>
<td>-6.03</td>
<td>-1.80</td>
<td>-3.73</td>
<td>-16.02</td>
<td>-21.10</td>
<td>-14.31</td>
</tr>
</tbody>
</table>

E0; C=3

<table>
<thead>
<tr>
<th>$\nu_k$</th>
<th>PE</th>
<th>NF</th>
<th>NB</th>
<th>NS</th>
<th>SK</th>
<th>MB</th>
<th>BC</th>
<th>AB</th>
<th>QC</th>
<th>ON</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n_k - 1$</td>
<td>-4.36</td>
<td>-7.70</td>
<td>-4.72</td>
<td>-5.65</td>
<td>-8.73</td>
<td>-1.13</td>
<td>-8.00</td>
<td>-2.81</td>
<td>5.72</td>
<td>-8.42</td>
</tr>
<tr>
<td>$n_k^{-1}c_k - 1$</td>
<td>-4.47</td>
<td>-7.22</td>
<td>-4.92</td>
<td>-5.73</td>
<td>-8.53</td>
<td>-2.22</td>
<td>-7.83</td>
<td>-4.30</td>
<td>0.12</td>
<td>-8.77</td>
</tr>
<tr>
<td>$\sqrt{n_k}$</td>
<td>-3.37</td>
<td>-4.17</td>
<td>-3.88</td>
<td>-4.08</td>
<td>-4.64</td>
<td>-3.87</td>
<td>-4.92</td>
<td>-4.30</td>
<td>-4.36</td>
<td>-5.88</td>
</tr>
</tbody>
</table>

Table 7.4 Differences between BIC’s for multinomial model and BIC’s for full model, as defined in (7.3)

<table>
<thead>
<tr>
<th>2-digit</th>
<th>PE</th>
<th>NF</th>
<th>NB</th>
<th>NS</th>
<th>SK</th>
<th>MB</th>
<th>BC</th>
<th>AB</th>
<th>QC</th>
<th>ON</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>16</td>
<td>20</td>
<td>34</td>
<td>42</td>
<td>54</td>
<td>66</td>
<td>105</td>
<td>116</td>
<td>166</td>
<td>368</td>
</tr>
<tr>
<td>E0</td>
<td>34</td>
<td>51</td>
<td>60</td>
<td>66</td>
<td>86</td>
<td>91</td>
<td>136</td>
<td>116</td>
<td>213</td>
<td>386</td>
</tr>
</tbody>
</table>

Table 7.5 Realized province sample sizes $\{n_k : k = 1, \ldots, 10\}$ for A1 and E0

The multinomial model for the sampling variances provides estimators of the sampling variances with smaller variances than the direct estimators of the sampling variances. An assumption that $c_{k,ts}$, the parameter for the working model in two digit code $t$ and one digit code $s$, is constant across groups of two digit codes permits construction of a less variable estimator. One possibility is to assume that $c_{k,ts} = c_k$ for all two digit codes in province $k$. An alternative is to construct an estimator under an assumption that $c_{k,ts} = c_{ts}$ and combine estimators across the provinces in a particular two digit code.
Figure 7.3 has plots of the estimates of \( \{c_{k,ts} : k = 1, \ldots, 10; t = 1, \ldots, D_s; s = 1, \ldots, 10\} \) for the 31 two digit codes in the one digit codes A-J. The different plotting symbols distinguish the different two digit codes. Each line connects points corresponding to a single two digit code. Qualitatively, the estimates of \( \{c_{k,ts} : k = 1, \ldots, 10; t = 1, \ldots, D_s; s = 1, \ldots, 10\} \) vary more in small provinces such as Prince Edward Island and New Foundland than in large provinces such as Ontario and Quebec. To compare the variability in the estimates of \( \{c_{k,ts} : k = 1, \ldots, 10; t = 1, \ldots, D_s; s = 1, \ldots, 10\} \) across provinces to the variability across two digit codes more formally, a linear model with indicators for two digit codes and indicators for provinces was fit to the estimates \( \{\hat{c}_{k,ts} : k = 1, \ldots, 10; t = 1, \ldots, D_s; s = 1, \ldots, 10\} \) using an assumption of a constant variance. The F-test for the province indicators is significant at the 0.001 level (F=7.7, p<0.001), while the F-test for the two digit code indicators is 0.95 leading to a p-value of 0.54. The pattern in Figure 7.3 and the results of the ANOVA suggest that an assumption of a common \( c_k \) for all two digit codes in province \( k \) is more consistent with the data than an assumption of a constant \( c_{ts} \) for all provinces in a particular two digit code. The LFS design also motivates use of different constants for each province because the provincial samples are independent. The LFS sample design does not provide a reason to use different constants for different two digit codes because the occupational classification is not part of the LFS design. As a consequence, the thirty-one two digit codes in the one digit codes A through J are used obtain a common estimate of the constant \( c_k \) for province \( k \).
Figure 7.3 Estimates of \( \{c_{k,ts} : k = 1, \ldots, 10; t = 1, \ldots, D_s; s = 1, \ldots, 10\} \) calculated separately for the 31 two digit codes in the one digit codes A-J.

The estimates, \( \{\hat{c}_k : k = 1, \ldots, K\} \), are constructed as in (4.4) with the initial estimates of \( \{\Gamma_{uu,k} : k = 1, \ldots, K\} \) based on \( \hat{p}_{T,ik}^0 \) in place of the true values. In the notation of Section 4.1.1, \(|A_c| = 31\). The first row of Table 7.6 has estimates \( \{\hat{c}_k : k = 1, \ldots, K\} \) based on the SPREE estimator of \( p_{T,ik} \). The SPREE estimator of \( p_{T,ik} \) is the initial estimator if the estimates of the coefficients on the Census interactions are restricted to equal 1.

Two estimators of the variance of \( \hat{c}_k \) are the “Taylor” estimator defined in (4.30) and the variance of the B bootstrap estimates. The assumption underlying the Taylor estimator of the variance of \( \hat{c}_k \) defined in (4.30) is

\[
\hat{c}_{k,ts} \sim \left( c_k, \frac{\sigma_{\hat{c}_k}^2}{(C_{ts} - 1)(n_{k,ts} - 1)} \right).
\] (7.4)
To check the assumption underlying the Taylor estimator of the variance of $\hat{c}_k$, we examine plots with standardized residuals,

$$\hat{r}_{c,k,ts} = [(C_{ts} - 1)(n_{k,ts} - 1)\hat{\sigma}_{c,k}^{-2}]^{0.5}(\hat{c}_{k,ts} - \hat{c}_k)$$

(7.5)
on the vertical axis and province sample sizes on the horizontal axis (Figure 7.4), where $\hat{\sigma}_{c,k}^{-2}$ is defined in (4.30). The residual for two digit code H7 in Nova Scotia is larger than 4.0, the residual for I1 in Manitoba is 3.9, and the residual for C0 in Prince Edward Island is 3.02. Otherwise, the residuals are smaller than 3 in absolute value. The variances of the residuals do not exhibit obvious trends as functions of the province sample sizes, and the residuals are roughly symmetric around zero. Other than the outliers, the plots in Figure 7.4 do not provide immediate reasons to reject the model (7.4). The bootstrap estimator of the variance of $\hat{c}_k$ is based on an assumption that the direct estimators of the covariance matrices have Wishart distributions and are independent of the direct estimators of the means. The plots in Figures 7.1 and 7.2 suggest a linear relationship between the direct estimators of the sampling variances and the direct estimators of the means. Because we suspect that the assumptions of the bootstrap estimator of the variance of $\hat{c}_k$ do not hold, we prefer the Taylor estimators of the standard errors of the estimators of $\hat{c}_k \ (k = 1, \ldots, K)$. The Taylor standard errors, defined in (4.30), are given in the last row of Table 7.6.
Figure 7.4 Standardized residuals defined in (7.4) on the vertical axis and province sample sizes on the horizontal axis

<table>
<thead>
<tr>
<th>PE</th>
<th>NF</th>
<th>NB</th>
<th>NS</th>
<th>SK</th>
<th>MB</th>
<th>BC</th>
<th>AB</th>
<th>QC</th>
<th>ON</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{c}_k$</td>
<td>1.311</td>
<td>1.293</td>
<td>1.169</td>
<td>1.145</td>
<td>1.115</td>
<td>1.235</td>
<td>1.436</td>
<td>1.530</td>
<td>1.670</td>
</tr>
<tr>
<td>SE</td>
<td>0.063</td>
<td>0.064</td>
<td>0.031</td>
<td>0.057</td>
<td>0.045</td>
<td>0.050</td>
<td>0.038</td>
<td>0.061</td>
<td>0.021</td>
</tr>
</tbody>
</table>

Table 7.6 Estimates of $c_k : k = 1, \ldots, 10$ and Taylor standard errors defined in (4.30)

Figures 7.5 and 7.6 show standardized residuals comparing the direct estimators of the variances to the estimators under the working model for A1 and E0, respectively. The
standardized residual for category $i$ and province $k$ is

$$
\hat{r}_{ik,e} = \frac{\hat{\sigma}_{e,ik}^2 - \hat{p}_{T,ik}(1 - \hat{p}_{T,ik})\hat{c}_k n_k^{-1}}{\sqrt{2(\hat{p}_{T,ik}(1 - \hat{p}_{T,ik})\hat{c}_k n_k^{-1})^2 n_k^{-1}}}.
$$

(7.6)

The standardization used to define $\hat{r}_{ik,e}$ is based on the fourth moments of a normal distribution. The labels in the plots indicate the province and three digit category for each two digit code. For example, the point labeled “MB2” in Figure 7.5 corresponds to three digit code A12 in Manitoba. In general, the residuals are symmetric around zero. In the simulations, the estimate of $c_k$ has a negative bias in small provinces. It is possible that the estimate of $c_k$ in Prince Edward Island is too small. However, the residuals for Prince Edward Island are negative in both A1 and E0. Many residuals with relatively large absolute values are in provinces containing cells for which the direct estimates differ from the synthetic estimates by more than two estimated standard deviations (e.g., Nova Scotia in A1 and Manitoba in A1 and E0).
Figure 7.5  Standardized residuals for evaluating goodness of fit of working model for sampling variances for A1. Residuals defined in (7.6) on y-axis; working model variance $\hat{\sigma}_k n_k^{-1} \hat{\sigma}_T,ik (1 - \hat{\sigma}_T,ik)$ on x-axis.
Figure 7.6 Standardized residuals for evaluating goodness of fit of working model for sampling variances for E0. Residuals defined in (7.6) on y-axis; working model variance $\hat{c}_k n_k^{-1} \hat{p}_T,ik (1 - \hat{p}_T,ik)$ on x-axis.

The preceding analysis deals with estimation of the sampling variances in the scale of the proportions. The procedure also requires estimators of the sampling variances in the scale of totals. In particular, in step 3 of the bootstrap data generating procedure described in Section 5.2.1, the bootstrap version of the direct estimator of the variance of the direct estimator of the province two digit total is $\hat{c}_k (\hat{M}_{k,ts}^{*(b)})^2 n_{k,ts}^{-1}$. The plots in Figure 7.7 and Figure 7.8 help justify the decision to use $\hat{c}_k (\hat{M}_{k,ts}^{*(b)})^2 n_{k,ts}^{-1}$ as the direct estimator of the variance of the direct estimator of the province two digit total in the bootstrap data generating algorithm. In Figure 7.7, the direct estimator of the variance of the direct estimator of the province two
digit total is plotted against $\hat{\mathcal{c}}_k \overline{\hat{\mathcal{M}}}_{k,ts}^2 n_{k,ts}^{-1}$ for the two digit codes in the one digit codes A and E. Figure 7.8 suggests that the estimators of the variances of the direct estimators of the province two digit totals are linearly related to $\{\hat{\mathcal{c}}_k n_{k,ts}^{-1} \overline{\hat{\mathcal{M}}}_{k,ts} : t = 1, \ldots, D_s ; s = 1, 5\}$. To evaluate the linear relationship more formally, the $\{\hat{\mathcal{V}}\{\overline{\hat{\mathcal{M}}}_{k,ts}\} : k = 1, \ldots, K ; t = 1, \ldots, D_s, s = 1, 5\}$ were regressed on $\hat{\mathcal{c}}_k \overline{\hat{\mathcal{M}}}_{k,ts}^2 n_{k,ts}^{-1}$ with no intercept and weights $(\hat{\mathcal{c}}_k \overline{\hat{\mathcal{M}}}_{k,ts}^2 n_{k,ts}^{-1})^{-2}$, where $\hat{\mathcal{V}}\{\overline{\hat{\mathcal{M}}}_{k,ts}\}$ is the direct estimator of the variance of the direct estimator of the province two digit total. The estimated slope from the regression is 0.955, and a standard error of the estimate under an assumption that the variance of the direct estimators of the variances of the direct estimator of the total in two digit code $ts$ and province $k$ is proportional to $(\hat{\mathcal{c}}_k \overline{\hat{\mathcal{M}}}_{k,ts}^2 n_{k,ts}^{-1})^2$ is 0.016. Figure 7.8 shows the standardized residuals from the regression. (The residuals are standardized by $0.11^{-1}(\hat{\mathcal{c}}_k \overline{\hat{\mathcal{M}}}_{k,ts}^2 n_{k,ts}^{-1})^{-1}$ so that the sum of squared standardized residuals is 1.) The standardized residuals are plotted with $\log(\hat{\mathcal{c}}_k \overline{\hat{\mathcal{M}}}_{k,ts}^2 n_{k,ts}^{-1})$ on the horizontal axis. The residuals are approximately symmetric around zero. The relationship between the estimated means and the estimated variances provides partial justification for the method used to generate the bootstrap versions of the direct estimators of the variances of the direct estimators of the marginal province two digit totals.
Figure 7.7  Direct estimators of variances of direct estimators of province 2-digit totals (y-axis), and squares of direct estimators of province 2-digit totals multiplied by $\hat{c}_k n_{k,ts}^{-1}$ (x-axis).
Figure 7.8  Residuals from weighted regression of direct estimators of variances on $\hat{c}_k \hat{M}_{k,t,s}^2 n_{k,t,s}^{-1}$. A1: open circle, A2: triangle, A3: plus, E0: $\times$, E1: diamond.

7.5 Estimates and Predictions for A1

The Census proportions and direct estimates for A1 (specialist managers) are in Tables 7.7 and 7.8, respectively. In the right plot of Figure 7.9, the LFS estimates of the proportions are plotted on the $y-$axis with the corresponding Census proportions on the $x-$axis. The different numbers used as the plotting symbols distinguish four different three digit codes in the two digit code A1. The three digit fractions in A1 fall into two groups; in both the LFS and the Census, the fractions in the three digit codes A12 and A14 are typically smaller than 0.20, while most of the proportions in the three digit codes A11 and A13 are larger than 0.28. Exceptions occur in Nova Scotia and Saskatchewan. In Nova Scotia, the LFS proportion for
three digit A11 is 0.137, and the fraction for three digit code A14 is 0.294. In Saskatchewan, both the LFS and Census totals in three digit code A14 exceed 0.20. In the Census, all of the proportions are greater than 0.10. The LFS estimates of the proportions in A14 in Prince Edward Island, A12 in New Foundland, and A12 in Manitoba are below 0.05.

The model (3.1) specifies the logits of the expected values of the direct estimators to be linearly related to the Census interactions, with a different intercept for the different categories. As an exploratory step, we plot the nonzero Census interactions on the x-axis with the corresponding logits of the LFS proportions on the y-axis (left plot in Figure 1). Qualitatively, the logits in three digit code A14 have a positive association with the corresponding Census interactions, while the associations for the other three digit codes are less clear. Only two of the Census interactions exceed 0.5 in absolute value. Because the Census interactions are small, the synthetic estimates are close to the national fractions regardless of the slopes assigned to the Census interactions.

The province two digit totals from the Census and the LFS are in the last row of Table 7.7 and the second to last row of Table 7.8, respectively. The estimated CV's of the LFS estimators of the province two digit totals are in the last row of Table 7.8. The province two digit totals observed in the LFS are uniformly smaller than the province two digit totals in the Census. Because the model and predictors are specified in the scale of proportions, the marginal level in the Census does not influence the model based predictors. Use of the multinomial model to construct the initial predictors ensures that the initial predictors preserve the LFS estimates of the province two digit totals. The raking operation ensures that the predictors preserve the LFS estimates of the national three digit totals.
Figure 7.9  Direct estimates and Census proportions for A1. Left plot: logits of direct estimators of proportions (y-axis), Census interactions (x-axis). Right plot: LFS proportions (y-axis), Census proportions (x-axis). Numbers correspond to three digit codes, \( i = 1, \ldots, 4 \).
<table>
<thead>
<tr>
<th></th>
<th>PE</th>
<th>NF</th>
<th>NB</th>
<th>NS</th>
<th>SK</th>
<th>MB</th>
<th>BC</th>
<th>AB</th>
<th>QC</th>
<th>ON</th>
</tr>
</thead>
<tbody>
<tr>
<td>A11</td>
<td>0.409</td>
<td>0.373</td>
<td>0.334</td>
<td>0.362</td>
<td>0.353</td>
<td>0.385</td>
<td>0.325</td>
<td>0.348</td>
<td>0.378</td>
<td>0.327</td>
</tr>
<tr>
<td>A12</td>
<td>0.172</td>
<td>0.138</td>
<td>0.157</td>
<td>0.148</td>
<td>0.094</td>
<td>0.118</td>
<td>0.142</td>
<td>0.165</td>
<td>0.185</td>
<td>0.187</td>
</tr>
<tr>
<td>A13</td>
<td>0.288</td>
<td>0.281</td>
<td>0.334</td>
<td>0.348</td>
<td>0.334</td>
<td>0.332</td>
<td>0.377</td>
<td>0.323</td>
<td>0.364</td>
<td>0.362</td>
</tr>
<tr>
<td>A14</td>
<td>0.131</td>
<td>0.208</td>
<td>0.175</td>
<td>0.141</td>
<td>0.220</td>
<td>0.165</td>
<td>0.156</td>
<td>0.164</td>
<td>0.073</td>
<td>0.124</td>
</tr>
</tbody>
</table>

\[ N_k \] 990 3700 6130 8630 7880 10720 47680 40270 80230 187710

Table 7.7 Census 2001 proportions and province two digit totals for A1
(specialist managers)

<table>
<thead>
<tr>
<th></th>
<th>PE</th>
<th>NF</th>
<th>NB</th>
<th>NS</th>
<th>SK</th>
<th>MB</th>
<th>BC</th>
<th>AB</th>
<th>QC</th>
<th>ON</th>
</tr>
</thead>
<tbody>
<tr>
<td>A11</td>
<td>0.390</td>
<td>0.287</td>
<td>0.298</td>
<td>0.137</td>
<td>0.201</td>
<td>0.376</td>
<td>0.369</td>
<td>0.297</td>
<td>0.388</td>
<td>0.365</td>
</tr>
<tr>
<td>A12</td>
<td>0.143</td>
<td>0.047</td>
<td>0.134</td>
<td>0.099</td>
<td>0.181</td>
<td>0.035</td>
<td>0.225</td>
<td>0.231</td>
<td>0.186</td>
<td>0.119</td>
</tr>
<tr>
<td>A13</td>
<td>0.422</td>
<td>0.499</td>
<td>0.474</td>
<td>0.470</td>
<td>0.368</td>
<td>0.390</td>
<td>0.322</td>
<td>0.334</td>
<td>0.363</td>
<td>0.386</td>
</tr>
<tr>
<td>A14</td>
<td>0.045</td>
<td>0.167</td>
<td>0.094</td>
<td>0.294</td>
<td>0.250</td>
<td>0.199</td>
<td>0.084</td>
<td>0.138</td>
<td>0.064</td>
<td>0.130</td>
</tr>
</tbody>
</table>

\[ \hat{M}_k \] 770 2243 4538 6091 5622 9220 35052 31323 72926 142835

CV of \[ \hat{M}_k \] 28% 24% 19% 17% 14% 13% 11% 11% 9% 7%

Table 7.8 May 2005 LFS proportions, province two digit totals (\[ \hat{M}_k \]),
and estimated coefficients of variation of the estimators of the
province two digit totals for A1

7.5.1 Model Estimates for A1

The two digit code A1 is one of three two digit codes with more than one three digit
code in the one digit code A. (In the notation of Section 7.2, \( D_s = 3 \).) As discussed in Section 7.2, we construct predictors under a reduced model in which the coefficient on the Census interaction is equal to 1, and \( \psi_{t1} = \psi_1 \) for each of the three two digit codes \( t \) in the one digit code A. Table 7.9 shows the estimates of the parameters of the reduced version of model (3.1)
for the two digit code A1. The standard errors of the estimators of the elements of $\lambda_o$ are the square roots of the diagonal elements of the covariance matrix in (A.8) of Appendix 2. The standard error of the estimator of $\psi_1$ is defined in (4.28). The estimate of $\psi_1$ is 2.69 times the corresponding standard error. The lower bound on the estimator of $\psi_1$ is $\xi_D = 0.0005$. The category effects are relative to category A11. In A1, the estimates of the category effects for three digit codes A12 and A24 differ significantly from the effect for three digit code A13. The category effect for three digit code A23 differs significantly from zero. The category effects are consistent with the pattern, evident in Figure 7.9, that the proportions in three digit codes A11 and A13 are typically larger than 0.28, while the proportions in three digit codes A12 and A14 are generally smaller than 0.20.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$\alpha_o,2$</th>
<th>$\alpha_o,3$</th>
<th>$\alpha_o,4$</th>
<th>$\alpha_o,2 - \alpha_o,4$</th>
<th>$\psi$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimate</td>
<td>-0.96</td>
<td>-0.10</td>
<td>-0.53</td>
<td>-0.43</td>
<td>0.0062</td>
</tr>
<tr>
<td>SE</td>
<td>0.15</td>
<td>0.11</td>
<td>0.15</td>
<td>0.18</td>
<td>0.0023</td>
</tr>
</tbody>
</table>

Table 7.9 Estimates of the parameters of the reduced model for A1 and Taylor standard errors. Coefficient on Census interactions, $\theta_o = 1$

To evaluate whether the assumption that $\theta_{o,t1} = 1$ is consistent with the data, the estimates of $\theta_{o,t1}$ ($t = 1, 2, 3$) are obtained for the three two digit codes in the one digit code A. The estimates of $\theta_{o,t1}$ ($t = 1, 2, 3$) and standard errors are given in Table 7.10. The standard error of the estimator of $\theta_{o,t1}$ is the square root of the diagonal element of $\hat{V}_{\lambda\lambda}$ defined in (A.8) of Appendix 2 corresponding to $\theta_{o,t1}$. For both A1 and A2, the estimates of $\theta_{o,t1}$ do not differ from either zero or one by more than two standard errors. The estimate of $\theta_{o,t1}$ for two digit code A3 differs from zero by more than two standard errors but differs from one by approximately one standard error. To test the hypothesis, $H_o : \theta_{o,t1} = 1$ for $t = 1, 2, 3$, we compute the test statistic,

$$W_\theta = \sum_{t=1}^{3} \frac{(\hat{\theta}_{o,t1} - 1)^2}{\hat{V}\{\hat{\theta}_{o,t1}\}}.$$
For the one digit code A, $W_\theta = 2.63$. Comparing the observed value of $W_\theta$ to a $\chi^2$ distribution with 3 degrees of freedom gives a p-value of 0.33. Therefore, we conclude that the data are consistent with the reduced model with $\theta_{o,t1} = 1$.

The effect of setting $\theta_{o,t1}$ equal to one is largely inconsequential in this example. As mentioned in the discussion of Table 7.7 and Figure 7.9, the Census table for A1 is close to an independence table, and the Census interactions for A1 are small. As a consequence, the parameter $\theta_{o,11}$ is poorly identified, and the synthetic estimators are close to the national fractions regardless of whether the estimate of $\theta_{o,11}$ is zero or 1. The differences between predictors calculated with $\hat{\theta}_{o,11} = 0.5$ and $\hat{\theta}_{o,11} = 1$ are negligible relative to the estimates of the MSE’s. Also, the estimates of the variances of the synthetic estimators of the proportions are highly correlated with the synthetic estimators of the proportions. As a result, the estimates of the MSE’s of predictors calculated with $\hat{\theta}_{o,11} = 0.5$ are similar to the MSE estimates obtained when $\hat{\theta}_{o,11}$ is restricted to equal 1.

The assumption that $\psi_{t1} = \psi_1$ for $t = 1, 2, 3$ allows us to obtain an estimator of $\psi_t$ with a smaller variance than the variance of an estimator constructed with any one of the two digit codes alone. Table 7.10 shows the separate estimates of $\psi_{t1}$ for $t = 1, 2, 3$ for the three two digit codes in A. The standard errors of $\hat{\psi}_{t1}$ ($t = 1, 2, 3$) are calculated as defined in (4.28). The estimated standard errors of the estimators of $\psi_{ts}$ in Table 7.10 are also large relative to the estimates of $\psi_t$.

To test the null hypothesis that $\psi_{t1} = \psi_1$ for $t = 1, 2, 3$, we compute the score statistic defined in (4.23) and obtain $S_\psi = 4.63$. Comparing $S_\psi$ to a $\chi^2$ distribution with two degrees of freedom gives a p-value of 0.099. A bootstrap test of the null hypothesis that the two digit codes A1, A2, and A3 share a common value of $\psi$ using the method described in Section 7.3 gives a p-value of 0.146. Neither test procedure provides strong evidence against the null hypothesis, and we do not reject the null hypothesis. We construct predictors under a model in which the value of $\psi_{ts}$ is assumed to be the same for the three two digit codes A1, A2, and A3.
Table 7.10 Estimates of $\theta_o$ and $\psi$ for A1, A2, and A3 with Taylor standard errors. The last column has the lower bounds for the estimators of $\psi$. The second column has the number of three digit codes in each two digit code.

<table>
<thead>
<tr>
<th>2-digit Code</th>
<th>$C$</th>
<th>$\hat{\theta}$</th>
<th>$[\hat{V}{\hat{\theta}}]^{0.5}$</th>
<th>$\hat{\psi}$</th>
<th>$[\hat{V}{\hat{\psi}}]^{0.5}$</th>
<th>$0.5[\hat{V}_0{\hat{\psi}}]^{0.5}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>4</td>
<td>0.54</td>
<td>0.39</td>
<td>0.0147</td>
<td>0.0089</td>
<td>0.0018</td>
</tr>
<tr>
<td>A2</td>
<td>2</td>
<td>0.62</td>
<td>1.12</td>
<td>0.0215</td>
<td>0.0155</td>
<td>0.0016</td>
</tr>
<tr>
<td>A3</td>
<td>10</td>
<td>0.82</td>
<td>0.17</td>
<td>0.0036</td>
<td>0.0021</td>
<td>0.0006</td>
</tr>
</tbody>
</table>

7.5.2 Predictions for A1

Table 7.12 contains the predictors of the proportions for A1 with the square roots of the bootstrap MSE estimators. The corresponding direct estimates and estimates of $p_{T,ik}$ are provided in Tables 7.11 and 7.13, respectively. Within each category, the direct estimates vary more across the provinces than either the predictors or the estimates of $p_{T,ik}$. In A11, the raking operation causes the predictors for Prince Edward Island, New Foundland, New Brunswick, Manitoba, and Alberta to exceed both the estimate of $p_{T,ik}$ and the direct estimate. In A13, the predictors in Prince Edward Island, Saskatchewan, Manitoba, and Ontario are below both the direct estimate and the estimate of $p_{T,ik}$. The predictors for the other cells are between the direct estimates and the estimates of $p_{T,ik}$.

In the two digit code A1, the direct estimates of the proportions for three digit codes A11 and A13 are closer to the national fractions for three digit codes A11 and A13 than to the national fractions for three digit codes A12 and A14 in all provinces except Nova Scotia. The national fractions for three digit codes A11, A12, A13, and A14 are 0.36, 0.16, 0.37, and 0.12, respectively. The direct estimate of the proportion for three digit code A11 in Nova Scotia (0.137) is closer to the national fractions for three digit codes A14 and A12 than to the national fraction for three digit code A11, and the direct estimate for three digit code A14 in Nova Scotia (0.294) is closer to the national fractions for three digit codes A11 and A13 than
to the national fraction for three digit code A14. Even though the order of the direct estimates for three digit codes A11 and A14 in Nova Scotia is opposite the order for the nation, the estimate of $p_{T,ik}$ for A11 is close to the national fraction for A11, and the estimate of $p_{T,ik}$ for A14 is close to the national fraction for A14 because the Census interactions are small. When forming the univariate predictors with the Beale estimator of $\gamma_k$, the weight assigned to the estimates of $p_{T,ik}$ in Nova Scotia is 0.811. Because the estimates of $p_{T,ik}$ are close to the national fractions, the predictor for A11 in Nova Scotia (0.311) is closer to the national fraction for A11 than to the national fraction for A12 or A14, and the predictor for A14 in Nova Scotia (0.164) is closer to the national fraction for A14 than to the national fraction for A11 or A13.
<table>
<thead>
<tr>
<th></th>
<th>PE</th>
<th>NF</th>
<th>NB</th>
<th>NS</th>
<th>SK</th>
<th>MB</th>
<th>BC</th>
<th>AB</th>
<th>QC</th>
<th>ON</th>
</tr>
</thead>
<tbody>
<tr>
<td>A11</td>
<td>0.390</td>
<td>0.287</td>
<td>0.298</td>
<td>0.137</td>
<td>0.201</td>
<td>0.376</td>
<td>0.369</td>
<td>0.297</td>
<td>0.388</td>
<td>0.365</td>
</tr>
<tr>
<td>(SE)</td>
<td>(0.126)</td>
<td>(0.126)</td>
<td>(0.080)</td>
<td>(0.052)</td>
<td>(0.055)</td>
<td>(0.064)</td>
<td>(0.053)</td>
<td>(0.051)</td>
<td>(0.046)</td>
<td>(0.033)</td>
</tr>
<tr>
<td>A12</td>
<td>0.143</td>
<td>0.047</td>
<td>0.134</td>
<td>0.099</td>
<td>0.181</td>
<td>0.035</td>
<td>0.225</td>
<td>0.231</td>
<td>0.186</td>
<td>0.119</td>
</tr>
<tr>
<td>(SE)</td>
<td>(0.094)</td>
<td>(0.048)</td>
<td>(0.056)</td>
<td>(0.050)</td>
<td>(0.053)</td>
<td>(0.026)</td>
<td>(0.047)</td>
<td>(0.047)</td>
<td>(0.038)</td>
<td>(0.021)</td>
</tr>
<tr>
<td>A13</td>
<td>0.422</td>
<td>0.499</td>
<td>0.474</td>
<td>0.470</td>
<td>0.368</td>
<td>0.390</td>
<td>0.322</td>
<td>0.334</td>
<td>0.363</td>
<td>0.386</td>
</tr>
<tr>
<td>(SE)</td>
<td>(0.104)</td>
<td>(0.120)</td>
<td>(0.091)</td>
<td>(0.072)</td>
<td>(0.066)</td>
<td>(0.061)</td>
<td>(0.051)</td>
<td>(0.055)</td>
<td>(0.048)</td>
<td>(0.032)</td>
</tr>
<tr>
<td>A14</td>
<td>0.045</td>
<td>0.167</td>
<td>0.094</td>
<td>0.294</td>
<td>0.250</td>
<td>0.199</td>
<td>0.084</td>
<td>0.138</td>
<td>0.064</td>
<td>0.130</td>
</tr>
<tr>
<td>(SE)</td>
<td>(0.047)</td>
<td>(0.086)</td>
<td>(0.054)</td>
<td>(0.074)</td>
<td>(0.063)</td>
<td>(0.055)</td>
<td>(0.031)</td>
<td>(0.038)</td>
<td>(0.023)</td>
<td>(0.021)</td>
</tr>
</tbody>
</table>

Table 7.11  Direct estimates of proportions and sampling standard deviations, A1

<table>
<thead>
<tr>
<th></th>
<th>PE</th>
<th>NF</th>
<th>NB</th>
<th>NS</th>
<th>SK</th>
<th>MB</th>
<th>BC</th>
<th>AB</th>
<th>QC</th>
<th>ON</th>
</tr>
</thead>
<tbody>
<tr>
<td>A11</td>
<td>0.398</td>
<td>0.358</td>
<td>0.322</td>
<td>0.312</td>
<td>0.311</td>
<td>0.378</td>
<td>0.336</td>
<td>0.328</td>
<td>0.379</td>
<td>0.356</td>
</tr>
<tr>
<td>(SE)</td>
<td>(0.048)</td>
<td>(0.046)</td>
<td>(0.043)</td>
<td>(0.042)</td>
<td>(0.039)</td>
<td>(0.043)</td>
<td>(0.038)</td>
<td>(0.038)</td>
<td>(0.035)</td>
<td>(0.030)</td>
</tr>
<tr>
<td>A12</td>
<td>0.165</td>
<td>0.126</td>
<td>0.149</td>
<td>0.135</td>
<td>0.112</td>
<td>0.093</td>
<td>0.165</td>
<td>0.183</td>
<td>0.181</td>
<td>0.142</td>
</tr>
<tr>
<td>(SE)</td>
<td>(0.036)</td>
<td>(0.031)</td>
<td>(0.033)</td>
<td>(0.030)</td>
<td>(0.025)</td>
<td>(0.025)</td>
<td>(0.029)</td>
<td>(0.030)</td>
<td>(0.029)</td>
<td>(0.021)</td>
</tr>
<tr>
<td>A13</td>
<td>0.318</td>
<td>0.320</td>
<td>0.374</td>
<td>0.390</td>
<td>0.359</td>
<td>0.362</td>
<td>0.373</td>
<td>0.340</td>
<td>0.374</td>
<td>0.379</td>
</tr>
<tr>
<td>(SE)</td>
<td>(0.044)</td>
<td>(0.043)</td>
<td>(0.043)</td>
<td>(0.041)</td>
<td>(0.041)</td>
<td>(0.039)</td>
<td>(0.038)</td>
<td>(0.036)</td>
<td>(0.028)</td>
<td></td>
</tr>
<tr>
<td>A14</td>
<td>0.119</td>
<td>0.196</td>
<td>0.154</td>
<td>0.163</td>
<td>0.217</td>
<td>0.166</td>
<td>0.126</td>
<td>0.149</td>
<td>0.066</td>
<td>0.123</td>
</tr>
<tr>
<td>(SE)</td>
<td>(0.031)</td>
<td>(0.039)</td>
<td>(0.033)</td>
<td>(0.032)</td>
<td>(0.038)</td>
<td>(0.033)</td>
<td>(0.026)</td>
<td>(0.029)</td>
<td>(0.018)</td>
<td>(0.019)</td>
</tr>
</tbody>
</table>

Table 7.12  Predictors of proportions and square roots of bootstrap MSE estimates, A1
Table 7.13 Estimates of \( \{ p_{T,ik} : i = 1, \ldots , 4; k = 1, \ldots , 10 \} \) and Taylor standard errors.

<table>
<thead>
<tr>
<th></th>
<th>PE</th>
<th>NF</th>
<th>NB</th>
<th>NS</th>
<th>SK</th>
<th>MB</th>
<th>BC</th>
<th>AB</th>
<th>QC</th>
<th>ON</th>
</tr>
</thead>
<tbody>
<tr>
<td>A11</td>
<td>0.382</td>
<td>0.348</td>
<td>0.309</td>
<td>0.335</td>
<td>0.326</td>
<td>0.357</td>
<td>0.298</td>
<td>0.322</td>
<td>0.349</td>
<td>0.301</td>
</tr>
<tr>
<td>(SE)</td>
<td>(0.024)</td>
<td>(0.023)</td>
<td>(0.022)</td>
<td>(0.023)</td>
<td>(0.023)</td>
<td>(0.021)</td>
<td>(0.022)</td>
<td>(0.023)</td>
<td>(0.023)</td>
<td>(0.021)</td>
</tr>
<tr>
<td>A12</td>
<td>0.166</td>
<td>0.133</td>
<td>0.151</td>
<td>0.142</td>
<td>0.090</td>
<td>0.113</td>
<td>0.135</td>
<td>0.158</td>
<td>0.177</td>
<td>0.179</td>
</tr>
<tr>
<td>(SE)</td>
<td>(0.019)</td>
<td>(0.016)</td>
<td>(0.017)</td>
<td>(0.016)</td>
<td>(0.011)</td>
<td>(0.013)</td>
<td>(0.016)</td>
<td>(0.018)</td>
<td>(0.020)</td>
<td>(0.020)</td>
</tr>
<tr>
<td>A13</td>
<td>0.322</td>
<td>0.314</td>
<td>0.369</td>
<td>0.385</td>
<td>0.370</td>
<td>0.369</td>
<td>0.415</td>
<td>0.359</td>
<td>0.403</td>
<td>0.399</td>
</tr>
<tr>
<td>(SE)</td>
<td>(0.021)</td>
<td>(0.021)</td>
<td>(0.023)</td>
<td>(0.023)</td>
<td>(0.023)</td>
<td>(0.023)</td>
<td>(0.024)</td>
<td>(0.023)</td>
<td>(0.024)</td>
<td>(0.024)</td>
</tr>
<tr>
<td>A14</td>
<td>0.130</td>
<td>0.205</td>
<td>0.171</td>
<td>0.138</td>
<td>0.214</td>
<td>0.162</td>
<td>0.151</td>
<td>0.161</td>
<td>0.0713</td>
<td>0.121</td>
</tr>
<tr>
<td>(SE)</td>
<td>(0.016)</td>
<td>(0.022)</td>
<td>(0.019)</td>
<td>(0.016)</td>
<td>(0.023)</td>
<td>(0.019)</td>
<td>(0.018)</td>
<td>(0.019)</td>
<td>(0.0091)</td>
<td>(0.015)</td>
</tr>
</tbody>
</table>

Figure 7.10 shows 95% prediction intervals calculated with the direct estimators (solid lines) and the predictors (dashed lines). Four sets of prediction intervals are plotted for each province (one set with two prediction intervals for each of the four three digit codes). An interval constructed with the direct estimator is \([ L_{ik,dir}, U_{ik,dir} ]\), where

\[
L_{ik,dir} = \max \{ \hat{p}_{ik} - 1.96 \hat{\sigma}_{e,ik,dir,md}, 0 \}, \tag{7.7}
\]

and

\[
U_{ik,dir} = \min \{ \hat{p}_{ik} + 1.96 \hat{\sigma}_{e,ik,dir,md}, 1 \}.
\]

An interval constructed with the predictor is \([ L_{ik,pred}, U_{ik,pred} ]\), where

\[
L_{ik,pred} = \max \{ \hat{p}_{ik,B} - t_{0.975,df} [\hat{MSE}_{ik}^{bs}]^{0.5}, 0 \}, \tag{7.8}
\]

\[
U_{ik,pred} = \min \{ \hat{p}_{ik,B} + t_{0.975,df} [\hat{MSE}_{ik}^{bs}]^{0.5}, 1 \},
\]

and \( t_{0.975,df} \) is the 97.5 percentile of a \( t \) distribution with 117 degrees of freedom. (The degrees of freedom is the sum of the degrees of freedom associated with each two digit code in the one digit code A: \( 9(3 + 1 + 9) = 117 \)).
Figure 7.10  Direct estimates and predictors of proportions in A1 with corresponding prediction intervals. Direct estimates: + and solid lines. Predictors: · and dashed lines.

The reduction in the MSE due to the prediction procedure is reflected in narrower
intervals in many of the provinces. The improvements are greatest in the small provinces. The lower endpoints of the prediction intervals based on the direct estimators are truncated at zero in six of the cells (A12 and A14 in Prince Edward Island and New Foundland, A14 in New Brunswick, and A22 in Manitoba). Truncation is not required for the prediction intervals calculated with the predictors and corresponding MSE estimators.

Previously, we observed that the proportions in categories A12 and A14 are generally smaller than the proportions in categories A11 and A13 in both the Census and the LFS. Nonetheless, many of the 95% intervals for differences of the form \( p_{ik} - p_{jk} \), where \( i = 1 \) or 2 and \( j = 1 \) or 4 constructed with the direct estimator of the covariance matrix contain zero. In most cases, the corresponding prediction intervals based on the bootstrap MSE estimator do not contain zero. For example, the direct estimate of the difference between the proportions in categories A11 and A12 in Prince Edward Island is 0.247, and an associated 95% confidence interval for the true difference is \([-0.138, 0.631]\). The predictor of the difference is 0.234, and a corresponding interval for the true difference based on the bootstrap MSE estimate is \([0.097, 0.371]\). The endpoints of the interval based on the predictor and bootstrap MSE estimate are \( \hat{p}_{11,B} - \hat{p}_{21,B} \pm t_{0.975,df}[\hat{MSE}_{diff,121}]^{0.5} \), where \( t_{0.975,df} \) is the 97.5 percentile of a \( t \) distribution with 117 degrees of freedom, and

\[
\hat{MSE}_{diff,121} = \frac{1}{B} \left\{ \sum_{b=1}^{B} (\hat{p}_{11,B}^{(b)} - \hat{p}_{11}^{(b)})^2 + (\hat{p}_{21,B}^{(b)} - \hat{p}_{21}^{(b)})^2 - 2(\hat{p}_{11,B}^{(b)} - \hat{p}_{11}^{(b)})(\hat{p}_{21,B}^{(b)} - \hat{p}_{21}^{(b)}) \right\} - \hat{b}_{11} - \hat{b}_{21}.
\]

### 7.5.3 Comparison of MSE’s and CV’s of Predictors to MSE’s and CV’s of Direct Estimators for A1

Ratios of the bootstrap MSE estimates to the estimates of the sampling variances of the proportions and totals are shown in Tables 7.14 and 7.15, respectively. The provinces are listed in decreasing order with respect to the estimated coefficient of variation of \( \hat{M}_{ik} \). (The bootstrap estimators of the MSE’s of the predictors of the proportions and totals are denoted by \( \hat{MSE}_{ik}^{bs} \) and \( \hat{MSE}_{ik,T}^{bs} \), respectively, and are defined in (5.36).) The MSE estimates for the predictors of the proportions are smaller than the corresponding estimates of the sampling
variances. In three digit code A12 of Ontario, the bootstrap MSE estimate is essentially equal to the estimated sampling variance. The MSE ratios tend to approach 1 as the province sizes increase because the sampling variances are smaller in the larger provinces and the predictors approach the direct estimators as the sampling variances decrease. The MSE estimates for the predictors of the totals are smaller than the corresponding estimates of the sampling variances for all cells except for A11 in Nova Scotia and A12 in Manitoba. The reductions in the estimated MSE's are usually smaller for totals than for proportions because variability in the direct estimators of the province two digit totals limits the possible reduction in the MSE in the scale of the totals.

Several of the estimated MSE ratios are noticeably large relative to the other MSE ratios in the same province and in provinces with similar sizes. The MSE ratios associated with the proportion and total in three digit code A14 in Prince Edward Island are 0.448 and 0.985, respectively. The MSE ratios in three digit code A12 in New Foundland are 0.419 and 0.723 for the proportion and total, respectively. Although the estimated MSE’s are smaller than the corresponding estimated sampling variances, the MSE ratios for the categories mentioned above are large relative to the other MSE ratios in the same provinces. For example, the MSE ratios for the proportions in the other categories in Prince Edward Island and New Foundland are between 0.129 and 0.177. Unusually large MSE ratios also occur in Manitoba and Nova Scotia. The MSE ratios for the proportion and total in A12 in Manitoba are 0.940 and 1.080, respectively. The MSE ratios for the proportion and total in A11 in Nova Scotia are 0.639 and 1.227, respectively.

Other estimated MSE’s are unusually small relative to the estimates of the sampling variances. In particular, the MSE ratios for three digit code A14 in Nova Scotia and three digit code A12 in Saskatchewan are smaller than the other MSE ratios in the same provinces. For example, the MSE ratio for the total in three digit code A14 in Nova Scotia is 0.236, while the MSE ratios for A11, A12, and A13 are 1.227, 0.541, and 0.567, respectively.

The unusually large and small MSE ratios noted above arise because of relationships between the estimated means and the estimated variances. As shown in Figure 7.1, the direct
estimates of the sampling variances are positively correlated with the diagonal elements of
\( n_k^{-1} \text{diag}(\hat{\mathbf{p}}_k - \hat{\mathbf{p}}_k \hat{\mathbf{p}}_k') \). As a consequence, when \( \hat{p}_{ik} \) is close to zero, the direct estimate of the sampling variance is close to zero. Also, when \( \hat{p}_{ik} \) is small, \( \hat{p}_{T,ik} > \hat{p}_{ik} \), and \( \hat{p}_{T,ik}(1 - \hat{p}_{T,ik}) > \hat{p}_{ik}(1 - \hat{p}_{ik}) \). While the direct estimates of the sampling variances are related to \( \{\hat{p}_{ik}(1 - \hat{p}_{ik}) : i = 1, \ldots, 4; k = 1, \ldots, 10\} \), the estimate of the MSE of the predictor is related to \( \hat{p}_{T,ik}(1 - \hat{p}_{T,ik}) \) through the estimate of the variance of \( u_{ik} \) and the estimate of the variance of \( \hat{p}_{T,ik} \). When \( 0.5 > \hat{p}_{T,ik} > \hat{p}_{ik} \), the components of the estimated MSE’s that are related to \( \hat{p}_{T,ik}(1 - \hat{p}_{T,ik}) \) can cause the estimated MSE’s to be unusually large relative to the sampling variances. This pattern explains the unusually large MSE ratios for A14 in Prince Edward Island, A12 in New Foundland, A11 in Nova Scotia and A12 in Manitoba. The relationships between means and variances also explains why the MSE ratios for A12 in Saskatchewan and for A14 in Nova Scotia are unusually small. In these two cells, \( 0.5 > \hat{p}_{ik} > \hat{p}_{T,ik} \), and the direct estimates of the sampling variances are relatively large. Analogous arguments explain the unusual MSE ratios for the totals for the cells mentioned above.

<table>
<thead>
<tr>
<th></th>
<th>PE</th>
<th>NF</th>
<th>NB</th>
<th>NS</th>
<th>SK</th>
<th>MB</th>
<th>BC</th>
<th>AB</th>
<th>QC</th>
<th>ON</th>
</tr>
</thead>
<tbody>
<tr>
<td>A11</td>
<td>0.144</td>
<td>0.136</td>
<td>0.284</td>
<td>0.499</td>
<td>0.440</td>
<td>0.505</td>
<td>0.546</td>
<td>0.591</td>
<td>0.822</td>
<td></td>
</tr>
<tr>
<td>A12</td>
<td>0.144</td>
<td>0.420</td>
<td>0.353</td>
<td>0.372</td>
<td>0.221</td>
<td>0.944</td>
<td>0.377</td>
<td>0.418</td>
<td>0.573</td>
<td>0.996</td>
</tr>
<tr>
<td>A13</td>
<td>0.177</td>
<td>0.129</td>
<td>0.228</td>
<td>0.356</td>
<td>0.393</td>
<td>0.453</td>
<td>0.566</td>
<td>0.487</td>
<td>0.560</td>
<td>0.784</td>
</tr>
<tr>
<td>A14</td>
<td>0.449</td>
<td>0.212</td>
<td>0.382</td>
<td>0.189</td>
<td>0.362</td>
<td>0.749</td>
<td>0.574</td>
<td>0.583</td>
<td>0.776</td>
<td></td>
</tr>
</tbody>
</table>

Table 7.14 Ratios of bootstrap estimates of MSE’s to sampling variances for proportions in A1.
Table 7.15  Ratios of bootstrap estimates of MSE’s to sampling variances
for totals in A1.

<table>
<thead>
<tr>
<th></th>
<th>PE</th>
<th>NF</th>
<th>NB</th>
<th>NS</th>
<th>SK</th>
<th>MB</th>
<th>BC</th>
<th>AB</th>
<th>QC</th>
<th>ON</th>
</tr>
</thead>
<tbody>
<tr>
<td>A11</td>
<td>0.512</td>
<td>0.388</td>
<td>0.522</td>
<td>1.228</td>
<td>0.840</td>
<td>0.587</td>
<td>0.608</td>
<td>0.723</td>
<td>0.742</td>
<td>0.956</td>
</tr>
<tr>
<td>A12</td>
<td>0.294</td>
<td>0.724</td>
<td>0.499</td>
<td>0.542</td>
<td>0.259</td>
<td>1.084</td>
<td>0.438</td>
<td>0.505</td>
<td>0.669</td>
<td>0.949</td>
</tr>
<tr>
<td>A13</td>
<td>0.394</td>
<td>0.310</td>
<td>0.415</td>
<td>0.566</td>
<td>0.658</td>
<td>0.695</td>
<td>0.837</td>
<td>0.703</td>
<td>0.755</td>
<td>0.915</td>
</tr>
<tr>
<td>A14</td>
<td>0.986</td>
<td>0.465</td>
<td>0.609</td>
<td>0.235</td>
<td>0.465</td>
<td>0.431</td>
<td>0.859</td>
<td>0.675</td>
<td>0.611</td>
<td>0.830</td>
</tr>
</tbody>
</table>

Table 7.16 shows the estimated coefficients of variation for the cells mentioned above with unusually large or small MSE ratios. The estimated CV of the direct estimator is \( \hat{\sigma}_{e,ik} \hat{p}_{ik}^{-1} \). The estimated CV of the predictor is \( (\widehat{MSE}_{ik})^{0.5} \hat{p}_{ik,B}^{-1} \). Because of the relationship between the means and the variances, the procedure leads to a reduction in the estimated coefficients of variation even for the cells with relatively large MSE ratios. The estimated CV’s of the direct estimators are larger than 70% for cells in which the direct estimator of the proportion is smaller than 0.05. The estimated CV’s of the predictors of the proportions in Table 7.16 are smaller than 30%.

<table>
<thead>
<tr>
<th>3-digit</th>
<th>Province</th>
<th>CV direct</th>
<th>CV predictor</th>
</tr>
</thead>
<tbody>
<tr>
<td>A14</td>
<td>PE</td>
<td>1.032</td>
<td>0.263</td>
</tr>
<tr>
<td>A12</td>
<td>NF</td>
<td>1.030</td>
<td>0.248</td>
</tr>
<tr>
<td>A11</td>
<td>NS</td>
<td>0.382</td>
<td>0.134</td>
</tr>
<tr>
<td>A12</td>
<td>MB</td>
<td>0.729</td>
<td>0.267</td>
</tr>
<tr>
<td>A14</td>
<td>NS</td>
<td>0.253</td>
<td>0.199</td>
</tr>
<tr>
<td>A12</td>
<td>SK</td>
<td>0.295</td>
<td>0.223</td>
</tr>
</tbody>
</table>

Table 7.16  Estimated Coefficients of Variation
7.5.4 Model Assessment for A1

To diagnose the fit of the model, we calculate standardized residuals of the form

\[ \hat{r}_{ik} = \frac{\hat{p}_{ik} - \hat{p}_{T,ik}}{(\hat{\sigma}^2_{a,ik,dir} + \hat{\psi}\hat{\sigma}^2_{b,ik})^{0.5}}, \]  

(7.9)

where \( \hat{\sigma}^2_{a,ik,dir} \) and \( \hat{\sigma}^2_{b,ik} \) are the diagonal elements of \( \hat{\Sigma}_{a,dir} \) and \( \hat{\Sigma}_{b} \) corresponding to three digit code \( i \) and province \( k \), and \( \hat{\psi} \) is estimate of \( \psi \) in Table 7.9. The matrices \( \Sigma_a \) and \( \Sigma_b \) are defined in (4.14) and (4.15), and \( \hat{\Sigma}_{a,dir} + \hat{\psi}\hat{\Sigma}_{b} \) is an estimator of the variance of \( \hat{p} - \hat{p}_T \). Rao (2003, pg. 111) suggests residuals similar to (7.9) in the context of a linear mixed model. The absolute values of the standardized residuals are smaller than 2 for most of the provinces. The standardized residual for A12 in Manitoba is -2.18. The standardized residuals for A11 and A14 in Nova Scotia are -3.12 and 2.07, respectively. A criticism of the residuals (7.9) is that the residuals are sensitive to the variance estimator in the denominator. If the working covariance matrix is used to compute the standardized residuals instead of the direct estimator, then the residual for A11 is -2.38, while the residual for A14 is 2.56 because the working model leads to a larger estimated variance for A11 and a smaller estimated variance for A14. The remaining residuals are smaller than 2 in absolute value if the working model is used to construct the variance in the denominator. Figure 7.11 shows the standardized residuals for A1 plotted against \( \hat{p}_{T,ik} \). To check for a relationship between \( \hat{p}_{T,ik} \) and the variance of the residuals, we also plot the absolute values of the residuals against \( \hat{p}_{T,ik} \) in Figure 7.12. The labels in the plots indicate the province and three digit code. (For example, NS4 is the residual for A14 in Nova Scotia.) Neither plot exhibits a systematic pattern for residuals as a function of \( \hat{p}_{T,ik} \).

<table>
<thead>
<tr>
<th>PE</th>
<th>NF</th>
<th>NB</th>
<th>NS</th>
<th>SK</th>
<th>MB</th>
<th>BC</th>
<th>AB</th>
<th>QC</th>
<th>ON</th>
</tr>
</thead>
<tbody>
<tr>
<td>i = 1</td>
<td>0.058</td>
<td>-0.471</td>
<td>-0.127</td>
<td>-3.116</td>
<td>-1.950</td>
<td>0.271</td>
<td>1.168</td>
<td>-0.432</td>
<td>0.718</td>
</tr>
<tr>
<td>i = 2</td>
<td>-0.239</td>
<td>-1.551</td>
<td>-0.272</td>
<td>-0.780</td>
<td>1.634</td>
<td>-2.179</td>
<td>1.797</td>
<td>1.430</td>
<td>0.190</td>
</tr>
<tr>
<td>i = 3</td>
<td>0.916</td>
<td>1.507</td>
<td>1.097</td>
<td>1.077</td>
<td>-0.023</td>
<td>0.316</td>
<td>-1.539</td>
<td>-0.393</td>
<td>-0.704</td>
</tr>
<tr>
<td>i = 4</td>
<td>-1.538</td>
<td>-0.427</td>
<td>-1.273</td>
<td>2.074</td>
<td>0.549</td>
<td>0.636</td>
<td>-1.674</td>
<td>-0.503</td>
<td>-0.261</td>
</tr>
</tbody>
</table>

Table 7.17  Standardized residuals for A1
Figure 7.11 Standardized residuals for A1 (y-axis); $\hat{p}_{T,ik}$ (x-axis)

Figure 7.12 Absolute values of standardized residuals for A1 (y-axis); $|\hat{p}_{T,ik}|$ (x-axis)
As a global measure of discrepancy between the direct estimators and the fitted model, we compute the quadratic form,

\[ Q = (\hat{p} - \hat{p}_T)'[\hat{\Sigma}_{a,dir} + \hat{\psi}\hat{\Sigma}_b]^{-1}(\hat{p} - \hat{p}_T). \] (7.10)

The matrices \( \hat{\Sigma}_{a,dir} \) and \( \hat{\Sigma}_b \) are estimates of \( \Sigma_a \) defined in (4.14) and \( \Sigma_b \) defined in (4.15), respectively. Under an assumption of normality, the statistic \( Q \) has a chi-squared distribution if the variance of \( \hat{p} - \hat{p}_T \) is known. The degrees of freedom of the chi-squared distribution is the rank of \( [\hat{\Sigma}_a + \hat{\psi}\hat{\Sigma}_b] \), and the rank is \( (C-1)(K-1) \) if \( \hat{\theta} \) is restricted to equal 1. The rank is \( (C-1)(K-1)-1 \) if \( \hat{\theta} \) is unrestricted. To account for estimation of \( \psi \), we compare the statistic \( Q \) to a chi-squared distribution with \( (C-1)(K-1)-1 \) or \( (C-1)(K-1)-2 \) degrees of freedom if \( \hat{\theta} \) is restricted to equal 1 or if \( \hat{\theta} \) is unrestricted, respectively. Because A1 has four categories \( (C = 4) \) and we assume that the coefficient on the Census interactions is equal to 1, we compare \( Q \) to a chi-squared distribution with \( 3(9)-1 = 26 \) degrees of freedom. The observed \( Q \) is 41.73. The probability that a chi-squared random variable with 26 degrees of freedom exceeds the observed \( Q \) is 0.026. It is not surprising that the observed \( Q \) is unusual with respect to a chi-squared distribution because several of the standardized residuals exceed 2 in absolute value. The estimator \( \hat{\Sigma}_{a,dir} \) may be unreliable because \( \hat{\Sigma}_{a,dir} \) depends on the direct estimator of \( \Sigma_{ee,k} \).

If the test statistic \( Q \) is constructed with the working model estimate of \( \Sigma_{ee,k} \) instead of the direct estimate of \( \Sigma_{ee,k} \), then \( Q = 35.36 \). The probability that a chi-squared random variable with 26 degrees of freedom exceeds 35.36 is 0.104. The test is approximate, and it is difficult to separate lack of fit of the model for the mean from lack of fit of the working model for the variances. The test is also sensitive to variability in the direct estimator of the sampling variance. Although the observed \( Q \) is unlikely under a chi-squared reference distribution, we do not reject the model on the basis of this approximate test.

In practice, only the data from previous Censuses are available to use for prediction. In this study, the Census 2006 data are also available. For consistency with a realistic situation, we use the Census 2001 data to construct the predictors. We compare the predictors to the May 2006 data as a way to evaluate the procedure. Under the assumption that minimal changes to the occupational structure of the population occur between May 2005 and the date of the
2006 Census, the predictors of the proportions for May 2005 should be close to the Census 2006 proportions.

The differences between the predictors for May 2005 and the Census 2006 proportions are smaller than the differences between the direct estimators for May 2005 and the Census 2006 proportions. All but one of the 95% prediction intervals for the predictors contain the Census 2006 proportions. (The 2006 Census proportion for A14 in Saskatchewan is smaller than the lower bound of the corresponding 95% prediction interval.) Six of the intervals for the direct estimators do not contain the corresponding Census 2006 proportions. The upper bounds of the intervals for the direct estimators of the proportions for A11 in Nova Scotia and Saskatchewan and for A12 in Ontario and Manitoba are smaller than the Census 2006 proportions. Similarly, the lower bounds of the intervals for the direct estimators in A14 in Nova Scotia and Ontario are larger than the Census 2006 proportions.

To summarize the distance between the predictors for May 2005 and the Census 2006 proportions, we compare the sum of the absolute values of the differences between the direct estimators and the Census 2006 proportions to the sum of the absolute values of the differences between the predictors and the Census 2006 proportions. The sum of the absolute differences calculated with the direct estimators is 2.66, while the sum of the absolute differences calculated with the predictors is 1.31.

### 7.6 Estimates and Predictions for E0

Tables 7.18 and 7.19 show the Census proportions and the direct estimates of the proportions, respectively, for the two digit code E0. As shown in the right plot of Figure 7.13, the LFS proportions in E0 are approximately linearly related to the Census proportions. The proportions in category E01 are smaller than the proportions in E02 and E03 in both the Census and the LFS. In the left plot of Figure 7.13, the logits of the direct estimators of the proportions are plotted against the Census interactions. The logits of the direct estimators are approximately linearly related to the Census interactions in each three digit code. The logits in Figure 7.13 form two approximately horizontal lines. Seven out of the nine logits
on the top line are for three digit code E02. In all of the provinces except Nova Scotia and Saskatchewan, the province two digit totals recorded in the Census are larger than the LFS estimates of the province two digit totals. The estimates of the coefficients of variation of the province two digit totals range from 0.064 in Ontario to 0.186 in Prince Edward Island. The direct estimator of the province two digit total in Alberta is the 4th largest; however, five of the estimated coefficients of variation for the direct estimators of the province two digit totals are smaller than the estimated coefficient of variation for Alberta. In particular, the direct estimator of the two digit total in Alberta is larger than the direct estimator of the two digit total in Manitoba and smaller than the direct estimator of the two digit total in British Columbia. The estimated coefficient of variation for the direct estimator of the two digit total in Alberta is larger than the estimated coefficients of variation for both Manitoba and Saskatchewan.
Figure 7.13 Direct estimates and Census proportions for E0. Left plot: logits of direct estimators of proportions (y-axis), Census interactions (x-axis). Right plot: LFS proportions (y-axis), Census proportions (x-axis). Numbers correspond to three digit codes, $i = 1, 2, 3$. 
Table 7.18 Census 2001 proportions and province two digit totals for E0
(judges, lawyers, psychologists, social workers, ministers of religion, and policy and program officers)

<table>
<thead>
<tr>
<th></th>
<th>PE</th>
<th>NF</th>
<th>NB</th>
<th>NS</th>
<th>SK</th>
<th>MB</th>
<th>BC</th>
<th>AB</th>
<th>QC</th>
<th>ON</th>
</tr>
</thead>
<tbody>
<tr>
<td>E01</td>
<td>0.167</td>
<td>0.117</td>
<td>0.162</td>
<td>0.178</td>
<td>0.162</td>
<td>0.173</td>
<td>0.213</td>
<td>0.207</td>
<td>0.231</td>
<td>0.213</td>
</tr>
<tr>
<td>E02</td>
<td>0.360</td>
<td>0.499</td>
<td>0.469</td>
<td>0.460</td>
<td>0.495</td>
<td>0.485</td>
<td>0.400</td>
<td>0.414</td>
<td>0.336</td>
<td>0.355</td>
</tr>
<tr>
<td>E03</td>
<td>0.473</td>
<td>0.384</td>
<td>0.369</td>
<td>0.362</td>
<td>0.342</td>
<td>0.342</td>
<td>0.386</td>
<td>0.379</td>
<td>0.433</td>
<td>0.432</td>
</tr>
</tbody>
</table>

Table 7.19 May 2005 LFS proportions, province two digit totals (\(\hat{M}_k\)), and estimated coefficients of variation of the estimators of the province two digit totals for E0

<table>
<thead>
<tr>
<th></th>
<th>PE</th>
<th>NF</th>
<th>NB</th>
<th>NS</th>
<th>SK</th>
<th>MB</th>
<th>BC</th>
<th>AB</th>
<th>QC</th>
<th>ON</th>
</tr>
</thead>
<tbody>
<tr>
<td>E01</td>
<td>0.102</td>
<td>0.138</td>
<td>0.130</td>
<td>0.248</td>
<td>0.185</td>
<td>0.164</td>
<td>0.256</td>
<td>0.157</td>
<td>0.330</td>
<td>0.214</td>
</tr>
<tr>
<td>E02</td>
<td>0.281</td>
<td>0.472</td>
<td>0.549</td>
<td>0.332</td>
<td>0.434</td>
<td>0.592</td>
<td>0.361</td>
<td>0.351</td>
<td>0.284</td>
<td>0.351</td>
</tr>
<tr>
<td>E03</td>
<td>0.617</td>
<td>0.389</td>
<td>0.321</td>
<td>0.420</td>
<td>0.381</td>
<td>0.244</td>
<td>0.383</td>
<td>0.492</td>
<td>0.386</td>
<td>0.435</td>
</tr>
</tbody>
</table>

\[CV\text{ of } \hat{M}_k\] 19% 17% 14% 13% 11% 10% 10% 12% 9% 6%

7.6.1 Model Estimates for E0

Table 7.20 contains estimates of the parameters of the reduced model for E0 in which the coefficient on the Census interactions is restricted to equal 1. The two digit codes E0 and E1 are used to estimate the common variance parameter \(\psi_5\). The variance of \(\hat{\lambda}\) is defined in (A.8) of Appendix 2. The variance of the estimator of \(\psi\) is defined in (4.28).

That the estimates of \(\alpha_{o,2}\) and \(\alpha_{o,3}\) are positive is consistent with the property that the national fractions for E02 and E03 are larger than the national fraction for E01. The relative
magnitudes of the estimates of $\alpha_{o,2}$ and $\alpha_{o,3}$ are opposite the ordering of the national fractions for categories E02 and E03. The national fraction for E03 is larger than the national fraction for E02, but the estimate of $\alpha_{o,2}$ is larger than the estimate of $\alpha_{o,3}$. In seven of the provinces, the Census proportion for E02 is larger than the Census proportion for E03. In the largest two provinces (Quebec and Ontario), the proportions for E03 are larger than the corresponding proportions for E02 in both the Census and the LFS. The national fraction for E03 exceeds the national fraction for E02 in the LFS because the LFS proportion for E03 is larger than the LFS proportion for E02 in the largest two provinces. Because the interactions in the Census are nontrivial, the category effects, $\alpha_{o,1}$ and $\alpha_{o,2}$, are not interpretable as main effects for E0.

The estimate of the common $\psi_5$ is 0.0033 (SE 0.0026). The lower bound for the estimate of $\psi_5$, $\xi_D = 0.00076$, where $\xi_D$ is defined in (4.22). Although pooling the information from E0 and E1 to estimate $\psi_5$ reduces the estimated standard error of $\hat{\psi}_5$, the combined estimate of $\psi_5$ does not differ from zero by more than two standard errors. Because a normal approximation for the distribution of $\hat{\psi}_5$ may be poor, we construct a bootstrap test of the null hypothesis that $\psi_5 = 0$. As described previously, we construct the bootstrap test by simulating from a bootstrap distribution in which $p_{ik,t5}^* = \hat{p}_{T,ik,t5}$ and computing the fraction of bootstrap estimates of $\psi_5$ that exceed the observed value of 0.0033. Of the 500 bootstrap estimates, 17 exceed 0.0033 (p-value = 0.034), an indication that the data are inconsistent with a model in which $\psi_5 = 0$.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$\alpha_{o,2}$</th>
<th>$\alpha_{o,3}$</th>
<th>$\alpha_{o,2} - \alpha_{o,3}$</th>
<th>$\psi$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimate</td>
<td>1.30</td>
<td>1.11</td>
<td>0.19</td>
<td>0.0033</td>
</tr>
<tr>
<td>SE</td>
<td>0.11</td>
<td>0.11</td>
<td>0.087</td>
<td>0.0028</td>
</tr>
</tbody>
</table>

Table 7.20 Estimates of the parameters of the reduced model for E0 and Taylor standard errors. Coefficient on Census interactions, $\theta_o = 1$

To evaluate whether the assumption that $\theta_{o,t5} = 1$ is reasonable, the estimates of $\theta_{o,t5}$
(t = 1, 2) are obtained for E0 and E1 separately (Table 7.21). The estimate of \( \theta_{o,t5} \) for E0 differs from zero by more than two estimated standard deviations but differs from one by approximately one estimated standard deviation. The estimate of \( \theta_{o,t5} \) for E1 does not differ from zero or one by more than two estimated standard deviations. To test the null hypothesis that \( \theta_{o,t5} = 1 \), we compute

\[
W_\theta = \frac{(\hat{\theta}_{15} - 1)^2}{\hat{V}\{\theta_{15}\}} + \frac{(\hat{\theta}_{25} - 1)^2}{\hat{V}\{\theta_{25}\}} = 3.73.
\]

Comparing 3.73 to a \( \chi^2 \) distribution with 2 degrees of freedom gives a p-value of 0.15. We conclude that the data are consistent with a reduced model in which \( \theta_{o,t5} = 1 \).

The score statistic for testing the null hypothesis that \( \psi_{t5} = \psi_5 \) for both two digit codes \( (t = 1, 2) \) in the one digit code E is \( S_\psi = 0.0015 \). A p-value based on a \( \chi^2(1) \) reference distribution is 0.97, and a bootstrap p-value, obtained using the test procedure described in Section 7.3 is 0.98. We fail to reject null hypothesis that \( \psi_{t5} = \psi_5 \) (t = 1, 2) and construct predictors under the reduced model.

<table>
<thead>
<tr>
<th>2-digit Code</th>
<th>C</th>
<th>( \hat{\theta} )</th>
<th>( \sqrt{\hat{V}{\hat{\theta}}} )</th>
<th>( \hat{\psi} )</th>
<th>( \sqrt{\hat{V}{\hat{\psi}}} )</th>
<th>( 0.5\sqrt{\hat{V}_0{\hat{\psi}}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>E0</td>
<td>3</td>
<td>1.29</td>
<td>0.31</td>
<td>0.0038</td>
<td>0.0054</td>
<td>0.0018</td>
</tr>
<tr>
<td>E1</td>
<td>3</td>
<td>0.51</td>
<td>0.29</td>
<td>0.0025</td>
<td>0.0032</td>
<td>0.0009</td>
</tr>
</tbody>
</table>

Table 7.21 Estimates of \( \theta_o \) and \( \psi \) for E0 and E1 with Taylor standard errors. The last column has the lower bounds for the estimators of \( \psi \). The second column has the number of three digit codes in each two digit code.

### 7.6.2 Predictions for E0

Tables 7.22 - 7.24 contain the predictors, the direct estimators, and the synthetic estimators for E0. The ratios of the estimated variances of \( \{u_k : k = 1,\ldots,10\} \) to the sampling variances under the working model range from 0.085 in Prince Edward Island to 0.847 in Ontario. Because the estimates of the variances of \( \{u_k : k = 1,\ldots,10\} \) are small
relative to the estimates of the sampling variances, the Beale estimates of $\gamma_k$ are smaller than 0.5 in all provinces except Ontario. It follows that the distance between the initial predictors and the synthetic estimators is smaller than the distance between the initial predictors and the direct estimators in all provinces except Ontario, where the weight assigned to the direct estimator in forming the initial predictor is 0.53.

The raking operation changes the relative distances between the predictors, the synthetic estimators, and the direct estimators. For 17 of the 30 cells, the distance between the benchmarked predictors and the direct estimators is smaller than the distance between the initial predictors and the direct estimators. The benchmarked predictors for three digit codes E02 in New Foundland, E01 in Saskatchewan, and E03 in British Columbia are closer to the direct estimators than to the synthetic estimators. After the raking operation, the predictors for E01 and E02 in Ontario are closer to the synthetic estimators than to the direct estimators. The predictor is between the direct estimator and the synthetic estimator for all cells except for E03 in Ontario, where the predictor is smaller than both the direct estimator and the synthetic estimator.
<table>
<thead>
<tr>
<th></th>
<th>PE</th>
<th>NF</th>
<th>NB</th>
<th>NS</th>
<th>SK</th>
<th>MB</th>
<th>BC</th>
<th>AB</th>
<th>QC</th>
<th>ON</th>
</tr>
</thead>
<tbody>
<tr>
<td>E01</td>
<td>0.102</td>
<td>0.138</td>
<td>0.130</td>
<td>0.248</td>
<td>0.185</td>
<td>0.164</td>
<td>0.256</td>
<td>0.157</td>
<td>0.330</td>
<td>0.214</td>
</tr>
<tr>
<td>(SE)</td>
<td>(0.055)</td>
<td>(0.052)</td>
<td>(0.047)</td>
<td>(0.062)</td>
<td>(0.045)</td>
<td>(0.037)</td>
<td>(0.044)</td>
<td>(0.046)</td>
<td>(0.044)</td>
<td>(0.028)</td>
</tr>
<tr>
<td>E02</td>
<td>0.281</td>
<td>0.472</td>
<td>0.549</td>
<td>0.332</td>
<td>0.434</td>
<td>0.592</td>
<td>0.361</td>
<td>0.351</td>
<td>0.284</td>
<td>0.351</td>
</tr>
<tr>
<td>(SE)</td>
<td>(0.076)</td>
<td>(0.081)</td>
<td>(0.070)</td>
<td>(0.064)</td>
<td>(0.058)</td>
<td>(0.059)</td>
<td>(0.048)</td>
<td>(0.052)</td>
<td>(0.036)</td>
<td>(0.030)</td>
</tr>
<tr>
<td>E03</td>
<td>0.617</td>
<td>0.389</td>
<td>0.321</td>
<td>0.420</td>
<td>0.381</td>
<td>0.244</td>
<td>0.383</td>
<td>0.492</td>
<td>0.386</td>
<td>0.435</td>
</tr>
<tr>
<td>(SE)</td>
<td>(0.089)</td>
<td>(0.078)</td>
<td>(0.059)</td>
<td>(0.065)</td>
<td>(0.054)</td>
<td>(0.049)</td>
<td>(0.046)</td>
<td>(0.061)</td>
<td>(0.043)</td>
<td>(0.030)</td>
</tr>
</tbody>
</table>

Table 7.22 Direct estimates of proportions and sampling standard deviations, E0

<table>
<thead>
<tr>
<th></th>
<th>PE</th>
<th>NF</th>
<th>NB</th>
<th>NS</th>
<th>SK</th>
<th>MB</th>
<th>BC</th>
<th>AB</th>
<th>QC</th>
<th>ON</th>
</tr>
</thead>
<tbody>
<tr>
<td>E01</td>
<td>0.180</td>
<td>0.134</td>
<td>0.174</td>
<td>0.209</td>
<td>0.184</td>
<td>0.189</td>
<td>0.243</td>
<td>0.215</td>
<td>0.282</td>
<td>0.227</td>
</tr>
<tr>
<td>(SE)</td>
<td>(0.025)</td>
<td>(0.021)</td>
<td>(0.024)</td>
<td>(0.028)</td>
<td>(0.024)</td>
<td>(0.024)</td>
<td>(0.027)</td>
<td>(0.027)</td>
<td>(0.029)</td>
<td>(0.023)</td>
</tr>
<tr>
<td>E02</td>
<td>0.332</td>
<td>0.473</td>
<td>0.459</td>
<td>0.415</td>
<td>0.460</td>
<td>0.485</td>
<td>0.370</td>
<td>0.379</td>
<td>0.301</td>
<td>0.340</td>
</tr>
<tr>
<td>(SE)</td>
<td>(0.032)</td>
<td>(0.034)</td>
<td>(0.033)</td>
<td>(0.033)</td>
<td>(0.033)</td>
<td>(0.033)</td>
<td>(0.030)</td>
<td>(0.030)</td>
<td>(0.027)</td>
<td>(0.025)</td>
</tr>
<tr>
<td>E03</td>
<td>0.489</td>
<td>0.393</td>
<td>0.367</td>
<td>0.376</td>
<td>0.356</td>
<td>0.326</td>
<td>0.387</td>
<td>0.407</td>
<td>0.417</td>
<td>0.433</td>
</tr>
<tr>
<td>(SE)</td>
<td>(0.035)</td>
<td>(0.033)</td>
<td>(0.032)</td>
<td>(0.032)</td>
<td>(0.031)</td>
<td>(0.030)</td>
<td>(0.030)</td>
<td>(0.032)</td>
<td>(0.031)</td>
<td>(0.026)</td>
</tr>
</tbody>
</table>

Table 7.23 Predictors of proportions and square roots of bootstrap MSE estimates, E0

<table>
<thead>
<tr>
<th></th>
<th>PE</th>
<th>NF</th>
<th>NB</th>
<th>NS</th>
<th>SK</th>
<th>MB</th>
<th>BC</th>
<th>AB</th>
<th>QC</th>
<th>ON</th>
</tr>
</thead>
<tbody>
<tr>
<td>E01</td>
<td>0.182</td>
<td>0.130</td>
<td>0.178</td>
<td>0.195</td>
<td>0.179</td>
<td>0.190</td>
<td>0.232</td>
<td>0.225</td>
<td>0.250</td>
<td>0.231</td>
</tr>
<tr>
<td>(SE)</td>
<td>(0.015)</td>
<td>(0.011)</td>
<td>(0.015)</td>
<td>(0.016)</td>
<td>(0.015)</td>
<td>(0.015)</td>
<td>(0.018)</td>
<td>(0.017)</td>
<td>(0.019)</td>
<td>(0.018)</td>
</tr>
<tr>
<td>E02</td>
<td>0.339</td>
<td>0.476</td>
<td>0.445</td>
<td>0.436</td>
<td>0.471</td>
<td>0.460</td>
<td>0.376</td>
<td>0.390</td>
<td>0.314</td>
<td>0.333</td>
</tr>
<tr>
<td>(SE)</td>
<td>(0.018)</td>
<td>(0.020)</td>
<td>(0.020)</td>
<td>(0.020)</td>
<td>(0.020)</td>
<td>(0.020)</td>
<td>(0.019)</td>
<td>(0.019)</td>
<td>(0.017)</td>
<td>(0.018)</td>
</tr>
<tr>
<td>E03</td>
<td>0.479</td>
<td>0.394</td>
<td>0.377</td>
<td>0.369</td>
<td>0.351</td>
<td>0.350</td>
<td>0.391</td>
<td>0.385</td>
<td>0.436</td>
<td>0.436</td>
</tr>
<tr>
<td>(SE)</td>
<td>(0.020)</td>
<td>(0.019)</td>
<td>(0.019)</td>
<td>(0.019)</td>
<td>(0.018)</td>
<td>(0.018)</td>
<td>(0.019)</td>
<td>(0.019)</td>
<td>(0.020)</td>
<td>(0.020)</td>
</tr>
</tbody>
</table>

Table 7.24 Estimates of \( \hat{p}_{T,ik} : i = 1, 2, 3; k = 1, \ldots, 10 \) and standard errors, E0

Prediction intervals for the direct estimators and the predictors are plotted in Figure 7.14. Three sets of intervals are plotted for each province (one set of two intervals for each three digit code). The prediction intervals are calculated as in (7.7) and (7.8), where the
degrees of freedom used in (7.8) is $df = 9(2 + 2) = 36$. The intervals for the predictors are narrower than the intervals for the direct estimators because the estimates of the MSE’s of the predictors are smaller than the direct estimates of the sampling variances. The lower interval endpoint based on the direct estimators for E01 in New Foundland is truncated at zero. The lower interval endpoints constructed with the predictors and the bootstrap MSE estimates are strictly positive. The Census 2001 proportions for E01 are smaller than the corresponding proportions for E02 and E03 in all provinces. Likewise, the direct estimates of the proportions for E01 are smaller than the corresponding proportions for E02 and E03 in all provinces except Quebec. In several provinces, the 95% interval for the difference between the proportion in E01 and the proportion in E02 or E03 constructed with the direct estimator contains zero, while the corresponding interval constructed with the predictor does not. For example, in Nova Scotia, the direct estimate of the difference between the proportions in E01 and E02 is $-0.084$, and a corresponding 95% confidence interval is $[-0.295, 0.126]$. The difference between the predictors for E01 and E02 is $-0.206$, and an associated 95% prediction interval based on the bootstrap MSE estimator is $[-0.311, -0.099]$. For this example, the direct estimate of the difference is outside of the 95% interval constructed with the predictors.
7.6.3 Comparison of MSE’s and CV’s of Predictors to MSE’s and CV’s of Direct Estimators for E0

Tables 7.25 and 7.26 contain ratios of the bootstrap estimates of the MSE’s to the direct estimates of the sampling variances for the proportions and totals, respectively, for E0. The estimates of the MSE’s of the predictors for both totals and proportions are all smaller than the corresponding estimates of the sampling variances. The MSE ratios for the totals are larger than the corresponding MSE ratios for the proportions. The estimated coefficients of variation of the direct estimators of the province two digit totals are given in the last row of each table. In general, the MSE ratios approach one as the estimated coefficients of variation
of the direct estimators of the province two digit totals decrease.

<table>
<thead>
<tr>
<th></th>
<th>PE</th>
<th>NF</th>
<th>NB</th>
<th>NS</th>
<th>SK</th>
<th>MB</th>
<th>BC</th>
<th>AB</th>
<th>QC</th>
<th>ON</th>
</tr>
</thead>
<tbody>
<tr>
<td>E01</td>
<td>0.217</td>
<td>0.167</td>
<td>0.262</td>
<td>0.201</td>
<td>0.285</td>
<td>0.441</td>
<td>0.377</td>
<td>0.332</td>
<td>0.446</td>
<td>0.690</td>
</tr>
<tr>
<td>E02</td>
<td>0.179</td>
<td>0.183</td>
<td>0.225</td>
<td>0.271</td>
<td>0.312</td>
<td>0.315</td>
<td>0.390</td>
<td>0.346</td>
<td>0.535</td>
<td>0.689</td>
</tr>
<tr>
<td>E03</td>
<td>0.156</td>
<td>0.179</td>
<td>0.290</td>
<td>0.240</td>
<td>0.324</td>
<td>0.365</td>
<td>0.424</td>
<td>0.268</td>
<td>0.524</td>
<td>0.720</td>
</tr>
</tbody>
</table>

\[
CV(\hat{M}_k) = \{19\% \ 17\% \ 14\% \ 13\% \ 11\% \ 10\% \ 10\% \ 12\% \ 9\% \ 6\% \}
\]

Table 7.25 Ratios of bootstrap estimates of MSE’s to sampling variances for proportions in E0.

<table>
<thead>
<tr>
<th></th>
<th>PE</th>
<th>NF</th>
<th>NB</th>
<th>NS</th>
<th>SK</th>
<th>MB</th>
<th>BC</th>
<th>AB</th>
<th>QC</th>
<th>ON</th>
</tr>
</thead>
<tbody>
<tr>
<td>E01</td>
<td>0.538</td>
<td>0.284</td>
<td>0.467</td>
<td>0.297</td>
<td>0.382</td>
<td>0.617</td>
<td>0.500</td>
<td>0.525</td>
<td>0.563</td>
<td>0.785</td>
</tr>
<tr>
<td>E02</td>
<td>0.585</td>
<td>0.545</td>
<td>0.490</td>
<td>0.683</td>
<td>0.633</td>
<td>0.500</td>
<td>0.635</td>
<td>0.669</td>
<td>0.711</td>
<td>0.860</td>
</tr>
<tr>
<td>E03</td>
<td>0.422</td>
<td>0.492</td>
<td>0.619</td>
<td>0.505</td>
<td>0.538</td>
<td>0.677</td>
<td>0.705</td>
<td>0.487</td>
<td>0.792</td>
<td>0.900</td>
</tr>
</tbody>
</table>

\[
CV(\hat{M}_k) = \{19\% \ 17\% \ 14\% \ 13\% \ 11\% \ 10\% \ 10\% \ 12\% \ 9\% \ 6\% \}
\]

Table 7.26 Ratios of bootstrap estimates of MSE’s to sampling variances for totals in E0.

Tables 7.27 and 7.28 show the estimated coefficients of variation of the direct estimators of the proportions and the predictors of the proportions, respectively, for E0. Five of the estimated coefficients of variation for the direct estimators are larger than 25%. The estimated coefficients of variation for the predictors are smaller than 15% for all cells except for E01 of New Foundland, where the estimated coefficient of variation is 15.9%.
Table 7.27  Estimated coefficients of variation for direct estimators of proportions in E0

<table>
<thead>
<tr>
<th></th>
<th>PE</th>
<th>NF</th>
<th>NB</th>
<th>NS</th>
<th>SK</th>
<th>MB</th>
<th>BC</th>
<th>AB</th>
<th>QC</th>
<th>ON</th>
</tr>
</thead>
<tbody>
<tr>
<td>E01</td>
<td>0.534</td>
<td>0.377</td>
<td>0.359</td>
<td>0.249</td>
<td>0.245</td>
<td>0.224</td>
<td>0.173</td>
<td>0.296</td>
<td>0.134</td>
<td>0.132</td>
</tr>
<tr>
<td>E02</td>
<td>0.272</td>
<td>0.171</td>
<td>0.127</td>
<td>0.192</td>
<td>0.134</td>
<td>0.099</td>
<td>0.134</td>
<td>0.148</td>
<td>0.128</td>
<td>0.084</td>
</tr>
<tr>
<td>E03</td>
<td>0.144</td>
<td>0.201</td>
<td>0.183</td>
<td>0.154</td>
<td>0.141</td>
<td>0.203</td>
<td>0.119</td>
<td>0.125</td>
<td>0.111</td>
<td>0.070</td>
</tr>
</tbody>
</table>

Table 7.28  Estimated coefficients of variation for predictors of proportions in E0

<table>
<thead>
<tr>
<th></th>
<th>PE</th>
<th>NF</th>
<th>NB</th>
<th>NS</th>
<th>SK</th>
<th>MB</th>
<th>BC</th>
<th>AB</th>
<th>QC</th>
<th>ON</th>
</tr>
</thead>
<tbody>
<tr>
<td>E01</td>
<td>0.142</td>
<td>0.159</td>
<td>0.137</td>
<td>0.132</td>
<td>0.132</td>
<td>0.130</td>
<td>0.112</td>
<td>0.125</td>
<td>0.104</td>
<td>0.104</td>
</tr>
<tr>
<td>E02</td>
<td>0.098</td>
<td>0.073</td>
<td>0.072</td>
<td>0.080</td>
<td>0.071</td>
<td>0.068</td>
<td>0.082</td>
<td>0.080</td>
<td>0.089</td>
<td>0.072</td>
</tr>
<tr>
<td>E03</td>
<td>0.072</td>
<td>0.084</td>
<td>0.086</td>
<td>0.084</td>
<td>0.086</td>
<td>0.092</td>
<td>0.077</td>
<td>0.078</td>
<td>0.075</td>
<td>0.059</td>
</tr>
</tbody>
</table>

7.6.4 Model Assessment for E0

The standardized residuals defined in (7.9) for E0 are shown in Table 7.29 and Figure 7.15. The absolute values of the standardized residuals are plotted in Figure 7.16. The labels in the plots indicate the three digit code and the province. (For example, MB2 is the label for three digit code E02 for Manitoba.) The residuals are approximately symmetric around zero. The residuals in E02 and E03 of Manitoba exceed 1.96 in absolute value. The plot of the absolute values of the residuals does not show evidence of a relationship between the mean and the variance for the residuals. The value of the statistic $Q$ defined in (7.10) for E0 is 17.50, which is a typical value with respect to a chi-squared distribution with 17 degrees of freedom. The residuals and the magnitude of $Q$ relative to the chi-squared reference distribution indicate that the data are consistent with the specified model.

The largest absolute residual is 2.15, which occurs in three digit code E02 in Manitoba.
Manitoba is one of the provinces in which the direct estimates of the sampling variances are relatively inconsistent with the working model (Figure 7.6). The residuals in Table 7.29 are calculated with the direct estimator of the sampling covariance matrix. An alternative is to compute residuals under the assumption that the working model for the sampling variance holds by replacing $\hat{\Sigma}_{a, \text{dir}}$ in (7.9) with $\hat{\Sigma}_a$. The residuals calculated with the working model estimates of the sampling variances differ little from the residuals in Table 7.29.

<table>
<thead>
<tr>
<th></th>
<th>PE</th>
<th>NF</th>
<th>NB</th>
<th>NS</th>
<th>SK</th>
<th>MB</th>
<th>BC</th>
<th>AB</th>
<th>QC</th>
<th>ON</th>
</tr>
</thead>
<tbody>
<tr>
<td>E01</td>
<td>-1.367</td>
<td>0.156</td>
<td>-0.956</td>
<td>0.831</td>
<td>0.133</td>
<td>-0.622</td>
<td>0.492</td>
<td>-1.378</td>
<td>1.760</td>
<td>-0.525</td>
</tr>
<tr>
<td>E02</td>
<td>-0.718</td>
<td>-0.044</td>
<td>1.436</td>
<td>-1.547</td>
<td>-0.610</td>
<td>2.151</td>
<td>-0.288</td>
<td>-0.693</td>
<td>-0.711</td>
<td>0.516</td>
</tr>
<tr>
<td>E03</td>
<td>1.501</td>
<td>-0.060</td>
<td>-0.885</td>
<td>0.759</td>
<td>0.539</td>
<td>-1.968</td>
<td>-0.160</td>
<td>1.695</td>
<td>-1.054</td>
<td>-0.037</td>
</tr>
</tbody>
</table>

Table 7.29 Standardized residuals defined in (7.9) for E0.

Figure 7.15 Standardized residuals for E0 (y-axis), $\hat{p}_{T,ik}$ (x-axis)
A comparison with the Census 2006 proportions lends further support for the predictors. The Census 2006 proportions are contained in the 95% intervals constructed with the predictors for all cells except E02 and E03 of Manitoba. The 95% intervals constructed with the direct estimators also do not contain the Census 2006 proportions for E02 and E03 in Manitoba. The Census 2006 proportion for E01 in Quebec is also not contained in the 95% interval for the direct estimator. The sum of the absolute values of the differences between the predictors and the Census 2006 proportions is 0.84. The sum of the absolute differences between the direct estimators and the Census 2006 proportions is 1.32.

### 7.7 Summary

The estimation and prediction procedures are illustrated using the two digit codes A1 (specialist managers) and E0 (judges, lawyers, psychologists, social workers, ministers of religion, and policy and program officers). For both two digit codes, the estimated MSE’s of
the predictors are smaller than the estimated sampling variances for most of the cells. The median of the estimated efficiency gains for A1 in Prince Edward Island, New Foundland, and New Brunswick is 78%, and the median efficiency gain among the corresponding cells for E0 is 81%. (The estimated efficiency gain is defined as \( \frac{\hat{\sigma}_2^{e,ik,md} - \hat{MSE}_{ik}^{bs}}{\hat{\sigma}_2^{e,ik,md}} \).) The median of the estimated efficiency gains for A1 in Nova Scotia, Saskatchewan, and Manitoba is 62%, and the median among the corresponding cells for E0 is 70%. The median of the estimated efficiency gains for A1 among the cells in British Columbia, Alberta, and Quebec is 44%, and the corresponding median estimated efficiency gain for E0 is 61%. The median of the estimated efficiency gains among the cells in Ontario are 20% for A1 and 31% for E0. The estimated efficiency gains tend to decrease as the province sizes increase because the estimated sampling variances tend to decrease as the province sizes increase. The estimated efficiency gains for E0 are generally larger than the estimated efficiency gains for A1 because the ratios of the estimated variances of \( \{ u_{ik} : i = 1,\ldots,C; k = 1,\ldots,K \} \) to the estimated variances of \( \{ e_{ik} : i = 1,\ldots,C; k = 1,\ldots,K \} \) are typically smaller for E0 than for A1. The standardized residuals for both A1 and E0 are approximately symmetric around zero and do not exhibit obvious patterns as functions of \( \hat{p}_{T,ik} \).

The comparison between the predictors and the 2006 Census data lends further credibility to the prediction procedure. The 2006 Census proportions are usually contained in 95% prediction intervals constructed with the predictors, and the predictors are generally closer to the 2006 Census proportions than the direct estimators.

The differences between the direct estimates and the estimates of \( p_{T,ik} \) vary more in A1 than in E0. Plots of the logits of the LFS proportions against the logits of the Census proportions for the different categories in E0 exhibit approximately parallel lines, which is consistent with the specification of the model for the mean. Also, the absolute values of the standardized residuals for E0 are smaller than two in all cells except E02 in Manitoba. In A1, the plots of the logits of the LFS proportions against the logits of the Census proportions are less consistent with the assumptions of the model. In A1, the residuals for three cells exceed two in absolute value, and the test statistic used for an approximate chi-squared test exceeds the
95 percentile of the chi-squared reference distribution. Differences between the data collection procedures used in the LFS and the Census may partially explain why the model for the mean seems to fit better in E0 than in A1. The two digit code A1 is for the occupational group described as “specialist managers.” Perhaps “specialist managers” is less precisely defined than the occupations that comprise the two digit code E0 (judges, lawyers, psychologists, etc.). We conjecture that respondents may interpret “specialist managers” differently when responding to the LFS than when responding to the Census, leading to some of the evidence of lack of fit for A1.

Qualitatively, the raking operation has a noticeable effect on the predictors for both two digit codes. In A1, the predictors for several categories are not between the direct estimators and the estimators of \( p_{T,ik} \). In E0, the benchmarking procedure shifts the predictors toward the direct estimators in 17 of the 30 cells.

For both two digit codes, several categories stand out because of unusually large or small estimated MSE ratios or because the standardized residuals are relatively large in absolute value. The provinces with categories that stand out (e.g., Nova Scotia and Manitoba for A1 and Prince Edward Island and Manitoba for E0) are also provinces where the procedures used to assess the goodness of fit of the working model for the sampling variances suggest lack of fit. We do not expect the working model for the sampling variances to hold for the LFS data due to the complex design and estimation procedures. Nonetheless, we suspect that the unusually large MSE ratios arise because the direct estimators of the sampling variances for these cells are too small. Relationships between the errors in the predictors and the direct estimators of the sampling variances also affect the properties of the standardized residuals. Refining the model for the sampling variances and the tools used to evaluate the fit of the sampling variance model are areas for future work.
As discussed in Section 1.2, estimation for the Labour Force Survey (LFS) poses several challenges. Predictors for the totals and proportions of a two-way table are desired that remain in the natural parameter spaces and preserve the direct estimators of the marginal totals. Because the LFS uses a complex sample design, an assumption of simple random sampling is not appropriate, and direct estimators for different areas are not identically distributed. Due to limited information about the variances of estimated variances as well as the need for benchmarking, standard MSE estimators are not directly applicable. In this section, we summarize how we dealt with these challenges and identify potential areas for improvement and future study.

8.1 Summary

A prediction procedure based on a mixed model for the direct estimators of the proportions is developed. In simulations, the predictors are more efficient than the direct estimators, and the empirical coverages of prediction intervals based on Taylor and bootstrap MSE estimators are close to the nominal 95% level. In an application of the method to the two digit codes A1 and E0 of the LFS, estimated efficiency gains are obtained for both proportions and totals. The results of the simulations and the application demonstrate the efficiency of the proposed methods. The methods and results of the seven main sections are summarized below.

Section 3: A model for the direct estimators

In Section 3, a mixed model is specified for the direct estimators of the proportions. A nonlinear model for the mean incorporates census data by including census interactions as covariates. The form of the expectation function respects the natural parameter space. The
specified covariance structure for the vector of random effects arises under a variety of distribu-
tions, including multinomial and Dirichlet distributions. An advantage of the assumption
that the covariance matrix of the vector of random effects is proportional to a known function
of the mean is that multiple two-way tables with different dimensions can be used to estimate
a common proportionality constant, $\psi$. This property is used in Section 7 where the three two
digit codes in the one digit code A are used to estimate the between area component of variance
denoted $\psi$. In the model, the sampling covariance matrix is an unknown, unstructured matrix.
This generality reflects the complexity of the LFS sample design and estimation procedures.
The minimum MSE convex combination of the direct estimator and the mean, $p_{T,ik}$, depends
on the ratio of the variance of the random effect to the variance of the sampling error.

Section 4: Parameter estimators

A multinomial model for the sampling variances is used as a working model for estimation
and prediction. The generalized least squares estimator of the mean under the multinomial
model is used to define a set of estimating equations. The multinomial model for the covariance
matrix of the vector of sampling errors is a mathematically convenient choice for two main
reasons. One, the covariance matrix of the vector of sampling errors under the multinomial
model is proportional to the variance of the random effects. Two, the multinomial covariance
matrix is proportional to the matrix of derivatives of the expectation function.

An iteratively reweighted least squares procedure is used to obtain estimators of the
model parameters. The estimator of the mean is the solution to a set of estimating equations
derived under a multinomial model. An estimator of the random effects variance is obtained
using a linear approximation for the estimator of the mean to derive approximations for mean
squares. A lower bound is used to ensure that the estimator of the variance of the random
effects is strictly positive.

Initial estimators for the iteration are obtained using an extension of the SPREE pro-
cedure (Purcell and Kish, 1980). The initial estimator of the mean is the maximum likelihood
estimator under an assumption that the vector of direct estimators of cell totals for a province is
a multinomial random variable. The initial estimator of the between area variance component,
\( \psi \), is a GLS estimator that depends on a linear approximation.

Initial predictors of the proportions and totals are estimators of the optimal linear predictors under a multinomial model for the covariance matrix of the vector of sampling errors. The initial predictors of the proportions sum to 1 and, therefore, preserve the direct estimators of the province two digit totals. A raking operation is used to benchmark the initial predictors to the national three digit totals. Raking is computationally simple, leads to predictors that remain in the parameter space, and has good efficiency in simulations that represent the LFS data.

**Section 5: MSE estimators**

A Taylor MSE estimator is derived using an extension of the Lahiri and Rao (1995) approach. (Lahiri and Rao (1995) show that the Prasad and Rao (1990) estimator retains the desired order of accuracy even if the random effects are not normally distributed.) The assumptions of Lahiri and Rao (1995) differ from the model and estimation framework of Section 3 and Section 4 in many ways. Lahiri and Rao (1995) consider a linear mixed model with normally distributed sampling errors and assume that the sampling variances are known. The estimator of the linear expected value in Lahiri and Rao (1995) is an estimator of the BLUP, and the estimator of the random effects variance is an unbiased method of moments estimator.

In our context, the sampling variances are unknown, the expectation function is non-linear, and a multinomial model is used as a working model for the covariance matrix of the vector of sampling errors. In the linear model with the BLUP, the error in the predictor constructed with the true parameters is uncorrelated with the error in the BLUP predictor. If the multinomial model for the sampling variances is true, then the error in the predictor constructed with the true parameters is approximately uncorrelated with the error in \( \hat{p}_{T,ik} \), and the Taylor MSE estimator has a simpler form (Section 5.1.1.1). The Taylor MSE estimator does not assume that the working model for the sampling covariance matrix is the true model and includes terms for the variances of the estimators of the working model parameters. The Taylor MSE estimator is an estimator of the variance of the initial convex combination of the
direct estimator and the estimator of $p_{T,ik}$.

A bootstrap MSE estimator is investigated as a way to account for raking. The bootstrap procedure is related to the wild bootstrap method (Liu, 1988; Hall and Maiti, 2006). In our implementation of the wild bootstrap, we specify bootstrap distributions such that the bootstrap versions of the true values and direct estimators remain in the natural parameter spaces. The bootstrap versions of the direct estimators of the totals are conditionally unbiased for the bootstrap versions of the true totals. The sampling variances for the totals under the bootstrap distribution are equal to the direct estimators of the sampling variances for the totals. The proposed data generation procedure does not preserve the full covariance matrix for the totals and distorts the second moments for the proportions. To account for variability in the direct estimators of the sampling variances, bootstrap versions of the direct estimators of the sampling variances are generated from Wishart distributions.

In Section 5.3, benefits and drawbacks of the Taylor and bootstrap MSE estimators are discussed. The Taylor MSE estimator uses the design consistent direct estimators of the sampling variances and is simple to compute. The bootstrap MSE estimator may be preferable when raking or the use of the lower bound for the estimator of $\psi$ has a considerable effect on the MSE.

Section 6: Simulation

The empirical properties of the proposed predictors and MSE estimators are investigated through two sets of simulations in Section 6. In one set of simulations, the data are generated from a model that represents a simple random sampling design, and the covariance matrix of the vector of sampling errors is a multinomial covariance matrix. The sampling error covariance for the second set of simulations approximates the covariance matrix of a two stage sample design and differs from a multinomial covariance matrix. For each sampling error model, two values of the between area variance parameter, $\psi$, are used.

For all of the models considered, the MC MSE’s of the predictors are smaller than the MC MSE’s of the direct estimators. The ratios of the MC MSE’s of the predictors to the MC MSE’s of the direct estimators increase as the province sample sizes increase because the
weight assigned to the direct estimator in the initial predictor approaches one. The MC MSE of the benchmarked predictor is larger than the MC MSE of the initial convex combination of the direct estimator and the estimator of $p_{T,ik}$. The effect of benchmarking on the MC MSE is usually less than 10% for the models considered. For the models considered, use of the working model to construct the predictors is more efficient than use of the direct estimators of the sampling covariance matrices. In our simulations, a multinomial model for the sampling covariance matrix is either true or close to true, and the variances of the direct estimators of the variances are large.

The Taylor and bootstrap MSE estimators have good empirical properties despite their theoretical flaws. The MC relative biases are usually smaller than 15% and are often smaller than 5%. The empirical coverages of nominal 95% prediction intervals for proportions and totals are between 93% and 95% for most of the simulation models.

**Section 7: LFS application**

Canada’s National Occupational Classification system organizes employment into occupations using a hierarchical structure. The ten occupations at the top level of the hierarchy are labeled with the one digit codes, A-J. Each one digit occupation is subdivided into more specific occupations labeled with two digit codes (eg., A1). The two digit occupations are further subdivided into categories labeled with three digit codes (eg., A11 - A14). The procedures of Sections 4 and 5 are applied to obtain predictors for two-way tables defined by the cross-classifications of three digit codes and provinces for the two digit codes A1 (specialist managers) and E0 (judges, lawyers, psychologists, social workers, ministers of religion, and policy and program officers).

The Labour Force Survey provides a direct estimate of the sampling covariance matrix for the proportions based on a jackknife estimate of the sampling covariance matrix for the totals. The direct estimates of the sampling variances and covariances are correlated with functions of the estimates of proportions. Predictors are constructed using the multinomial model as a working model for the sampling variances. Because we are concerned about bias due to misspecification of the model for the sampling variances, MSE estimates are constructed
with the direct estimate of the sampling covariance matrix.

Estimates of the model parameters are obtained under the reduced model in which the estimates of the coefficients assigned to the Census interactions are restricted to equal 1. The two digit codes in a single one digit code are combined to estimate the variance of the random effects under an assumption that the variance parameter $\psi$ is the same for all two digit codes in a particular one digit code.

The estimated MSE’s of the predictors of the totals and proportions are smaller than the estimated sampling variances for most of the cells in the two-way tables for A1 and E0. Estimated coefficients of variation for the predictors of the proportions are uniformly smaller than the estimated coefficients of variation for the direct estimators for both A1 and E0. The estimated gains in efficiency are on the order of 70% – 80% in the small provinces, where the weight assigned to the synthetic estimator in forming the initial predictor is on the order of 85% – 90%. The percent differences between the estimated MSE’s of the predictors and the estimated sampling variances are larger for E0 than for A1 because the estimated ratios of the variances of the random effects to the sampling variances are typically smaller for E0 than for A1. The reductions in the MSE’s due to the prediction procedure are reflected in narrower confidence interval widths.

Standardized residuals, an approximate chi-squared test, and a comparison with the 2006 Census are used to assess the fit of the specified model. For A1, the chi-squared test indicates potential lack of fit. An examination of the residuals for A1 suggests that the model fits well for most of the cells and that the departures from the specified model are greatest in Nova Scotia. The chi-squared test and the standardized residuals are functions of the inverse of the direct estimate of the sampling covariance matrix. The direct estimates of the sampling variances and covariances are small when the direct estimates of the proportions are small. Evidence of lack of fit may result from instability in the direct estimator of the sampling covariance matrix rather than lack of fit of the model. The absolute differences between the predictors of the proportions in Nova Scotia and the Census 2006 proportions are smaller than the absolute differences between the direct estimates of the proportions and the Census 2006
proportions. The comparison with the 2006 Census lends strong support for the predictors in A1. For E0, the standardized residuals, the approximate chi-squared test, and the comparison with the 2006 Census suggest that the model fits well.

8.2 Future Study

In Section 1.2, we identified several challenges arising in the LFS application. Benchmarked predictors are required that preserve the natural parameter spaces for totals and proportions. The direct estimates of the sampling variances are correlated with the direct estimates of the means, and the variances of the direct estimators of the variances are unknown. Because of the raking operation and the use of the estimated variances to form the predictors, MSE estimation is not straightforward. The results of the simulation of Section 6 and the application of Section 7 demonstrate that the proposed methods are reasonable. Naturally, our investigation raises further questions and reveals areas for future work. In this section, we revisit some of the challenges discussed in Section 1.2 and identify potential areas for improvement and future study.

Restricted Parameter Spaces and Benchmarking

In Section 3, the model for the mean is an extension of the model underlying structure preserving estimation (SPREE). Our interest in the SPREE model is driven largely by the interests of the investigators at Statistics Canada. In future work, we plan to explore other model forms for the mean. For example, in Section 6.7, a model for the mean is suggested in which logits of the expected values of the true proportions are assumed to be proportional to the logits of the Census proportions. The model of Section 6.7 has fewer parameters than the SPREE model, and limited investigations with the LFS data suggest that the simpler model fits well for many two-way tables of interest.

A raking operation is used for the LFS data to obtain predictors that preserve the direct estimators of the marginal totals and remain in the parameter space. An alternative benchmarking procedure based on an augmented model is proposed in Section 4.2.3. In Section 6.7, the efficiencies of the augmented model predictors of Section 4.2.3 are compared to the
efficiencies of the raked predictors. The MC MSE’s of predictors based on the augmented model are smaller than the MC MSE’s of the raked predictors for a range of sample sizes. Further study of the augmented model predictor is needed to better understand the properties of this benchmarking procedure.

**Estimated Sampling Variances**

In the LFS application, the direct estimators of the sampling variances are correlated with the direct estimators of the means. Because the distribution of the direct estimator of the sampling variance is unknown, validating a model for the sampling variances is not straightforward. In the simulations of Section 6, the MC MSE’s of predictors constructed under a working model for the sampling variances are often smaller than the MC MSE’s of predictors calculated with the direct estimators of the sampling variances. Because MSE estimators based on the working model are biased if the model is not true, the direct estimators of the sampling variances are used for MSE estimation.

In Section 7, it is conjectured that some estimated MSE’s are larger than the direct estimates of the sampling variances because the direct estimates of the sampling variances for those cells are too small. If the direct estimator of the sampling covariance matrix is unreliable, then modeling the sampling variances may lead to efficiency gains. Incorporating the direct estimators into the model structure is a potential area for future work.

**MSE Estimation**

Two estimators of the MSE are discussed in Section 5. First, an estimator of the MSE of the initial (not benchmarked) predictor is derived using Taylor linearization. A bootstrap procedure is proposed to account for the effect of benchmarking.

The proposed bootstrap procedure has several problems. The data generating method reproduces the direct estimators of the variances of the totals but does not preserve the variances of the proportions. Also, in the bootstrap procedure, a direct estimator of the sampling variances is generated from a Wishart distribution to account for variability in the direct estimators of the sampling variances. A comparison of alternative distributions to be used in the bootstrap is an area for future study.
APPENDIX 1  Linear approximation to the operation that converts totals to proportions

To define the linear approximation to the operation that converts \( \hat{M}_k \) to \( \hat{p}_k \), we note that by a first order Taylor expansion about the true total \( M_{ik} \),

\[
\hat{p}_{ik} - p_{ik} \approx -T_k^{-1} p_{ik}(\hat{M}_k - T_k) + T_k^{-1}(\hat{M}_k - M_{ik}) + O_p(||T_k^{-1} a_{ik}||^2) = l'_{ik} a_k,
\]

where

\[
l'_{ik} = [T_k^{-1}(d'_{C,i} - p_{ik}J'_C)], \tag{A.1}
\]

\( d_{j,l} \) is the \( j \)-dimensional column vector with a 1 in the \( l \)th entry (\( 1 \leq l \leq j \)) and zeros elsewhere. In vector form,

\[
e_k \approx L_k a_k, \tag{A.2}
\]

where \( L_k \) is the \( C \times C \) matrix with \( l'_{ik} \) in row \( i \) for \( i = 1, \ldots, C \). Let \( \hat{L}_k \) be an estimator of \( L_k \) obtained by evaluating the true proportions and totals in (A.1) at the direct estimators, and let \( \hat{L} \) be the block diagonal matrix with \( \hat{L}_k \) as the \( k \)th block. Then, a direct estimator of the sampling variance, \( e \), based on the linear approximation in (A.2) is given by

\[
\hat{\Sigma}_{ee} = \hat{L} \hat{V}_{JK,T} \hat{L}', \tag{A.3}
\]

where \( \hat{V}_{JK,T} \) denotes the LFS jackknife estimator of the sampling variance of the direct estimators of the totals.
APPENDIX 2 Justification of the Linear Approximation for $\hat{\lambda}$

Let $\hat{\psi} = \hat{\psi}_K$ and $\hat{\lambda} = \hat{\lambda}_K$ denote the estimators of $\psi$ and $\lambda_o$ based on $K$ provinces obtained on the final step of the iterative procedure of Section 4.1. For simplicity, assume that $\{c_k : k = 1, \ldots\}$ is a sequence of known constants and $\{\Sigma_{ee,k} : k = 1, \ldots\}$ is a sequence of known matrices. (In practice $c_k$ and $\Sigma_{ee,k}$ are unknown and estimated.) Define

$$S_{w,k}(\delta) = (X_k^{(1)})'(c_k n_k^{-1} + \psi)^{-1}(\hat{p}_k^{(1)} - p_{T,k}^{(1)}(\lambda)),$$

where $X_k^{(1)}$ is a $(C - 1) \times C$ matrix with $x'_{ik} = (I[i = 2], \ldots, I[i = C], (\alpha \beta_{cen})_i)$ in row $i$ for $i = 1, \ldots, C - 1$, $\hat{p}_k^{(1)} = (\hat{p}_{2k}, \ldots, \hat{p}_{Ck})'$, $p_{T,k}^{(1)} = (p_{T,2k}(\lambda), \ldots, p_{T,Ck}(\lambda))'$, $\delta = (\lambda', \psi)'$, and (as defined previously) $\lambda = (\alpha_2, \ldots, \alpha_C, \theta)'$. The sequence of estimators $\{\hat{\delta}_K : K = 1, \ldots\}$ satisfies

$$K^{-1} \sum_{k=1}^{K} S_{w,k}(\hat{\delta}_K) = 0,$$

where $\hat{\delta}_K = (\hat{\lambda}_K', \hat{\psi}_K)'$. Assume that $\hat{\delta}_K - \delta_o = o_p(1)$, where $\delta_o = (\lambda_o', \psi_o)'$ is the vector of true parameters. Also, assume

$$\lim_{K \to \infty} K^{-0.5} \sum_{k=1}^{K} W_k' (\hat{p}_k^{(1)} - p_{T,k}^{(1)}(\lambda_o)) \xrightarrow{L} N(0, V_{ww,\infty}),$$

(A.5)

where $V_{ww,\infty} = \lim_{K \to \infty} K^{-1} \sum_{k=1}^{K} W_k' (\psi^{(1)}(\Gamma_{uu,k}^{(1)} + \Sigma_{ee,k}^{(1)}) W_k)$. In particular,

$$K^{-0.5} \sum_{k=1}^{K} S_{w,k}(\delta_o) \xrightarrow{L} N(0, V_{\infty}),$$

(A.6)

where $V_{\infty} = \lim_{K \to \infty} K^{-1} \sum_{k=1}^{K} (X_k^{(1)})'(c_k n_k^{-1} + \psi)^{-2}(\psi_o \Gamma_{uu,k}^{(1)} + \Sigma_{ee,k}^{(1)}) X_k^{(1)}$. The matrices $\Gamma_{uu,k}^{(1)}$ and $\Sigma_{ee,k}^{(1)}$ denote $(C - 1) \times (C - 1)$ submatrices of $\Gamma_{uu,k}$ and $\Sigma_{ee,k}$ obtained by omitting
the first row and column of $\Gamma_{uu,k}$ and $\Sigma_{ee,k}$. The variance of the limiting distribution, $V_{\infty}$, is assumed to be positive definite. Then,

$$K^{0.5}(\hat{\lambda}_K - \lambda_o) \xrightarrow{L} N(0, V_{\lambda\lambda,\infty}), \quad (A.7)$$

where $V_{\lambda\lambda,\infty} = H_{\infty}^{-1}V_{\infty}H_{\infty}^{-1}$. A sequence of consistent estimators of $V_{\lambda\lambda,\infty}$ is $\{\hat{V}_{\lambda\lambda,K} : K = 1, \ldots\}$, where

$$\hat{V}_{\lambda\lambda,K} = \hat{H}_{xx,K}^{-1}\hat{V}_{ss,K}\hat{H}_{xx,K}, \quad (A.8)$$

and

$$\hat{V}_{ss,K} = K^{-1}\sum_{k=1}^{K} (X_{k}^{(1)})'(c_k n_k^{-1} + \tilde{\psi}_K)^{-2}(\tilde{\psi}_K \hat{\Gamma}_{uu,k}^{(1)} + \Sigma_{ee,k})X_{k}^{(1)}.$$

Justification of (A.7)

Let $S_w(\delta) = K^{-1}\sum_{k=1}^{K} S_{w,k}(\delta)$. The estimator $\hat{\delta}_K$ satisfies $S_w(\hat{\delta}_K) = 0$. The partial derivatives of $S_w(\delta)$ with respect to $\lambda$ and $\psi$ evaluated at $\lambda_o$ are given by,

$$\frac{\partial S_w(\delta_o)}{\partial \lambda} = -K^{-1}\sum_{k=1}^{K} (X_{k}^{(1)})'(c_k n_k^{-1} + \psi_o)^{-1}\Gamma_{uu,k}^{(1)}X_{k}^{(1)} = O_p(1),$$

and

$$\frac{\partial S_w(\delta_o)}{\partial \psi} = -K^{-1}\sum_{k=1}^{K} (X_{k}^{(1)})'(c_k n_k^{-1} + \psi_o)^{-2}(\tilde{p}_K - p_{T,k}(\lambda_o)) = O_p(K^{-0.5}), \quad (A.9)$$

where the orders follow from (A.5) and (A.4). By Taylor’s theorem,

$$0 = S_w(\hat{\delta}_K) = S_w(\delta_o) + \frac{\partial S_w(\delta^*_K)}{\partial \lambda}(\hat{\lambda}_K - \lambda_o) + \frac{\partial S_w(\delta^*_K)}{\partial \psi}(\hat{\psi}_K - \psi_o),$$

where $\delta^*_K = (\lambda^*_K, \psi^*_K)'$ is on the line segment joining $\hat{\delta}_K$ and $\delta_o$. By continuity of the derivatives of $S_w(\delta)$ and the order of the approximation in (A.9),

$$K^{0.5}S_w(\delta_o) = -K^{0.5}\frac{\partial S_w(\delta_o)}{\partial \lambda}(\hat{\lambda}_K - \lambda_o) + o_p(1),$$
and

$$K^{0.5}(\lambda_K - \lambda_o) = -K^{0.5}[\frac{\partial S_w(\delta_o)}{\partial \lambda}]^{-1} S_w(\delta_o) + o_p(1)$$

$$= K^{0.5}[H_\infty]^{-1} S_w(\delta_o) + o_p(1).$$

The result in (A.7) then follows from (A.5).
APPENDIX 3 Initial Estimator of $\lambda_0$

The initial estimator of $\lambda_0$ is the maximum likelihood estimator under an assumption that the vectors of direct estimators of totals for the provinces are independent multinomial random vectors. The likelihood function under a model that specifies the vectors of direct estimators of totals to have independent multinomial distributions is

$$L(\lambda) \propto \prod_{k=1}^{K} \prod_{i=1}^{C} p_{T,ik}(\lambda)^{\hat{M}_{ik}}. \quad (A.10)$$

To define the score function corresponding to the likelihood (A.10), let $p_T(\lambda) = \left( p_{T,1}(\lambda), \ldots, p_{T,K}(\lambda) \right)$, where $p_{T,k}(\lambda) = \left( p_{T,1k}(\lambda), \ldots, p_{T,Ck}(\lambda) \right)$, and let $\hat{p}$ be the vector of direct estimators of proportions listed in the corresponding order. Let $\hat{D}_{prov}$ be a diagonal matrix with the vector $(\hat{M}_{1J_C}', \ldots, \hat{M}_{KJ_C}')$ on the diagonal, where $J_C$ denotes a $C \times 1$ vector of ones. The score function corresponding to the model (A.10) is

$$s(\lambda) = X' \hat{D}_{prov}(\hat{p} - p_T(\lambda)), \quad (A.11)$$

where $X$ is the $CK \times C$ matrix with $x_{ik}' = (I[i = 2], \ldots, I[i = C], (\alpha\beta)_{ik}^{cen})$ in row $C(k - 1) + i$ (the row corresponding to category $i$ and province $k$). The first $C - 1$ elements of the $C \times 1$ vector $s(\lambda)$ are \{\sum_{k=1}^{K} \hat{M}_{ik} - \hat{M}_{ik}p_{T,ik}(\lambda), i = 2, \ldots, C\}, and the last element is \{\sum_{k=1}^{K} \hat{M}_{ik} \sum_{i=1}^{C} (\alpha\beta)_{ik}^{cen} (\hat{p}_{ik} - p_{T,ik}(\lambda))\}. The initial estimator $\hat{\lambda}_0$ satisfies $s(\hat{\lambda}_0) = 0$.

The initial estimator $\hat{p}_{T,ik} = p_{T,ik}(\hat{\lambda}_0)$ is approximately linear in the direct estimators of the cell totals. To justify the linear approximation, we assume that $\hat{\lambda}_0$ converges in probability to $\lambda_0$ as $K$, increases. By a Taylor expansion of the score function in (A.11) around the true parameter $\lambda_0$ and the derivative of $p_T(\lambda)$ with respect to $\lambda$ is,

$$\hat{p}_T - p_T \approx W_{p,u} \hat{D}_{prov}u + W_{p,d}a, \quad (A.12)$$
where $\tilde{p}_T^0 = ((\tilde{p}_{T,1}^0)', \ldots, (\tilde{p}_{T,K}^0)')$, $\tilde{p}_T^0, k = (p_{T,1k}(\hat{\lambda}^0), \ldots, p_{T,Ck}(\hat{\lambda}^0))'$, $p_T = p_T(\lambda_o)$,

\[
W'_{p,u} = \Gamma_{uu}X'(X' D_{prov} \Gamma_{uu}X)^{-1} X',
\]
\[
W'_{p,d} = W'_{p,u}[I_{CK} - \text{diag}(p_T)(I_K \otimes J'_C) \otimes J_C],
\]

$D_{prov} = \text{diag}([T_1, \ldots, T_K]) \otimes I_C$. 
APPENDIX 4

Distortion of Covariances in Bootstrap Data Generating Procedure

The method proposed in section 5.5 to generate the bootstrap versions of the direct estimators of the cell totals from the bootstrap true totals and the original estimated sampling variances does not preserve the correlation structure in the direct estimator of the sampling covariance matrix for the totals. The data generation method, commonly termed Normal to Anything (NORTA), has been studied extensively (e.g., Chen, 2001; Cario and Nelson, 1997). Several modifications to NORTA have been proposed to reduce the distortion of the correlations, some of which use simulations to solve systems of nonlinear equations. Erhardt (2009) proposes a method to generate multivariate count data with an arbitrary correlation structure. The Erhardt (2009) method is implemented in the R package corcounts, and we considered using the Erhardt (2009) method in our bootstrap data generating procedure. To our knowledge, the corrections to NORTA are computationally demanding and are not guaranteed to converge. Therefore, we prefer the simple but approximate NORTA method. In the bootstrap data generating procedure of Section 5.5, the totals generated from NORTA are converted to proportions. The covariance matrix for the proportions under the bootstrap distribution is not equal to the original estimate of the covariance matrix even if the covariance matrix of the generated totals has the specified structure. This section provides a limited investigation of the nature of the distortion for totals and proportions using a covariance matrix that was generated from the two stage simulation model (described in Chapter 6) with $\psi = 0.003$.

**Totals**

As defined (4.6), $\hat{\Sigma}_{aa,k,md}$ denotes the modified direct estimator of the sampling covariance matrix of the totals, and let $M^*_k$ be the vector of bootstrap versions of the true totals for
the $k^{th}$ province. The objective is to generate $\hat{M}_k^*$ such that

$$E_x[\hat{M}_k^* | M_k^*, \hat{\Sigma}_{aa,k,md}] = M_k^*,$$

$$V_x\{\hat{M}_k^* | M_k^*, \hat{\Sigma}_{aa,k,md} \} = \hat{\Sigma}_{aa,k,md},$$

and the elements of $\hat{M}_k^*$ are all nonnegative. If $\hat{\Sigma}_{aa,k,md}$ is diagonal, then the method proposed in section 5.5 attains the desired result. Otherwise, the off diagonal elements of $V_x\{\hat{M}_k^* | M_k^*, \hat{\Sigma}_{aa,k,md} \}$ are not equal to the off-diagonal elements of $\hat{\Sigma}_{aa,k,md}$.

We compare the original estimator, $\hat{\Sigma}_{aa,k,md}$ to the variance of the generated $\hat{M}_k^* - M_k^*$ for one set of estimates obtained in the two stage simulation with $\psi = 0.003$. The bootstrap data generating procedure was run 50,000 times starting with the sample estimates to obtain $\hat{M}_k^{*(b)}$ and $M_k^{*(b)}$ for $b = 1, \ldots, 50,000$. The estimates of $p_{T,ik}$, the marginal two digit totals, and the realized province sample sizes used to generate $\hat{M}_k^{*(b)}$ and $M_k^{*(b)}$ for $b = 1, \ldots, 50,000$ are provided in Table A.1. The off-diagonal elements of the matrices in the left set of columns of Table A.2 are the correlations associated with $\hat{\Sigma}_{aa,k,md}$ for $k = 2, 6,$ and 10 (New Foundland, Manitoba, and Ontario). The diagonal elements of the matrices in the left set of columns of Table A.2 are the square roots of the diagonal elements of $\Sigma_{aa,k,md}$. The right set of columns of Table A.2 contains standard deviations and correlations from the of the empirical covariance matrix,

$$V_{sim,k} = \frac{1}{4999} \sum_{b=1}^{50000} (\hat{M}_k^{*(b)} - M_k^{*(b)} - \bar{M}_d)(\hat{M}_k^{*(b)} - M_k^{*(b)} - \bar{M}_d)' ,$$  \hspace{1cm} (A.13)

where

$$\bar{M}_d = \frac{1}{50000} \sum_{b=1}^{50000} (\hat{M}_k^{*(b)} - M_k^{*(b)}).$$
Table A.1 Estimates $\hat{p}_{T,ik}$ and $\hat{M}_k$ and realized sample sizes from a single simulation run from the two-stage simulation with $\psi = 0.003$

<table>
<thead>
<tr>
<th></th>
<th>PE</th>
<th>NF</th>
<th>NB</th>
<th>NS</th>
<th>SK</th>
<th>MB</th>
<th>BC</th>
<th>AB</th>
<th>QC</th>
<th>ON</th>
</tr>
</thead>
<tbody>
<tr>
<td>$i = 1$</td>
<td>0.444</td>
<td>0.415</td>
<td>0.376</td>
<td>0.394</td>
<td>0.402</td>
<td>0.424</td>
<td>0.390</td>
<td>0.368</td>
<td>0.420</td>
<td>0.371</td>
</tr>
<tr>
<td>$i = 2$</td>
<td>0.150</td>
<td>0.123</td>
<td>0.140</td>
<td>0.089</td>
<td>0.133</td>
<td>0.108</td>
<td>0.146</td>
<td>0.129</td>
<td>0.162</td>
<td>0.165</td>
</tr>
<tr>
<td>$i = 3$</td>
<td>0.273</td>
<td>0.264</td>
<td>0.315</td>
<td>0.313</td>
<td>0.326</td>
<td>0.311</td>
<td>0.306</td>
<td>0.351</td>
<td>0.340</td>
<td>0.340</td>
</tr>
<tr>
<td>$i = 4$</td>
<td>0.133</td>
<td>0.197</td>
<td>0.168</td>
<td>0.204</td>
<td>0.139</td>
<td>0.158</td>
<td>0.159</td>
<td>0.152</td>
<td>0.078</td>
<td>0.125</td>
</tr>
<tr>
<td>$\hat{M}_k$</td>
<td>533</td>
<td>1496</td>
<td>3912</td>
<td>5953</td>
<td>6838</td>
<td>10910</td>
<td>30931</td>
<td>33299</td>
<td>76153</td>
<td>141505</td>
</tr>
<tr>
<td>$n_k$</td>
<td>9</td>
<td>12</td>
<td>25</td>
<td>36</td>
<td>55</td>
<td>71</td>
<td>79</td>
<td>76</td>
<td>118</td>
<td>213</td>
</tr>
</tbody>
</table>

Qualitatively, the differences between the correlations of the original covariance matrix on the left side of Table A.2 and the correlations of the simulated covariance matrix on the right side of Table A.2 are not large. The data generating procedure maintains the order and sign of the correlations of the original matrix.

As verification that the method preserves the variances of the totals, the ratios of the diagonal elements of $V_{sim,k}$ to the diagonal elements of $\Sigma_{aa,k,md}$ are reported in Table A.3 with corresponding MC standard errors in parentheses. The entries in Table A.3 are the squares of the ratios of the standard errors given in Table A.2. With the exception of three digit code $i = 2$ in Ontario, the ratios in Table A.3 are within two MC standard errors of one.
<table>
<thead>
<tr>
<th></th>
<th>Starting Matrix</th>
<th>Simulated Matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( i = 1 )</td>
<td>( i = 2 )</td>
</tr>
<tr>
<td>New Foundland (( k = 2 ))</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>( i = 1 )</td>
<td>213</td>
</tr>
<tr>
<td></td>
<td>( i = 2 )</td>
<td>-0.2618</td>
</tr>
<tr>
<td></td>
<td>( i = 3 )</td>
<td>0.5909</td>
</tr>
<tr>
<td></td>
<td>( i = 4 )</td>
<td>-0.0525</td>
</tr>
<tr>
<td>Manitoba (( k = 6 ))</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>( i = 1 )</td>
<td>959</td>
</tr>
<tr>
<td></td>
<td>( i = 2 )</td>
<td>-0.1016</td>
</tr>
<tr>
<td></td>
<td>( i = 3 )</td>
<td>0.0679</td>
</tr>
<tr>
<td></td>
<td>( i = 4 )</td>
<td>-0.1631</td>
</tr>
<tr>
<td>Ontario (( k = 10 ))</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>( i = 1 )</td>
<td>6336</td>
</tr>
<tr>
<td></td>
<td>( i = 2 )</td>
<td>-0.0073</td>
</tr>
<tr>
<td></td>
<td>( i = 3 )</td>
<td>-0.1233</td>
</tr>
<tr>
<td></td>
<td>( i = 4 )</td>
<td>-0.1342</td>
</tr>
</tbody>
</table>

Table A.2 Standard deviations and correlations of the original estimator of the covariance matrix of the sampling errors in the totals (starting matrix) and the covariance matrix of the simulated totals (simulated matrix).
<table>
<thead>
<tr>
<th></th>
<th>PE</th>
<th>NF</th>
<th>NB</th>
<th>NS</th>
<th>SK</th>
<th>MB</th>
<th>AB</th>
<th>BC</th>
<th>QC</th>
<th>ON</th>
</tr>
</thead>
<tbody>
<tr>
<td>$i = 1$</td>
<td>0.9972</td>
<td>1.0034</td>
<td>1.0045</td>
<td>1.0024</td>
<td>0.9989</td>
<td>1.0003</td>
<td>1.0029</td>
<td>1.0041</td>
<td>0.9938</td>
<td>0.9993</td>
</tr>
<tr>
<td>(SE)</td>
<td>(0.0076)</td>
<td>(0.0074)</td>
<td>(0.0077)</td>
<td>(0.0072)</td>
<td>(0.0067)</td>
<td>(0.0067)</td>
<td>(0.0066)</td>
<td>(0.0068)</td>
<td>(0.0065)</td>
<td>(0.0064)</td>
</tr>
<tr>
<td>$i = 2$</td>
<td>1.001</td>
<td>1.003</td>
<td>1.0043</td>
<td>1.0039</td>
<td>0.9978</td>
<td>1.0061</td>
<td>1.0010</td>
<td>0.9970</td>
<td>0.9887</td>
<td>1.0182</td>
</tr>
<tr>
<td>(SE)</td>
<td>(0.011)</td>
<td>(0.010)</td>
<td>(0.0087)</td>
<td>(0.0087)</td>
<td>(0.0071)</td>
<td>(0.0074)</td>
<td>(0.0074)</td>
<td>(0.0075)</td>
<td>(0.0071)</td>
<td>(0.0068)</td>
</tr>
<tr>
<td>$i = 3$</td>
<td>0.9997</td>
<td>1.0086</td>
<td>0.9872</td>
<td>0.9952</td>
<td>1.0015</td>
<td>0.9961</td>
<td>0.9982</td>
<td>1.0055</td>
<td>0.9909</td>
<td>0.9959</td>
</tr>
<tr>
<td>(SE)</td>
<td>(0.0083)</td>
<td>(0.0087)</td>
<td>(0.0079)</td>
<td>(0.0071)</td>
<td>(0.0069)</td>
<td>(0.0065)</td>
<td>(0.0068)</td>
<td>(0.0068)</td>
<td>(0.0066)</td>
<td>(0.0064)</td>
</tr>
<tr>
<td>$i = 4$</td>
<td>1.009</td>
<td>0.998</td>
<td>0.9858</td>
<td>1.0050</td>
<td>1.0010</td>
<td>0.9852</td>
<td>1.0092</td>
<td>0.9979</td>
<td>0.9959</td>
<td>1.0014</td>
</tr>
<tr>
<td>(SE)</td>
<td>(0.022)</td>
<td>(0.013)</td>
<td>(0.0082)</td>
<td>(0.0067)</td>
<td>(0.0085)</td>
<td>(0.0075)</td>
<td>(0.0077)</td>
<td>(0.0070)</td>
<td>(0.0080)</td>
<td>(0.0069)</td>
</tr>
</tbody>
</table>

Table A.3  Ratios of diagonal elements of (A.13) to the diagonal elements of the generating matrix for totals

**Proportions**

In the discussion of the simulation results, it was conjectured that the bootstrap and Taylor estimators of the leading terms in the MSE’s of the predictors of the proportions differ because the conditional covariance matrix of $\hat{p}_k^* - p_k^*$ given the observed data differs from the original estimator, $\hat{\Sigma}_{ee,k,md}$, defined in (4.7). The organization of Tables A.4 and A.5 is analogous to the organization of Tables A.2 and A.3. The left set of columns of Table A.4 contains the standard deviations and correlations of $\hat{\Sigma}_{ee,k,md}$, and the right set of columns contains the empirical standard deviations and correlations of $\hat{p}_k^{*(b)} - p_k^{*(b)}$. The ratios of the sample variances of $\{\hat{p}_k^{*(b)} - p_k^{*(b)}\}$ to the diagonal elements of $\hat{\Sigma}_{ee,k}$ are shown in Table A.5.

In the examples in Table A.4, the data generating method preserves the order and sign of the correlations. The distortion of the variances for the proportions is considerably larger than the distortion for totals, especially in the smaller provinces. The empirical variances of the generated proportions differ from the diagonal elements of $\hat{\Sigma}_{ee,k}$ by more than 10% in all categories of Prince Edward Island, in three digit codes $i = 3$ and $i = 4$ of New Foundland, and in three digit code $i = 4$ of Nova Scotia.
<table>
<thead>
<tr>
<th></th>
<th>Starting Matrix</th>
<th></th>
<th></th>
<th></th>
<th>Simulated Matrix</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><em>i</em> = 1</td>
<td><em>i</em> = 2</td>
<td><em>i</em> = 3</td>
<td><em>i</em> = 4</td>
<td><em>i</em> = 1</td>
<td><em>i</em> = 2</td>
<td><em>i</em> = 3</td>
<td><em>i</em> = 4</td>
</tr>
<tr>
<td>New Foundland (<em>k</em> = 2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>i</em> = 1</td>
<td>0.1054</td>
<td>-0.6325</td>
<td>0.2500</td>
<td>-0.4629</td>
<td>0.1073</td>
<td>-0.5768</td>
<td>0.0977</td>
<td>-0.4577</td>
</tr>
<tr>
<td><em>i</em> = 2</td>
<td>-0.6325</td>
<td>0.0833</td>
<td>0.3162</td>
<td>-0.2928</td>
<td>-0.5768</td>
<td>0.0830</td>
<td>0.2088</td>
<td>-0.2734</td>
</tr>
<tr>
<td><em>i</em> = 3</td>
<td>0.2500</td>
<td>0.3162</td>
<td>0.1054</td>
<td>-0.9258</td>
<td>0.0977</td>
<td>0.2088</td>
<td>0.0955</td>
<td>-0.8209</td>
</tr>
<tr>
<td><em>i</em> = 4</td>
<td>-0.4629</td>
<td>-0.2928</td>
<td>-0.9258</td>
<td>0.1708</td>
<td>-0.4577</td>
<td>-0.2734</td>
<td>-0.8209</td>
<td>0.1503</td>
</tr>
<tr>
<td>Manitoba (<em>k</em> = 6)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>i</em> = 1</td>
<td>0.0664</td>
<td>-0.4429</td>
<td>-0.4156</td>
<td>-0.6100</td>
<td>0.0658</td>
<td>-0.4354</td>
<td>-0.4378</td>
<td>-0.5855</td>
</tr>
<tr>
<td><em>i</em> = 2</td>
<td>-0.4429</td>
<td>0.0343</td>
<td>0.0000</td>
<td>-0.0891</td>
<td>-0.4354</td>
<td>0.0342</td>
<td>-0.0041</td>
<td>-0.0994</td>
</tr>
<tr>
<td><em>i</em> = 3</td>
<td>-0.4156</td>
<td>0.0000</td>
<td>0.0424</td>
<td>-0.2700</td>
<td>-0.4378</td>
<td>-0.0041</td>
<td>0.0439</td>
<td>-0.2769</td>
</tr>
<tr>
<td><em>i</em> = 4</td>
<td>-0.6100</td>
<td>-0.0891</td>
<td>-0.2700</td>
<td>0.0550</td>
<td>-0.5855</td>
<td>-0.0994</td>
<td>-0.2769</td>
<td>0.0541</td>
</tr>
<tr>
<td>Ontario (<em>k</em> = 10)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>i</em> = 1</td>
<td>0.0387</td>
<td>-0.2613</td>
<td>-0.5585</td>
<td>-0.2917</td>
<td>0.0387</td>
<td>-0.2671</td>
<td>-0.5580</td>
<td>-0.2883</td>
</tr>
<tr>
<td><em>i</em> = 2</td>
<td>-0.2613</td>
<td>0.0297</td>
<td>-0.2813</td>
<td>-0.2760</td>
<td>-0.2671</td>
<td>0.0299</td>
<td>-0.2780</td>
<td>-0.2791</td>
</tr>
<tr>
<td><em>i</em> = 3</td>
<td>-0.5585</td>
<td>-0.2813</td>
<td>0.0392</td>
<td>-0.2958</td>
<td>-0.5580</td>
<td>-0.2780</td>
<td>0.0391</td>
<td>-0.2952</td>
</tr>
<tr>
<td><em>i</em> = 4</td>
<td>-0.2917</td>
<td>-0.2760</td>
<td>-0.2958</td>
<td>0.0311</td>
<td>-0.2883</td>
<td>-0.2791</td>
<td>-0.2952</td>
<td>0.0310</td>
</tr>
</tbody>
</table>

Table A.4  Standard deviations and correlations of the original estimator of the covariance matrix of the sampling errors in the proportions (starting matrix) and the covariance matrix of the simulated proportions (simulated matrix).
Table A.5  Ratios of variances of \( \hat{p}_{ik} - \hat{p}_{ik} \) to diagonal elements of \( \hat{\Sigma}_{ee,k,md} \)

<table>
<thead>
<tr>
<th></th>
<th>PE</th>
<th>NF</th>
<th>NB</th>
<th>NS</th>
<th>SK</th>
<th>MB</th>
<th>BC</th>
<th>AB</th>
<th>QC</th>
<th>ON</th>
</tr>
</thead>
<tbody>
<tr>
<td>i = 1</td>
<td>0.8711</td>
<td>1.0371</td>
<td>0.9303</td>
<td>1.0727</td>
<td>1.0062</td>
<td>0.9835</td>
<td>1.0264</td>
<td>1.0053</td>
<td>0.9649</td>
<td>1.0005</td>
</tr>
<tr>
<td>(SE)</td>
<td>(0.0051)</td>
<td>(0.0063)</td>
<td>(0.0054)</td>
<td>(0.0065)</td>
<td>(0.0063)</td>
<td>(0.0060)</td>
<td>(0.0064)</td>
<td>(0.0062)</td>
<td>(0.0060)</td>
<td>(0.0063)</td>
</tr>
<tr>
<td>i = 2</td>
<td>0.6489</td>
<td>0.9932</td>
<td>0.9899</td>
<td>0.9815</td>
<td>1.0137</td>
<td>0.9953</td>
<td>0.9791</td>
<td>0.9904</td>
<td>0.9660</td>
<td>1.0102</td>
</tr>
<tr>
<td>(SE)</td>
<td>(0.0041)</td>
<td>(0.0084)</td>
<td>(0.0073)</td>
<td>(0.0076)</td>
<td>(0.0069)</td>
<td>(0.0067)</td>
<td>(0.0068)</td>
<td>(0.0063)</td>
<td>(0.0065)</td>
<td></td>
</tr>
<tr>
<td>i = 3</td>
<td>1.1870</td>
<td>0.8213</td>
<td>0.9080</td>
<td>0.9819</td>
<td>0.9863</td>
<td>1.0724</td>
<td>0.9855</td>
<td>0.9879</td>
<td>0.9972</td>
<td>0.9947</td>
</tr>
<tr>
<td>(SE)</td>
<td>(0.0083)</td>
<td>(0.0045)</td>
<td>(0.0076)</td>
<td>(0.0069)</td>
<td>(0.0062)</td>
<td>(0.0060)</td>
<td>(0.0063)</td>
<td>(0.0062)</td>
<td>(0.0063)</td>
<td></td>
</tr>
<tr>
<td>i = 4</td>
<td>0.5842</td>
<td>0.7741</td>
<td>1.0051</td>
<td>1.2396</td>
<td>0.9770</td>
<td>0.9679</td>
<td>0.9773</td>
<td>1.0080</td>
<td>0.9817</td>
<td>0.9968</td>
</tr>
<tr>
<td>(SE)</td>
<td>(0.0051)</td>
<td>(0.0053)</td>
<td>(0.0076)</td>
<td>(0.0088)</td>
<td>(0.0068)</td>
<td>(0.0065)</td>
<td>(0.0066)</td>
<td>(0.0068)</td>
<td>(0.0074)</td>
<td>(0.0067)</td>
</tr>
</tbody>
</table>

Part of the reason that the variance of \( \hat{p}_{ik} - \hat{p}_{ik} \) differs from the \( i \)th diagonal element of \( \hat{\Sigma}_{ee,k,md} \) is related to the linear approximation used to convert the direct estimator of the covariance matrix for the totals to the direct estimator of the covariance matrix for the proportions. A linear approximation for

\[
e_k^{(b)} = \hat{p}_k - p_e^{(b)}
\]

is

\[
\hat{e}_k^{(b)} = M_k^{-1} (\hat{M}_k^{(b)} - p_k^{(b)} \hat{M}_k^{(b)}).
\]

(A.14)

For small \( \hat{\psi} \), an approximation for the variance of \( \hat{e}_k^{(b)} \) is

\[
\hat{\Sigma}_{ee,k} = \hat{L}_{T,k} \hat{\Sigma}_{aa,k,md} \hat{L}_{T,k}^T.
\]

(A.15)

where \( \hat{L}_{T,k} \) is the matrix of the linear approximation defined in (A.2) of Appendix 1 evaluated at the synthetic estimator, \( \hat{T}_{ik} = \hat{p}_{T,ik} \hat{M}_k \). The original direct estimators of the sampling variances are obtained by evaluating the matrix of the linear approximation at the direct estimators of the totals instead of at the synthetic estimators.

Table A.6 contains the ratios of the variances of \( e_k^{(b)} \) to the diagonal elements of \( \hat{\Sigma}_{ee,k} \) defined in (A.15), where the \( e_k^{(b)} \) are generated from the covariance matrix for the totals used for Table A.5. (The MC standard errors for Table A.6 are similar in magnitude to the MC
In Table A.5, the ratio of the variance of \( \epsilon_{4k}^* \) to \( \tilde{\sigma}_{e,4k,md}^2 \) in Nova Scotia is 1.24, while the corresponding ratio in Table A.6 is 1.03. (Note that the denominator in Table A.5 is \( \tilde{\sigma}_{e,ik,md}^2 \), while the denominators in Table A.6 are the diagonal elements of \( \tilde{\Sigma}_{ee,k} \) defined in (A.15)). The variances of \( \epsilon_{ik}^* \) in the smallest two provinces differ significantly from the corresponding diagonal elements of \( \tilde{\Sigma}_{ee,k} \).

The ratios of the variances of \( e_k^* \) to the diagonal elements of \( \tilde{\Sigma}_{ee,k} \) are closer to one in the large provinces than in the small provinces, in part, because the linear approximation (A.14) improves as the province sample sizes increase. Table A.7 shows the ratios of the variances of the elements of the linearized variable defined in (A.14) to the diagonal elements of \( \tilde{\Sigma}_{ee,k} \) defined in (A.15). The variance of variables of the linearized variables for Prince Edward Island in are significantly closer to the diagonal elements of \( \tilde{\Sigma}_{ee,k} \) than are the variances of the elements of \( e_k^* \). The comparison of Table A.7 to Table A.6 demonstrates that the small sample size effect on the quality of the linear approximation in the small provinces largely explains why the ratios in Table A.6 differ significantly from 1 in the small provinces.

Table A.8 contains the ratios of the variances of the linearized variables \( e_k^* \) to the diagonal elements of the original direct estimator of the covariance matrix. As we expect, variances of the linearized variables for Prince Edward Island differ more from the diagonal elements of the original \( \tilde{\Sigma}_{ee,k} \) defined in (4.7) than they do from the diagonal elements of \( \tilde{\Sigma}_{ee,k} \) defined in (A.15). (The ratios in Table A.8 are farther from 1 than the corresponding ratios in Table A.7.)

The differences between the variances of the elements of \( e_k^* \) and the diagonal elements of \( \tilde{\Sigma}_{ee,k} \) are due to more than MC variability. The variance of \( e_k^* \) is

\[
E_s[ V_s(\tilde{e}_k^* | \mathbf{u}_k^*) ] = E_s[ L_k^* \tilde{\Sigma}_{aa,k,md} \tilde{L}_k^* ],
\]

where \( L_k^* \) is the matrix of the linear approximation defined in (A.2) evaluated at \( M_{ik}^* \). For small \( \psi \),

\[
E[ L_k^* \tilde{\Sigma}_{aa,k,md} \tilde{L}_k^* ] \approx \tilde{\Sigma}_{ee,k}.
\]

The variances of the elements of \( e_k^* \) also differ from the diagonal elements of \( \tilde{\Sigma}_{ee,k} \) because
variances of the elements of \( \tilde{e}_k^{(b)} \) depend on covariances between \( \tilde{M}_{ik}^{(b)} - M_{ik}^{(b)} \) and \( \tilde{M}_{jk}^{(b)} - M_{jk}^{(b)} \) for \( i \neq j \), and the data generating procedure for the totals does not preserve the original covariances of \( \tilde{\Sigma}_{aa,k,md} \).

<table>
<thead>
<tr>
<th></th>
<th>PE</th>
<th>NF</th>
<th>NB</th>
<th>NS</th>
<th>SK</th>
<th>MB</th>
<th>BC</th>
<th>AB</th>
<th>QC</th>
<th>ON</th>
</tr>
</thead>
<tbody>
<tr>
<td>( i = 1 )</td>
<td>0.7214</td>
<td>0.9908</td>
<td>0.9226</td>
<td>0.9660</td>
<td>0.9752</td>
<td>0.9797</td>
<td>0.9991</td>
<td>0.9692</td>
<td>0.9815</td>
<td>0.9971</td>
</tr>
<tr>
<td>( i = 2 )</td>
<td>0.6443</td>
<td>0.9700</td>
<td>0.9879</td>
<td>0.9639</td>
<td>0.9936</td>
<td>0.9952</td>
<td>0.9723</td>
<td>0.9810</td>
<td>0.9753</td>
<td>0.9951</td>
</tr>
<tr>
<td>( i = 3 )</td>
<td>1.0364</td>
<td>0.7846</td>
<td>0.8984</td>
<td>0.9761</td>
<td>0.9816</td>
<td>1.0066</td>
<td>0.9744</td>
<td>0.9783</td>
<td>1.0006</td>
<td>1.0015</td>
</tr>
<tr>
<td>( i = 4 )</td>
<td>0.4212</td>
<td>0.7599</td>
<td>0.9748</td>
<td>1.0329</td>
<td>0.9228</td>
<td>0.9739</td>
<td>0.9744</td>
<td>0.9747</td>
<td>0.9929</td>
<td>0.9997</td>
</tr>
</tbody>
</table>

Table A.6 Ratios of variances of \( e_k^{(b)} \) to diagonal elements of \( \tilde{\Sigma}_{ee,k} \).

<table>
<thead>
<tr>
<th></th>
<th>PE</th>
<th>NF</th>
<th>NB</th>
<th>NS</th>
<th>SK</th>
<th>MB</th>
<th>BC</th>
<th>AB</th>
<th>QC</th>
<th>ON</th>
</tr>
</thead>
<tbody>
<tr>
<td>( i = 1 )</td>
<td>1.0403</td>
<td>1.0402</td>
<td>0.9988</td>
<td>1.0088</td>
<td>0.9919</td>
<td>1.0079</td>
<td>1.0136</td>
<td>0.9977</td>
<td>1.0034</td>
<td>1.0039</td>
</tr>
<tr>
<td>( i = 2 )</td>
<td>0.9168</td>
<td>0.9847</td>
<td>1.0120</td>
<td>0.9880</td>
<td>0.9951</td>
<td>1.0074</td>
<td>0.9930</td>
<td>0.9931</td>
<td>1.0050</td>
<td>1.0032</td>
</tr>
<tr>
<td>( i = 3 )</td>
<td>1.0914</td>
<td>0.9762</td>
<td>0.9836</td>
<td>1.0056</td>
<td>1.0065</td>
<td>1.0060</td>
<td>1.0060</td>
<td>1.0008</td>
<td>1.0131</td>
<td>1.0101</td>
</tr>
<tr>
<td>( i = 4 )</td>
<td>0.9740</td>
<td>0.9636</td>
<td>0.9950</td>
<td>1.0106</td>
<td>0.9758</td>
<td>1.0116</td>
<td>1.0101</td>
<td>0.9882</td>
<td>1.0089</td>
<td>1.0064</td>
</tr>
</tbody>
</table>

Table A.7 Ratios of variances of \( \tilde{e}_k^{(b)} \) to diagonal elements of \( \tilde{\Sigma}_{ee,k} \).

<table>
<thead>
<tr>
<th></th>
<th>PE</th>
<th>NF</th>
<th>NB</th>
<th>NS</th>
<th>SK</th>
<th>MB</th>
<th>BC</th>
<th>AB</th>
<th>QC</th>
<th>ON</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.2499</td>
<td>1.0925</td>
<td>1.0033</td>
<td>1.1244</td>
<td>1.0120</td>
<td>1.0202</td>
<td>1.0418</td>
<td>1.0178</td>
<td>1.0040</td>
<td>1.0095</td>
</tr>
<tr>
<td>2</td>
<td>0.9209</td>
<td>1.0036</td>
<td>1.0143</td>
<td>0.9884</td>
<td>1.0155</td>
<td>1.0081</td>
<td>0.9946</td>
<td>0.9968</td>
<td>1.0109</td>
<td>1.0060</td>
</tr>
<tr>
<td>3</td>
<td>1.2606</td>
<td>1.0110</td>
<td>0.9927</td>
<td>1.0123</td>
<td>1.0095</td>
<td>1.0789</td>
<td>1.0259</td>
<td>1.0019</td>
<td>1.0241</td>
<td>1.0152</td>
</tr>
<tr>
<td>4</td>
<td>1.3235</td>
<td>0.9713</td>
<td>1.0364</td>
<td>1.2123</td>
<td>1.0083</td>
<td>1.0199</td>
<td>1.0118</td>
<td>1.0035</td>
<td>1.0089</td>
<td>1.0087</td>
</tr>
</tbody>
</table>

Table A.8 Ratios of variances of \( \tilde{e}_k^* \) to direct estimators of sampling variances
A way to generate variables so that the variance of $\hat{e}_k^{*\{b\}}$ is approximately equal to the original direct estimator of the sampling covariance matrix for the proportions is to generate the totals from a different covariance matrix. Define

$$\hat{\Sigma}_{aa,k} = \hat{p}_{T,k}(\hat{p}_{T,k})' \hat{V}\{\hat{M}_k\} + \hat{M}_k^2 \hat{\Sigma}_{ee,k,md}.$$  

Then, $\hat{L}_{T,k} \hat{\Sigma}_{aa,k} \hat{L}_{T,k}' = \hat{\Sigma}_{ee,k,md}$, and the variance of $\hat{e}_k^{*\{b\}}$ is approximately equal to the original $\hat{\Sigma}_{ee,k,md}$ of (4.7).

Table A.9 shows the ratios of the variances of $\hat{e}_k^{*\{b\}}$ to the diagonal elements of $\hat{\Sigma}_{ee,k,md}$ when the totals are generated from $\hat{\Sigma}_{aa,k}$. As we expect, the ratios in Table A.9 are closer to 1 than the corresponding ratios in A.8. The differences between the variances of $\hat{e}_k^{*\{b\}}$ generated from $\hat{\Sigma}_{aa,k}$ and the diagonal elements of $\hat{\Sigma}_{ee,k,md}$ arise from the distortion of the covariance matrix for the totals and the effect of the variance of $u_{ik}$, as discussed above.

<table>
<thead>
<tr>
<th></th>
<th>PE</th>
<th>NF</th>
<th>NB</th>
<th>NS</th>
<th>SK</th>
<th>MB</th>
<th>BC</th>
<th>AB</th>
<th>QC</th>
<th>ON</th>
</tr>
</thead>
<tbody>
<tr>
<td>$i = 1$</td>
<td>1.0810</td>
<td>1.0559</td>
<td>1.0025</td>
<td>0.9945</td>
<td>1.0094</td>
<td>0.9909</td>
<td>1.0077</td>
<td>1.0020</td>
<td>1.0122</td>
<td>1.0136</td>
</tr>
<tr>
<td>$i = 2$</td>
<td>0.9030</td>
<td>0.9998</td>
<td>1.0156</td>
<td>0.9953</td>
<td>1.0122</td>
<td>1.0053</td>
<td>0.9899</td>
<td>1.0114</td>
<td>1.0014</td>
<td>0.9879</td>
</tr>
<tr>
<td>$i = 3$</td>
<td>1.1273</td>
<td>0.9505</td>
<td>1.0021</td>
<td>1.0095</td>
<td>1.0036</td>
<td>1.0062</td>
<td>0.9920</td>
<td>1.0000</td>
<td>1.0031</td>
<td>1.0012</td>
</tr>
<tr>
<td>$i = 4$</td>
<td>0.9639</td>
<td>0.9711</td>
<td>1.0128</td>
<td>1.0077</td>
<td>0.9870</td>
<td>0.9850</td>
<td>1.0062</td>
<td>1.0013</td>
<td>1.0081</td>
<td>1.0038</td>
</tr>
</tbody>
</table>

Table A.9  Ratios of variances of $\hat{e}_k^{*\{b\}}$ to direct estimators of sampling variances
APPENDIX 5
Distortion of Covariances in Bootstrap Data Generating Procedure

In Section 5, we propose two estimators of the MSE’s of the predictors of the proportions. A closed form MSE estimator (referred to as a Taylor MSE) is based on a linear approximation for the MSE of the “initial predictor” (the predictor that is not benchmarked to the direct estimators of the marginal totals). As an attempt to account for raking, a bootstrap MSE estimator is proposed. Each MSE estimator has benefits and drawbacks. In this appendix, we compare the bootstrap and Taylor estimates obtained in the LFS application. First, we compare the bootstrap and Taylor estimates of the variances of the model parameters, \( \hat{c}_k \), \( \hat{\psi} \), and \( \hat{\lambda} \). We then compare the properties of the bootstrap and Taylor estimates of the MSE’s of the predictors of the proportions. We find that the raking operation contributes to differences between the two types of MSE estimates. Because the final benchmarking operation has noticeable effects on the predictors for several cells in the LFS application, we use the bootstrap MSE estimators for the data analysis of Section 7.

Taylor and Bootstrap Variances of Model Parameters

Table A.10 contains Taylor and bootstrap standard errors of the estimators of the variance parameters \( \{c_k : k = 1, \ldots, 10\} \). The bootstrap estimator of the variance of \( \hat{c}_k \) is the sample variance of the 2000 bootstrap estimates of \( c_k \). The Taylor estimator of the variance of \( \hat{c}_k \) is defined in (4.30). The ratios of the bootstrap estimates of the variances of \( \{\hat{c}_k : k = 1, \ldots, 10\} \) to the corresponding Taylor estimates of the variances range from 0.34 (Nova Scotia) to 3.57 (Quebec). The Taylor standard error defined in (4.30) uses an assumption that the variance of \( \left[ (C_{ts} - 1)(n_{k,ts} - 1) \right]^{0.5} \hat{c}_{k,ts} \) is constant. Use of the \( (C_{ts} - 1)(n_{k,ts} - 1) \) for the standardization is based on the variance under a Wishart distribution. The residual plots
in Figure 7.4 used to validate the model underlying the Taylor estimator of the variance of \( \hat{c}_k \) contains several outliers. Otherwise, the plot does not suggest serious departures from the assumptions underlying the estimator (4.30). The bootstrap version of the sampling variance has a Wishart distribution that is independent of the bootstrap versions of the direct estimators of the proportions. We suspect that some of the differences between the Taylor and bootstrap variance estimators for \( \hat{c}_k \) arise because the bootstrap variance estimator uses the Wishart assumption more heavily. In the simulations, the bootstrap variance estimator performs poorly when the Wishart assumption does not hold. Therefore, we prefer the Taylor estimators of the variances of \( \{ \hat{c}_k : k = 1, \ldots, 10 \} \).

<table>
<thead>
<tr>
<th></th>
<th>PE</th>
<th>NF</th>
<th>NB</th>
<th>NS</th>
<th>SK</th>
<th>MB</th>
<th>BC</th>
<th>AB</th>
<th>QC</th>
<th>ON</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{c}_k )</td>
<td>1.311</td>
<td>1.293</td>
<td>1.169</td>
<td>1.145</td>
<td>1.115</td>
<td>1.235</td>
<td>1.436</td>
<td>1.530</td>
<td>1.670</td>
<td>1.497</td>
</tr>
<tr>
<td>Taylor SE</td>
<td>0.063</td>
<td>0.064</td>
<td>0.031</td>
<td>0.057</td>
<td>0.045</td>
<td>0.050</td>
<td>0.038</td>
<td>0.061</td>
<td>0.021</td>
<td>0.029</td>
</tr>
<tr>
<td>Bootstrap SE</td>
<td>0.061</td>
<td>0.049</td>
<td>0.030</td>
<td>0.033</td>
<td>0.029</td>
<td>0.038</td>
<td>0.034</td>
<td>0.036</td>
<td>0.040</td>
<td>0.037</td>
</tr>
</tbody>
</table>

Table A.10 Estimates of \( c_k : k = 1, \ldots, 10 \) with Taylor and bootstrap standard error estimates

The estimates of the parameters of model (1), Taylor MSE estimates, and bootstrap MSE estimates are shown in Tables A.11 and A.12. In A1, the bootstrap estimate of the variance of \( \hat{\psi} \) is larger than the Taylor variance estimate. The estimator of \( \hat{\psi} \) uses inverses of matrices that are functions of \( \hat{p}_{T,ik} \). In three digit code A36, the \( \{ \hat{p}_{T,ik} : k = 1, \ldots, 10 \} \) are relatively small. We conjecture that the bootstrap estimates of the inverses of the matrices used to compute the bootstrap versions of the estimates of \( \psi \) for the one digit code A are highly variable because the small probabilities in A36 and the unstable estimates of the inverses lead to extreme bootstrap estimates of \( \psi \). In E0, the two variance estimates are similar. The bootstrap and Taylor estimates of the variance of \( \hat{\lambda} \) are essentially the same.
Table A.11  Estimates of $\psi$ for A1 and E0 with bootstrap and Taylor standard error estimates

<table>
<thead>
<tr>
<th>2-digit</th>
<th>$\hat{\psi}$</th>
<th>bootstrap</th>
<th>Taylor</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>0.0062</td>
<td>0.0027</td>
<td>0.0023</td>
</tr>
<tr>
<td>E0</td>
<td>0.0033</td>
<td>0.0026</td>
<td>0.0028</td>
</tr>
</tbody>
</table>

Table A.12  Estimates of $\lambda_\circ$ for A1 and E0 with bootstrap and Taylor standard error estimates. E0 has 3 categories, so $\alpha_{\circ,4}$ is not in the model for E0.

<table>
<thead>
<tr>
<th>2-digit</th>
<th>$\alpha_{\circ,2}$</th>
<th>$\alpha_{\circ,3}$</th>
<th>$\alpha_{\circ,4}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1 Estimate</td>
<td>-0.96</td>
<td>-0.10</td>
<td>-0.53</td>
</tr>
<tr>
<td>A1 Taylor SE</td>
<td>0.15</td>
<td>0.11</td>
<td>0.15</td>
</tr>
<tr>
<td>A1 Bootstrap SE</td>
<td>0.15</td>
<td>0.11</td>
<td>0.16</td>
</tr>
<tr>
<td>E0 Estimate</td>
<td>1.30</td>
<td>1.11</td>
<td>-</td>
</tr>
<tr>
<td>E0 Taylor SE</td>
<td>0.11</td>
<td>0.11</td>
<td>-</td>
</tr>
<tr>
<td>E0 Bootstrap SE</td>
<td>0.11</td>
<td>0.11</td>
<td>-</td>
</tr>
</tbody>
</table>

Taylor and Bootstrap Estimates of Prediction MSE’s

In Figures A.1 and A.2, the bootstrap estimates of the MSE’s for the proportions for A1 and E0, respectively, are plotted on the vertical axis with the corresponding Taylor MSE estimates on the horizontal axis. The solid line in each figure is a 45 degree line through the origin. In A1, the bootstrap MSE estimates are uniformly larger than the corresponding Taylor MSE estimates. In E0, most of the Taylor MSE estimates are larger than the corresponding bootstrap MSE estimates. The MSE estimates for E0 are closer to the 45 degree line than the MSE estimates for A1.

Tables A.13 and A.14 show the percent differences between the bootstrap MSE es-
estimates and the Taylor MSE estimates relative to the Taylor MSE estimates for A1 and E0, respectively. The percent difference is defined

$$100 \left( \frac{\hat{\text{MSE}}_{ik}^{bs} - \hat{\text{MSE}}_{2,ik,B}}{\hat{\text{MSE}}_{2,ik,B}} \right).$$

(A.16)

The percent differences are larger in absolute value for A1 than for E0. The largest percent difference in A1 of 33% occurs in three digit code A11 in Nova Scotia. The absolute values of the percent differences in E0 are smaller than 10% and most are smaller than 5%. The percent difference with the largest absolute value occurs in E03 of Prince Edward Island.

Figure A.1 Taylor MSE estimates (x-axis). Bootstrap MSE estimates (y-axis). Solid line is $y = x$. Two digit code A1.
Table A.13 Percent differences between bootstrap MSE’s and Taylor MSE’s relative to Taylor MSE’s for A1

<table>
<thead>
<tr>
<th></th>
<th>PE</th>
<th>NF</th>
<th>NB</th>
<th>NS</th>
<th>SK</th>
<th>MB</th>
<th>BC</th>
<th>AB</th>
<th>QC</th>
<th>ON</th>
</tr>
</thead>
<tbody>
<tr>
<td>A11</td>
<td>23.26</td>
<td>20.80</td>
<td>23.89</td>
<td>33.22</td>
<td>19.78</td>
<td>27.53</td>
<td>24.05</td>
<td>23.49</td>
<td>13.63</td>
<td>20.65</td>
</tr>
<tr>
<td>A13</td>
<td>17.53</td>
<td>13.84</td>
<td>13.05</td>
<td>20.01</td>
<td>18.06</td>
<td>22.36</td>
<td>19.43</td>
<td>16.19</td>
<td>9.52</td>
<td>7.89</td>
</tr>
<tr>
<td>A14</td>
<td>21.49</td>
<td>17.99</td>
<td>16.17</td>
<td>11.46</td>
<td>18.50</td>
<td>17.80</td>
<td>18.03</td>
<td>18.00</td>
<td>12.99</td>
<td>8.85</td>
</tr>
</tbody>
</table>

Figure A.2 Taylor MSE estimates (x-axis). Bootstrap MSE estimates (y-axis). Solid line is $y = x$. Two digit code E0.
Table A.14 Percent differences between bootstrap MSE’s and Taylor MSE’s relative to Taylor MSE’s for E0

<table>
<thead>
<tr>
<th></th>
<th>PE</th>
<th>NF</th>
<th>NB</th>
<th>NS</th>
<th>SK</th>
<th>MB</th>
<th>BC</th>
<th>AB</th>
<th>QC</th>
<th>ON</th>
</tr>
</thead>
<tbody>
<tr>
<td>E01</td>
<td>-0.60</td>
<td>-3.43</td>
<td>-6.49</td>
<td>1.47</td>
<td>-6.31</td>
<td>1.35</td>
<td>-2.54</td>
<td>-3.79</td>
<td>2.82</td>
<td>-4.65</td>
</tr>
<tr>
<td>E02</td>
<td>5.95</td>
<td>2.86</td>
<td>-1.42</td>
<td>2.25</td>
<td>-1.91</td>
<td>0.65</td>
<td>-2.37</td>
<td>-3.78</td>
<td>-2.68</td>
<td>-6.20</td>
</tr>
<tr>
<td>E03</td>
<td>7.27</td>
<td>0.51</td>
<td>1.79</td>
<td>-1.15</td>
<td>-0.35</td>
<td>-0.90</td>
<td>-2.60</td>
<td>4.60</td>
<td>-5.37</td>
<td></td>
</tr>
</tbody>
</table>

The difference between the two estimates of the variance of $\psi$ partly explain the differences between the bootstrap and Taylor MSE estimates for A1. To investigate the effect of the estimate of the variance of $\psi$ on the differences between the bootstrap and Taylor MSE estimators, the Taylor MSE estimators were recomputed using the bootstrap estimate of the variance of $\psi$. Changing the estimator of the variance of $\psi$ increases the Taylor MSE estimators in A1. A comparison of Tables A.15 and A.13 shows that the effect of using the bootstrap estimate of $\psi$ on the percent differences between the bootstrap and Taylor MSE’s is minimal.

Table A.15 Percent differences between bootstrap and Taylor MSE estimates with Taylor MSE estimates evaluated at the bootstrap estimate of the variance of $\psi$.

<table>
<thead>
<tr>
<th></th>
<th>PE</th>
<th>NF</th>
<th>NB</th>
<th>NS</th>
<th>SK</th>
<th>MB</th>
<th>BC</th>
<th>AB</th>
<th>QC</th>
<th>ON</th>
</tr>
</thead>
<tbody>
<tr>
<td>A14</td>
<td>20.79</td>
<td>17.24</td>
<td>14.91</td>
<td>10.05</td>
<td>16.84</td>
<td>15.97</td>
<td>15.75</td>
<td>15.78</td>
<td>10.42</td>
<td>6.38</td>
</tr>
</tbody>
</table>

Another factor that contributes to the differences between the bootstrap and Taylor MSE estimates is the raking operation. The bootstrap MSE estimates account for the final raking step, while the Taylor MSE’s are estimates of the MSE’s of the initial predictors that
A bootstrap estimator of the MSE of the initial predictor is obtained by replacing $\tilde{p}_{ik,B}^{(b)}$ (bootstrap version of benchmarked predictor) with $\hat{p}_{\text{pred},ik,B}^{(b)}$ (bootstrap version of initial predictor) in the definition of the bootstrap MSE estimator. Let $\hat{\text{MSE}}_{2,ik,NR}^{bs}$ denote the bootstrap estimate of the MSE of the initial predictor. In Figures A.3 and A.4, the percent difference defined in (A.16) is plotted on the vertical axis with the corresponding percent difference between $\hat{\text{MSE}}_{2,ik}^{bs}$ and $\hat{\text{MSE}}_{2,ik,NR}^{bs}$ on the horizontal axis. The plot for A1 reveals a positive association between the two percent differences. The correlations between the two percent differences for A1 is 0.66. The positive association is weaker for E0 than for A1. The correlation between the two percent differences for E0 is 0.0299. The percent differences between the bootstrap and Taylor MSE estimates tend to increase as the percent differences between the bootstrap estimates of the benchmarked and non-benchmarked predictors increase. However, the raking only appears to explain about one third of the difference between the bootstrap and Taylor estimated variances.
Figure A.3  Percent difference between bootstrap estimates of MSE’s of benchmarked predictors and Taylor MSE estimates (y-axis). Percent difference between bootstrap estimates of MSE’s of benchmarked predictors and bootstrap estimates of MSE’s of initial predictors (x-axis) for A1
Figure A.4 Percent difference between bootstrap estimates of MSE’s of benchmarked predictors and Taylor MSE estimates (y-axis). Percent difference between bootstrap estimates of MSE’s of benchmarked predictors and bootstrap estimates of MSE’s of initial predictors (x-axis) for E0

A third source of differences between the bootstrap and Taylor MSE estimates is the estimate of the bias of the leading term. Tables A.16 and A.18 contain the ratios of the bootstrap estimates of the biases of the Taylor estimates of the leading terms to the Taylor estimates of the leading terms for A1 and E0, respectively. Tables A.17 and A.19 contain the ratios of the Taylor estimates of the biases of the Taylor estimates of the leading terms to the Taylor estimates of the leading terms. In A1, the bootstrap estimates of the biases are larger in magnitude than the Taylor estimates of the biases, and the ratios of the two bias estimates
tend to decrease as the province sizes increase. In E0, the bootstrap estimates of the bias are uniformly smaller in magnitude than the Taylor estimates of the bias. The Taylor estimates of the bias are all negative for both E0 and A1. The bootstrap estimates of the biases for E0 in Prince Edward Island, New Foundland, and E01 of Nova Scotia are positive. In the simulations, both bias estimators are inaccurate. The bootstrap estimate of the bias of the estimator of the leading term is better than the Taylor estimator when $\psi$ is small and the lower bound leads to a positive bias in the estimator of the leading term. Because $\hat{\psi} = 0.0033$ for E0, we prefer the bootstrap bias estimate for E0. For A1, the bootstrap estimator of the bias is larger in magnitude than the Taylor estimator of the bias. Using the bootstrap bias estimate for A1 is a conservative choice in this example.

<table>
<thead>
<tr>
<th></th>
<th>PE</th>
<th>NF</th>
<th>NB</th>
<th>NS</th>
<th>SK</th>
<th>MB</th>
<th>BC</th>
<th>AB</th>
<th>QC</th>
<th>ON</th>
</tr>
</thead>
<tbody>
<tr>
<td>A11</td>
<td>-0.125</td>
<td>-0.127</td>
<td>-0.133</td>
<td>-0.133</td>
<td>-0.129</td>
<td>-0.130</td>
<td>-0.133</td>
<td>-0.131</td>
<td>-0.129</td>
<td>-0.118</td>
</tr>
<tr>
<td>A12</td>
<td>-0.124</td>
<td>-0.133</td>
<td>-0.132</td>
<td>-0.134</td>
<td>-0.147</td>
<td>-0.139</td>
<td>-0.135</td>
<td>-0.133</td>
<td>-0.127</td>
<td>-0.104</td>
</tr>
<tr>
<td>A13</td>
<td>-0.129</td>
<td>-0.131</td>
<td>-0.131</td>
<td>-0.132</td>
<td>-0.132</td>
<td>-0.132</td>
<td>-0.131</td>
<td>-0.131</td>
<td>-0.128</td>
<td>-0.113</td>
</tr>
<tr>
<td>A14</td>
<td>-0.126</td>
<td>-0.118</td>
<td>-0.126</td>
<td>-0.132</td>
<td>-0.121</td>
<td>-0.127</td>
<td>-0.126</td>
<td>-0.126</td>
<td>-0.137</td>
<td>-0.114</td>
</tr>
</tbody>
</table>

Table A.16 Ratios of bootstrap estimates of biases of estimates of leading terms to Taylor estimates of leading terms for A1

<table>
<thead>
<tr>
<th></th>
<th>PE</th>
<th>NF</th>
<th>NB</th>
<th>NS</th>
<th>SK</th>
<th>MB</th>
<th>BC</th>
<th>AB</th>
<th>QC</th>
<th>ON</th>
</tr>
</thead>
<tbody>
<tr>
<td>A11</td>
<td>-0.009</td>
<td>-0.009</td>
<td>-0.016</td>
<td>-0.024</td>
<td>-0.024</td>
<td>-0.021</td>
<td>-0.025</td>
<td>-0.025</td>
<td>-0.028</td>
<td>-0.030</td>
</tr>
<tr>
<td>A12</td>
<td>-0.009</td>
<td>-0.015</td>
<td>-0.018</td>
<td>-0.020</td>
<td>-0.014</td>
<td>-0.031</td>
<td>-0.021</td>
<td>-0.022</td>
<td>-0.027</td>
<td>-0.028</td>
</tr>
<tr>
<td>A13</td>
<td>-0.011</td>
<td>-0.009</td>
<td>-0.015</td>
<td>-0.019</td>
<td>-0.021</td>
<td>-0.023</td>
<td>-0.027</td>
<td>-0.025</td>
<td>-0.027</td>
<td>-0.029</td>
</tr>
<tr>
<td>A14</td>
<td>-0.013</td>
<td>-0.011</td>
<td>-0.019</td>
<td>-0.010</td>
<td>-0.017</td>
<td>-0.018</td>
<td>-0.030</td>
<td>-0.026</td>
<td>-0.031</td>
<td>-0.029</td>
</tr>
</tbody>
</table>

Table A.17 Ratios of Taylor estimates of biases of estimates of leading terms to Taylor estimates of leading terms for A1
To compare the bootstrap and Taylor MSE estimates after eliminating the effect of raking and the effect of estimating the bias of the leading term, define
\[
\widehat{\text{MSE}}_{1,ik}^{bs} = B^{-1} \sum_{b=1}^{B} (\hat{p}_{\text{pred},ik,B}^{(b)} - \hat{p}_{ik,B}^{(b)})^2
\]
and the equivalent Taylor MSE estimator,
\[
\widehat{\text{MSE}}_{1,ik}^{T} = \widehat{\text{MSE}}_{1,ik}^{bs} + \hat{g}_{3,ik}.
\]
Tables A.20 and A.21 contain the relative differences
\[
\frac{\widehat{\text{MSE}}_{1,ik}^{bs} - \widehat{\text{MSE}}_{1,ik}^{T}}{\widehat{\text{MSE}}_{1,ik}^{T}}
\]
for A1 and E0, respectively. Removing the effects of raking and estimating the biases of the estimates of the leading terms greatly reduces the percent differences between the bootstrap and Taylor MSE estimates for A1 and for most of the cells in E0. The percent decrease in the
Taylor estimate of the MSE relative to the bootstrap estimate of the MSE in Ontario for E0 is larger when one does not account for the effect of raking or the effect of estimating the leading term.

<table>
<thead>
<tr>
<th>PE</th>
<th>NF</th>
<th>NB</th>
<th>NS</th>
<th>SK</th>
<th>MB</th>
<th>BC</th>
<th>AB</th>
<th>QC</th>
<th>ON</th>
</tr>
</thead>
<tbody>
<tr>
<td>A11</td>
<td>4.30</td>
<td>1.56</td>
<td>3.99</td>
<td>13.22</td>
<td>-0.63</td>
<td>7.20</td>
<td>5.34</td>
<td>6.27</td>
<td>-0.41</td>
</tr>
<tr>
<td>A12</td>
<td>0.54</td>
<td>4.66</td>
<td>9.26</td>
<td>2.03</td>
<td>-5.60</td>
<td>11.51</td>
<td>-3.69</td>
<td>-4.10</td>
<td>-1.23</td>
</tr>
<tr>
<td>A13</td>
<td>-0.02</td>
<td>-1.60</td>
<td>-3.08</td>
<td>2.85</td>
<td>1.16</td>
<td>4.37</td>
<td>3.02</td>
<td>0.49</td>
<td>-3.03</td>
</tr>
<tr>
<td>A14</td>
<td>5.71</td>
<td>1.15</td>
<td>0.28</td>
<td>-2.62</td>
<td>0.13</td>
<td>-0.18</td>
<td>6.16</td>
<td>3.51</td>
<td>2.72</td>
</tr>
</tbody>
</table>

Table A.20 Relative differences between bootstrap and Taylor MSE estimators for A1 after eliminating effects of raking and estimating the biases of the estimators of the leading terms.

<table>
<thead>
<tr>
<th>PE</th>
<th>NF</th>
<th>NB</th>
<th>NS</th>
<th>SK</th>
<th>MB</th>
<th>BC</th>
<th>AB</th>
<th>QC</th>
<th>ON</th>
</tr>
</thead>
<tbody>
<tr>
<td>E01</td>
<td>-1.14</td>
<td>-3.73</td>
<td>-7.34</td>
<td>1.79</td>
<td>-7.22</td>
<td>-0.92</td>
<td>-5.28</td>
<td>-6.22</td>
<td>-2.01</td>
</tr>
<tr>
<td>E02</td>
<td>5.15</td>
<td>1.33</td>
<td>-0.19</td>
<td>1.96</td>
<td>-1.53</td>
<td>0.83</td>
<td>-2.67</td>
<td>-3.96</td>
<td>-1.85</td>
</tr>
<tr>
<td>E03</td>
<td>6.45</td>
<td>0.82</td>
<td>2.33</td>
<td>-0.56</td>
<td>0.43</td>
<td>-2.01</td>
<td>-2.07</td>
<td>-4.36</td>
<td>2.97</td>
</tr>
</tbody>
</table>

Table A.21 Relative differences between bootstrap and Taylor MSE estimators for E0 after eliminating effects of raking and estimating the biases of the estimators of the leading terms.

The above analysis focuses on the proportions. Tables A.22 and A.23 show the percent changes between the bootstrap and Taylor MSE estimates for the totals for A1 and E0, respectively. The percent differences in most cells are larger for proportions than for totals. Examples of cells where the percent changes for totals exceed the percent changes for proportions occur in Quebec and Ontario. That the bootstrap and Taylor MSE estimates are more similar for totals than for proportions in most cells is consistent with the results of the simulation. The bootstrap and Taylor estimates of the leading terms are more similar for totals than for pro-
portions because the bootstrap data generating procedure preserves the direct estimates of the variances of the totals but distorts the variances of the proportions.

<table>
<thead>
<tr>
<th></th>
<th>PE</th>
<th>NF</th>
<th>NB</th>
<th>NS</th>
<th>SK</th>
<th>MB</th>
<th>BC</th>
<th>AB</th>
<th>QC</th>
<th>ON</th>
</tr>
</thead>
<tbody>
<tr>
<td>A11</td>
<td>6.17</td>
<td>5.74</td>
<td>6.23</td>
<td>3.06</td>
<td>7.80</td>
<td>13.67</td>
<td>16.72</td>
<td>13.01</td>
<td>11.15</td>
<td>31.51</td>
</tr>
<tr>
<td>A12</td>
<td>3.11</td>
<td>5.34</td>
<td>17.75</td>
<td>16.47</td>
<td>12.15</td>
<td>21.77</td>
<td>16.15</td>
<td>21.42</td>
<td>18.24</td>
<td>7.91</td>
</tr>
<tr>
<td>A13</td>
<td>3.51</td>
<td>9.50</td>
<td>1.59</td>
<td>5.30</td>
<td>10.39</td>
<td>13.21</td>
<td>5.01</td>
<td>14.53</td>
<td>7.29</td>
<td>6.88</td>
</tr>
<tr>
<td>A14</td>
<td>12.54</td>
<td>9.11</td>
<td>5.14</td>
<td>13.81</td>
<td>12.07</td>
<td>12.88</td>
<td>12.10</td>
<td>13.86</td>
<td>12.51</td>
<td>12.05</td>
</tr>
</tbody>
</table>

Table A.22 Percent differences between bootstrap and Taylor MSE estimates for totals in A1

<table>
<thead>
<tr>
<th></th>
<th>PE</th>
<th>NF</th>
<th>NB</th>
<th>NS</th>
<th>SK</th>
<th>MB</th>
<th>BC</th>
<th>AB</th>
<th>QC</th>
<th>ON</th>
</tr>
</thead>
<tbody>
<tr>
<td>E01</td>
<td>-0.99</td>
<td>-1.01</td>
<td>3.66</td>
<td>7.74</td>
<td>3.61</td>
<td>4.74</td>
<td>6.03</td>
<td>-0.58</td>
<td>19.58</td>
<td>7.82</td>
</tr>
<tr>
<td>E02</td>
<td>2.35</td>
<td>-5.23</td>
<td>7.23</td>
<td>-2.56</td>
<td>-5.66</td>
<td>8.79</td>
<td>5.99</td>
<td>-6.73</td>
<td>-1.23</td>
<td>7.48</td>
</tr>
<tr>
<td>E03</td>
<td>3.86</td>
<td>0.24</td>
<td>-8.07</td>
<td>3.23</td>
<td>1.13</td>
<td>1.92</td>
<td>-0.09</td>
<td>7.15</td>
<td>-3.28</td>
<td>7.87</td>
</tr>
</tbody>
</table>

Table A.23 Percent differences between bootstrap and Taylor MSE estimates for totals in E0
BIBLIOGRAPHY


