



AgEcon SEARCH
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search
<http://ageconsearch.umn.edu>
aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

Aggregated Fractional Regression Estimation: Some Monte Carlo Evidence

Jingyu Song | song173@purdue.edu

Michael S. Delgado | delgado2@purdue.edu

Paul V. Preckel | preckel@purdue.edu

Department of Agricultural Economics, Purdue University, West Lafayette, IN 47907

May 2017

*Selected Paper prepared for presentation at the 2017 Agricultural & Applied Economics Association
Annual Meeting, Chicago, Illinois, July 30-August 1*

Copyright 2017 by Jingyu Song, Michael S. Delgado, and Paul V. Preckel. All rights reserved. Readers may make verbatim copies of this document for non-commercial purposes by any means, provided that this copyright notice appears on all such copies.

Abstract -- We propose a fractional regression framework for problems where individual level fractional outcomes are desired but only aggregate level outcomes are available. Our model is based on the quasi-maximum likelihood method, and links aggregated fractional outcomes with individual attributes to predict individual level fractions. To assess the finite sample performance of our estimation framework, we design two Monte Carlo simulation schemes, one with spatial clustering patterns across individuals, and the other without. We test both schemes with a single-outcome setup and a multi-outcome setup. Our results show that the bias and root mean squared error (RMSE) decrease consistently as the sample size grows for all cases, indicating the reliability of our proposed estimation strategy under different settings. Our estimation framework is generally applicable to cases in which only aggregated level fractional outcomes are available but individual level outcomes are wanted.

Keywords -- aggregate level outcome; individual level share; quasi-maximum likelihood; Monte Carlo simulation; spatial clustering

1. Introduction

There exists a number of economic research questions where the outcome variable of interest, y , is in the fractional form ($0 \leq y \leq 1$). Examples include 401(k) pension plan participation rate (Papke and Wooldridge, 1996), brand valuation (Dubin, 2007), financial asset portfolio shares (Mullahy, 2015), etc. Papke and Wooldridge offer a fractional response regression approach in their 1996 paper tailored for this type of questions. Using a Bernoulli log-likelihood function, the quasi-maximum likelihood estimator (QMLE) of the parameter estimates β are obtained by maximizing the total likelihood:

$$\max_b \sum_{i=1}^N l_i(b) \quad (1)$$

where $l_i(b) = y_i \log[G(x_i b)] + (1 - y_i) \log[1 - G(x_i b)]$, and the $G(\cdot)$ function denotes the univariate conditional mean, i.e. $E(y_i | x_i) = G(x_i b)$.

One advantage of this approach compared with log-odds type procedures is that the dependent variable of interest, the fractional outcomes, can take on extreme values of the bounded range -- zero and one (Papke and Wooldridge, 1996). Following Gourieroux et al. (1984), the QMLE estimator is consistent as long as the likelihood expression is a member of the linear exponential family. The estimation approach is also robust to distributional misspecification (Papke and Wooldridge, 1996).

Mullahy (2015) demonstrates the fractional logit version of the Papke and Wooldridge (1996) model. For the univariate case, the conditional mean becomes:

$$E(s | \mathbf{x}) = G(\mathbf{x}; \boldsymbol{\omega}) = \frac{\exp(\mathbf{x}\boldsymbol{\omega})}{1 + \exp(\mathbf{x}\boldsymbol{\omega})}. \quad (2)$$

Mullahy also expands the univariate case to a more general form of a multivariate fractional regression model. In the multivariate case, the conditional mean can be expressed as:

$$E(s_k | \mathbf{x}) = G_k(\mathbf{x}; \boldsymbol{\beta}) = \frac{\exp(\mathbf{x}\boldsymbol{\beta}_k)}{\sum_{m=1}^M \exp(\mathbf{x}\boldsymbol{\beta}_m)} \quad k = 1, 2, \dots, M. \quad (3)$$

This multivariate structure is suitable for cases where multiple outcome categories are of interest simultaneously. For instance, Mullahy demonstrates in his article models assessing different financial asset categories -- each category takes a share of the entire financial asset portfolio and all categories add up to one.

A similar yet different class of questions is not covered by previous literature and has not been studied widely. In this type of fractional response cases, outcomes are only observable at a more aggregated level (than the individual observational level), while conditioning variables are

available at the individual level. In other words, there is a mismatch between the observable aggregate level outcomes and the individual level conditioning variables. We are interested in finding the slope coefficients, and predicting the individual level fractional outcomes using available data on fractional outcomes at the aggregate level and the conditioning variables at the individual level. However, the above mentioned estimation approaches cannot be used directly due to this mismatch.

One important empirical example representative of this aggregated fractional response problem is the fine-scale land allocation problem. Typically, cropland allocation data are made available by national census or survey instruments, and information on total land shares in various crops at the state/province level are provided. These data, however, do not indicate the distribution of cropland within the states/provinces, especially not for cropland allocation over a wide geographic area across states or countries. This lack of finer than state/province level land use data brings challenges to applied research. Studies have shown that using aggregate level data may mask the heterogeneity across locations that bears critical implications for national and international research and policy (Auffhammer et al., 2013; Hendricks et al., 2014). Estimates at a finer than state/province level are needed.

To enable the estimation, we utilize the relationship between the aggregate and the individual levels, and bring the individual level attribute data up to match the aggregate fractions, so that the univariate/multivariate fractional logit framework can be constructed to estimate the coefficients and predict the individual level outcomes. In contrast to the previous studies, an additional aggregation step is required -- we aggregate outcomes across individual observations to match with the aggregate level outcomes, and then perform the estimation as Papke and Wooldridge (1996) and Mullahy (2015).

This paper serves as the first attempt in dealing with aggregated fractional regression estimation. Following the idea of aggregating individual level shares to match fractional outcomes at the aggregate level, we develop a quasi-maximum likelihood estimation framework that uses individual attribute factors and observed aggregate level fractional outcomes to determine individual level outcomes via aggregation. One advantage of this framework is that the sample size is reduced drastically via aggregation -- instead of evaluating at the individual level, we are now estimating at the aggregate level.

In what follows, we first describe our estimation strategy, formalize the theoretical fractional regression model with aggregation, and derive the likelihood function, in Section 2. Next, in Section 3, we discuss a Monte Carlo simulation framework designed to validate our proposed approach and assess its finite sample performance. We also present an alternative simulation scheme where spatial clustering is introduced to further test the performance of our approach. We present the Monte Carlo results for both simulation setups in Section 4. Section 5 concludes.

2. Theoretical Framework with Aggregation of Outcomes

Consider the following case: fractional outcomes at the aggregate level are available; attributes/conditioning variables at the individual level are available; each individual observation contributes to the aggregate level fractional outcome depending on the way it is related to the aggregate level. Our goal is to predict the fraction that is allocated to a particular outcome at the individual level, given the available individual attribute measurements and the aggregate level fractional outcomes. We introduce an aggregation step to accommodate to this mismatch: we add up individual level fractional outcomes to the aggregate level, and enable the estimation with a fractional response model. Detailed derivations and estimation steps are described in Song et al. (2016). Here, we briefly review the key points of our framework.

Assume there are j aggregate level structures, $j = \{1, 2, \dots, J\}$; within each aggregate structure, there are k fractional outcomes, $k = \{1, 2, \dots, K\}$, and the k fractional outcomes add up to one. y_{jk} stands for the observed aggregate level fraction in aggregate structure j that is in outcome k , such that $0 \leq y_{jk} \leq 1$. Z_{ijk} denotes the unobserved individual level fraction of individual i in aggregate structure j that is in outcome k . Since the number of individual observations in each aggregate structure may vary, we set $i = \{1, 2, \dots, I_j\}$. Let \mathbf{X}_{ij} be an N -dimensional vector of observable individual attributes for individual i in aggregate structure j , and we are interested in estimating the parameters β in the conditional mean for individual level fraction Z_{ijk} :

$$E[Z_{ijk} | \mathbf{X}_{ij}] = G_{ijk}(\mathbf{W}_{ij}(\mathbf{X}_{ij}), \beta_k) \quad (4)$$

where $\mathbf{W}(\cdot): \mathbb{R}^N \rightarrow \mathbb{R}^M$ reflects transformations of the fundamental explanatory variables (such as linear, quadratic, or with interaction), $G(\cdot): \mathbb{R}^M \rightarrow \mathbb{R}$, $0 < G(\cdot) < 1$ is a function that maintains the unit interval restriction on the conditional mean. Following Mullahy (2015), we parameterize

$G(\cdot)$ using a logistic functional form, and the predicted fraction of individual i in outcome k in aggregate structure j becomes:

$$G_{ijk}(\mathbf{W}_{ij}(\mathbf{X}_{ij}), \beta_k) = \frac{\exp(\mathbf{W}_{ij}(\mathbf{X}_{ij})\beta_k)}{\sum_{k=1}^K \exp(\mathbf{W}_{ij}(\mathbf{X}_{ij})\beta_k)} \quad \text{where } \beta_1 = 0. \quad (5)$$

The $\beta_1 = 0$ normalization facilitates parameter identification relative to the base case outcome. We extend equation (5), which is defined at the individual level, to the aggregate level via an aggregation structure – which is, the predicted fraction in outcome k in aggregate structure j is equal to the sum over individual level weighted fractions. The predicted fraction of outcome k in structure j is:

$$H_{jk} = \frac{\sum_{i \in I_j} G_{ijk}(\mathbf{W}_{ij}(\mathbf{X}_{ij}), \beta_k) A_{ij}}{\sum_{i \in I_j} A_{ij}} \quad (6)$$

where A_{ij} is the weight parameter of individual i for aggregate structure j . In other words, function (6) aggregates our predicted individual fractions to the aggregate level, converting individual level information to the more aggregated level, so that the individual level attribute data can be used to explain the aggregate level outcomes. Given H_{jk} , the quasi-log-likelihood function to be maximized with respect to the parameters β_k is:

$$\mathcal{L} = \sum_{j=1}^J \sum_{k=1}^K y_{jk} \ln H_{jk}. \quad (7)$$

This framework is generally applicable to cases in which only aggregate level data is available for the outcome, but individual level estimates are desired. If one specific aggregate level outcome is of interest, or if we only have available data for one particular outcome, then instead of having multiple aggregate level outcomes, we would go back to the Papke and Wooldridge (1996) univariate case. With all the other aggregate level outcomes treated as the base case, the $G(\cdot)$ function can be expressed as:

$$G_{ij}(\mathbf{W}_{ij}(\mathbf{X}_{ij}), \beta) = \frac{\exp(\mathbf{W}_{ij}(\mathbf{X}_{ij})\beta)}{1 + \exp(\mathbf{W}_{ij}(\mathbf{X}_{ij})\beta)} \quad (8)$$

and the predicted fraction at the aggregate level for the interested outcome in aggregate structure j becomes:

$$H_j = \frac{\sum_{i \in I_j} G_{ij}(\mathbf{W}_{ij}(\mathbf{X}_{ij}), \beta) A_{ij}}{\sum_{i \in I_j} A_{ij}}. \quad (9)$$

The quasi-log-likelihood function to be maximized is:

$$\mathcal{L} = \sum_{j=1}^J y_j \ln H_j. \quad (10)$$

In the context of the land use example, fraction of one specific crop in a state/province can be regarded as one aggregate level observed outcome/covariate. Data on total fractions of different crops (such as corn, soybean, wheat, etc.) are available only at the state/province level. The individual level refers to the finer than state/province level, and we name it the grid-cell level. Each aggregate level is consisted of several grid cells. We do not know the fraction of each crop at the grid-cell level, yet we observe the grid-cell level land attributes (such as temperature, precipitation, slope, soil pH, and so on). We can use the fact that grid-cell level land area adds up to the total state/province area to facilitate the aggregation and estimate land shares in each crop for each grid cell.

3. Monte Carlo Designs

We implement Monte Carlo simulations to assess the finite sample performance of our proposed estimation strategy. We consider two designs: one is with spatial clustering, the other is without. We call the Monte Carlo design without spatial clustering the original case, and the one with spatial clustering the alternative case -- with the alternative case where individuals within the same aggregate structure are spatially clustered, we are able to check whether spatial clustering has an impact on our estimation performance. For both cases, we estimate two representative setups: a single-outcome one, with two possible outcomes within each aggregate structure, one is our outcome of interest, the other is the base case; and a multi-outcome one, where multiple outcomes are of interest within each aggregate structure.

We start with the description of the original Monte Carlo design. For the single-outcome case, we assume there are three independent variables including the intercept (denoted as $x_0 \sim x_2$). In empirical analysis, some variables tend to have relatively large variations within and across aggregate structures (such as within and across state/province temperature influencing land shares in different crops), while others have relatively small fluctuations across observations (such as slope in the land use example). We try to characterize both variable types in our simulation. To proceed, we use x_1 to denote the variable with within and across aggregate structure variations; and x_2 to mimic the variable with relatively small variation. In order to capture the variation in x_1 across aggregate structures, we first set a base value for each aggregate structure, which takes value between 0 and 30 (mimicking actual temperature in degrees Celsius). We then generate a random value between $(-2, 2)$ to represent the variation within each aggregate structure (this can

be interpreted as the difference in temperature within each state/province, and the largest difference within a state/province equals 4 degrees Celsius). Both the base variation across aggregate structures and the variation within each structure follow a uniform distribution. Lastly, we add the two parts up to construct variable x_1 . Variable x_2 represents the variable with less variation. We generate it based on a $(0, 1)$ uniform distribution (mimicking slope). We also add an error term following the logistic distribution (location = 0, scale = 0.005) to reflect measurement errors.

We let the number of individual level observations within each aggregate structure vary across the sample. Exact number of observations in each aggregate structure is a randomly generated integer taking a value between 500 and 3000 (representing number of grid cells in a state/province). We assume that the true coefficient values are known: $\beta_0 = 2, \beta_1 = -0.15, \beta_2 = 1$. For simplicity, we set the weight of each individual, A_{ij} , to 100, rather than letting it vary across individuals. Based on the true theta values and the independent variables, we formulate the aggregate level outcomes, which, in empirical examples, are reported and publicly available. Then, we perform the estimation procedure based on equation (10), and repeat the process for a large number of times ($MC = 1000$). To measure the performance of the estimation approach, we compare the differences between estimated coefficient values and the pre-set true coefficient values. Measures considered for the comparison include average bias and average Root Mean Squared Error (RMSE). Average bias is defined as the average difference between the estimated coefficient values and the true coefficient values over 1000 replications. Average RMSE is calculated by taking the square root of the average of the squared difference between the estimated coefficients and the true coefficients over 1000 replications.

For the multi-outcome case, we assume there are three different outcomes that are of interest (plus the base case). For the independent variables, we use the same setup as is used in the single-outcome case; and we let the number of individuals within each aggregate structure vary across the full sample. Again, the minimum number of individuals in an aggregate structure is set at 500, and the maximum is 3000. The pre-set true coefficient values for the first outcome category are: $\beta_{0cat1} = 2, \beta_{1cat1} = -0.15, \beta_{2cat1} = 1$; for the second outcome category: $\beta_{0cat2} = 1.8, \beta_{1cat2} = -0.15, \beta_{2cat2} = 1$; and for the third: $\beta_{0cat3} = 1.5, \beta_{1cat3} = -0.18, \beta_{2cat3} = 1$. Similar to the single-outcome case, the individual level fractions are estimated, and the framework

is replicated for 1000 times. As the last step, we calculate the same two measures, average bias and average RMSE, over the 1000 replications.

We also consider an alternative Monte Carlo design. Taking the land use allocation case for example, in practice, grid cells with similar properties may be clustered geographically because of their similarities in land attributes and the climate. To capture this clustering and assess its potential impact on the predictive power of our framework, we modify our original Monte Carlo simulation setup described above, incorporate spatial clustering, and re-assess the finite sample performance of our framework for both the single-outcome and the multi-outcome cases.

To include spatial clustering, the central idea is to create a weight matrix, and use it to update the independent variables and the associated individual level fractions, so that the clustering pattern is captured by the individual level outcomes. R 3.1.0 provides a convenient package named `spdep` to create weight matrices. By multiplying the independent variables with the corresponding weight matrix, neighboring individuals get similar values. Therefore, we first create the three independent variables in the same way described above for the original setup. We then use the `spdep` package to create a weight matrix for each aggregate structure (with neighbor type: queen). For each weight matrix, the number of columns/rows equals the square root of the number of observations in that aggregate structure so that we keep the comparability for matrix operations. We multiply the previously generated variables by the weight matrix to get new variables that contain spatial clustering patterns. To ensure that the square root of the number of observations is an integer, we modify its generating step. Instead of picking an integer value between 500 and 3000 as the number of observations for an aggregate structure, we constrain the number of individuals to the squared value of an integer between 20 and 55. This guarantees that the number of rows/columns for the weight matrix is still an integer when we take the square root. All the other parts of this alternative Monte Carlo design remain the same as the original version demonstrated previously.

For both the original and the alternative Monte Carlo designs, we test number of aggregate structures $M = 20, 100, 250, 500$ for the single-outcome and the multi-outcome cases, respectively. As a technical note, for all four cases, we use the BFGS optimization method in the `optimx` package in R 3.1.0 as the estimation algorithm. Since it is well-known that the performance of the BFGS method is improved when analytical gradients are supplied (Nash and Varadhan, 2011; Nash, 2014), we input the analytical gradient of the likelihood function into the

R optimization routines. All reported simulations are conducted on a high-performance Linux cluster using dual 8-core Intel Xeon-E5 CPUs.

4. Results

We present the Monte Carlo simulation results for the original single-outcome setup in Table 1, and the results for the original multi-outcome case in Table 2.

Results show that for both setups, as the sample size increases, both the average bias and the average RMSE decrease and approach zero. There is a clear trend that increasing the sample size improves estimation results. We also record the completion time for each setup. It serves as an illustration of the tradeoff between performance improvement brought by increasing the sample size and the increase in computational time caused by larger sample size. Our results indicate that as the sample size grows, computational time grows even faster. Computational burden outgrows the benefits brought by the increase in the sample size in magnitude. Therefore, a reasonably large sample needs to be chosen to satisfy estimation needs without bringing in too heavy computational burdens.

Results for the alternative setup with spatial clustering are shown in Table 3 for the single-outcome case and Table 4 for the multi-outcome case. Similar to the original setups, there is a clear trend that increasing the sample size improves estimation results. As more aggregate structures are added to the simulation, both the average bias and the average RMSE decrease. In terms of the completion time, the alternative setup takes longer than the original one as the additional weight matrix generation process takes time. As the sample size grows, computational time grows even faster. It outgrows the benefits brought by increasing the sample size in magnitude.

5. Conclusions

We propose a novel approach in predicting individual level fractions using aggregate level data and individual level attribute variables. We evaluate the finite sample performance of our framework using Monte Carlo simulations. Two Monte Carlo designs are provided, one with spatial clustering, the other without; and we test both a single-outcome case and a multi-outcome case for both setups. Results show that our method performs well and produces reliable estimates with both small samples and relatively large samples. As sample size gets larger, the coefficient estimates get closer to the true values. However, there is a clear tradeoff between computational

time and improvement in estimation performance. The framework can be applied to any case where aggregate level fractional outcomes are known but individual level outcomes are desired.

Acknowledgements

This research was supported in part by computational resources provided by Information Technology at Purdue – Rosen Center for Advanced Computing, Purdue University, West Lafayette, Indiana.

References

- Auffhammer, M., Hsiang, S.M., Schlenker, W., and Sobel, A. (2013). Using weather data and climate model output in economic analyses of climate change. *Review of Environmental Economics and Policy*, 7(2), 181-198.
- Dubin, J.A. (2007). Valuing intangible asset with a nested logit market share model. *Journal of Econometrics*, 139: 285-302.
- Gourieroux, C., Monfort, A., and Trognon, A. (1984). Pseudo maximum likelihood methods: Theory. *Econometrica*, 52(3), 681-700.
- Hendricks, N.P., Smith, A., and Sumner, D.A. (2014). Crop supply dynamics and the illusion of partial adjustment. *American Journal of Agricultural Economics*, 96(5), 1469-1491.
- Mullahy, J. (2015). Multivariate fractional regression estimation of econometric share models. *Journal of Econometric Methods*, 4(1): 71-100.
- Nash, J. C. (2014). On Best Practice Optimization Methods in R. *Journal of Statistical Software*, 60(2), 1-14.
- Nash, J.C., and Varadhan, R. (2011). Unifying Optimization Algorithms to Aid Software System Users: optimx for R. *Journal of Statistical Software*, 43(9), 1-14.
- Papke, L.E., and Wooldridge, J.M. (1996). Econometric methods for fractional response variables with an application to 401(k) plan participation rates. *Journal of Applied Econometrics*, 11, 619-632.
- Song, J., Delgado, M.S., Preckel, P.V., and Villoria, N.B. (2016). Pixel Level Cropland Allocation and the Impacts of Biophysical Factors, *working paper*.

Table 1. Monte Carlo results on bias (estimates versus true values) and RMSE for the univariate model

Number of Aggregate Structures	Bias			RMSE			Running Time (hours)
	x_0	x_1	x_2	x_0	x_1	x_2	
20	-4.7245E-04	-2.9251E-06	1.0447E-03	3.8870E-03	4.7289E-05	9.3131E-03	1.6490E+00
100	-2.6911E-04	-4.0084E-07	5.5227E-04	1.7074E-03	2.0576E-05	4.0777E-03	5.7855E+00
250	-8.8727E-05	1.9334E-06	1.1243E-04	8.8227E-04	1.0897E-05	2.1078E-03	9.4445E+00
500	-7.7593E-05	2.0047E-06	8.8575E-05	6.2981E-04	7.7029E-06	1.4948E-03	2.6813E+01

Table 2. Monte Carlo results on bias (estimates versus true values) and RMSE for the multivariate model

Outcome	Number of Aggregate Structures	Bias			RMSE			Running Time (hours)
		x_0	x_1	x_2	x_0	x_1	x_2	
<i>One</i>	20	8.8005E-03	1.2389E-04	-2.3205E-02	1.1698E-02	1.4412E-04	2.9276E-02	1.5402E+01
	100	1.1551E-03	1.8534E-05	-3.1225E-03	3.5265E-03	3.3843E-05	8.1832E-03	7.8068E+01
	250	-2.0746E-04	-2.4331E-06	5.7426E-04	1.5877E-03	1.4783E-05	3.6398E-03	2.2927E+02
	500	6.6883E-05	2.1771E-06	-1.9253E-04	8.7402E-04	8.6725E-06	2.0206E-03	4.4194E+02
<i>Two</i>	20	9.5604 E-03	1.2697E-04	-2.4728E-02	1.2641E-02	1.4798E-04	3.1215E-02	
	100	1.4380E-03	1.9463E-05	-3.6750E-03	3.4967E-03	3.2724E-05	7.9834E-03	
	250	-3.0003E-04	-2.9782E-06	7.7557E-04	1.7746E-03	1.4487E-05	3.9161E-03	
	500	1.0221E-04	2.1779E-06	-2.5036E-04	9.5757E-04	8.6994E-06	2.1575E-03	
<i>Three</i>	20	1.8211 E-02	1.5925E-04	-4.2153E-02	2.0914E-02	1.8075E-04	4.8207E-02	
	100	2.6512E-03	2.4188E-05	-6.1015E-03	4.2962E-03	3.5580E-05	9.5673E-03	
	250	-4.5035E-04	-3.4637E-06	1.0949E-03	2.1935E-03	1.7155E-05	4.8875E-03	
	500	2.8095E-04	2.8671E-06	-5.8996E-04	1.2679E-03	9.9205E-06	2.7839E-03	

Table 3. Monte Carlo results on bias (estimates versus true values) and RMSE for the univariate alternative model

Number of Aggregate Structures	Bias			RMSE			Running Time (hours)
	x_0	x_1	x_2	x_0	x_1	x_2	
20	-2.9089E-04	2.4060E-06	5.0456E-04	4.2752E-03	1.0211E-05	8.7256E-03	5.5901E+00
100	-6.3032E-04	1.7920E-06	1.1976E-03	2.5023E-03	4.9919E-06	5.0897E-03	2.9435E+01
250	-5.1147E-05	2.4797E-06	1.9540E-05	1.2273E-03	3.7499E-06	2.5127E-03	6.0566E+01
500	-4.2633E-05	2.5378E-06	1.1292E-06	1.0961E-03	3.3378E-06	2.2396E-03	1.2840E+02

Table 4. Monte Carlo results on bias (estimates versus true values) and RMSE for the multivariate alternative model

Outcome	Number of Aggregate Structures	Bias			RMSE			Running Time (hours)
		x_0	x_1	x_2	x_0	x_1	x_2	
<i>One</i>	20	1.5406E-02	2.2167E-05	-3.1762E-02	1.8742E-02	2.8191E-05	3.8518E-02	1.5886E+01
	100	3.6380E-03	6.1631E-06	-7.5017E-03	7.0301E-03	1.0473E-05	1.4431E-02	8.9454E+01
	250	1.0147E-03	2.5254E-06	-2.0952E-03	2.1874E-03	4.0676E-06	4.4767E-03	2.3696E+02
	500	9.5652E-04	2.1459E-06	-1.9593E-03	2.7485E-03	3.9598E-06	5.6159E-03	4.5251E+02
<i>Two</i>	20	1.5769E-02	2.2205E-05	-3.2479E-02	1.8647E-02	2.8290E-05	3.8317E-02	
	100	3.8915E-03	6.1912E-06	-7.9929E-03	7.5718E-03	1.0519E-05	1.5499E-02	
	250	1.1493E-03	2.4734E-06	-2.3542E-03	2.3680E-03	4.1216E-06	4.4767E-03	
	500	8.2835E-04	1.9613E-06	-1.6915E-03	2.1759E-03	3.5880E-06	4.4581E-03	
<i>Three</i>	20	2.3351E-02	2.6051E-05	-4.7648E-02	2.6543E-02	3.3501E-05	5.4138E-02	
	100	4.5488E-03	6.3592E-06	-9.2880E-03	7.0910E-03	1.0197E-05	1.4499E-02	
	250	1.5330E-03	2.5602E-06	-3.0995E-03	2.9107E-03	4.4806E-06	5.9088E-03	
	500	4.7571E-04	1.6816E-06	-9.6118E-04	2.9607E-03	3.8693E-06	5.9948E-03	