Statistics and Applications {ISSN 2454-7395(online)}
Volume 20, No. 1, 2022 (New Series), pp 17–31

Adjusted Design Effect Model for Individual Variables in Survey Data

Mohammed Quazi and Yan Lu
Department of Mathematics and Statistics
University of New Mexico, Albuquerque, New Mexico, USA

Received: 25 July 2020; Revised: 06 December 2020; Accepted: 28 January 2021

Abstract

In this study we have extended longitudinal generalized variance functions (LGVF) for grouped variables variance estimation to individual variables variance estimation. Effect of survey design and change of population over time are incorporated into modeling to estimate variance of a survey statistic. Six such LGVF models are studied and results produced span over 15 years of Current Population Survey's (CPS) March Supplement data from socio-economic category. In addition to this, variables grouped together are also studied. 18 binary variables are considered. Simulation shows that individual variable variance estimation outperforms grouped variable variance estimation.

Key words: Longitudinal generalized variance function; Adjusted design effect model; Individual and grouped variable variance estimation.

MSC: 62D05

1. Introduction

Cost and labor involved in computing the estimated variances for thousands of estimates every year could be saved if computation could be simplified using generalized variance functions (GVF). This paper extends the results of Zhang, Cheng and Lu (2019) to individual variables variance estimation for a large-scale, complex survey data. Individual variables variance estimation will help to narrow the focus to only the variables of interest in survey data whereas grouped estimation brings in unwanted variability into the model and makes it harder to estimate individual parameters with high accuracy. Literature review in the area shows that the relative variance (relvar) of a survey statistic is a function of the population total. This idea is supported by Johnson and King (1987), Valliant (1987) and McIllece (2016). As for the GVF, Wolter (2007) discusses the application of GVF to estimate variance of a survey statistic. This method of variance estimation has been in use for Current Population Survey (CPS) data by US Census Bureau as well. Sampling error of GVF estimators for Current Employment Survey (CES) is evaluated by Cho, Eltinge, Gershunskaya and Huff (2002).

Corresponding Author: Mohammed Quazi
Email: mquazi@unm.edu

We use the March Supplement CPS data for the state of New Mexico (NM) from the socio-economic category from year 2003 to 2017. Figure 1 shows the sampling scheme used in this study. New Mexico households from same neighborhood in the data are divided into Ultimate Sampling Units (USUs) where each USU contains four households. We believe households in the same neighborhood have similar economic statuses. 100 USUs are randomly picked from each year's data. Four individuals from each USU are then sampled using probability proportional to size (PPS) method. The simulation process is explained in detail in section 4. Unemployment compensation, education benefits and disability payments are three of the eighteen variables studied – all the variables are shown in table 1 on page 23.

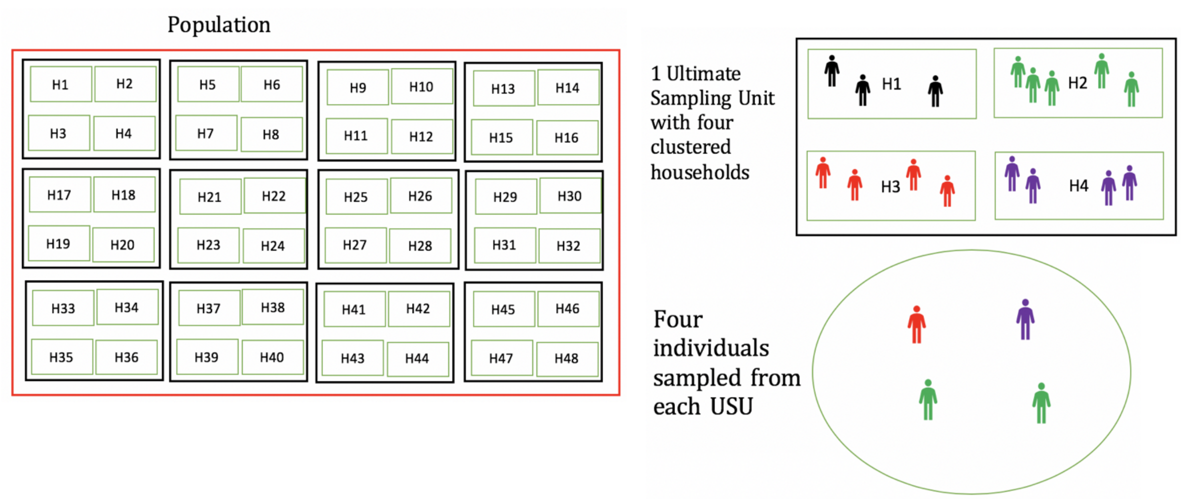

Figure 1: Sampling scheme employed in this study. USUs are black rectangles within the population. Oval shows individuals sampled using PPS method.

In the next section, section 2, longitudinal generalized variance functions (LGVFs) are briefly explained. Longitudinal adjusted design effect model (LADE) is discussed in section 3. Simulation process is discussed in section 4. Section 5, the final section, provides a succinct conclusion.

2. Longitudinal Generalized Variance Functions

We begin this section by providing a brief description of GVFs. Parameter of interest in this study is the proportion of people who benefit from the respective categories in table 1 on page 23. Let $\hat{p}$ be the estimated proportion of people benefiting for each such category. Let $\hat{T}$ be the estimated total number of people in the state of NM. T is estimated by $Horvitz - Thompson$ estimator

$$\hat{T}_t = \sum_h \sum_{i \in S_{th}} \left[\frac{M_{th} M_{thi} \bar{y}_{thi}}{n_{th} M_{thi}} \right] \quad , \tag{1}$$

where M_{th} is the total number of people sampled by CPS in stratum h for year t, technically, it is the total number of people sampled by CPS in NM for year t. For an instance, $M_{2010h} =$

2108, with a maximum number of people sampled from a single, unique household ("h_seq") being 20. M_{thi} is the number of secondary sampling units (SSU) in i^{th} primary sampling unit (PSU) for year t within stratum h. M_{thi} equals 4 for our study as shown in Figure 1. n_{th} is the number of PSUs picked in the sample within stratum h, this quantity equals 100 in this study. $\bar{y}_{thi}$ is the average of responses for year t and i^{th} PSU within stratum h. $\hat{T}_{thi} = M_{thi}\bar{y}_{thi}$ is the estimated total number of responses for year t and i^{th} PSU within stratum h.

From the formulation in Zhang *et al.* (2019), var($\hat{T}$) can be estimated by using GVF

$$\widehat{\text{var}}(\hat{T}) = \hat{a}\hat{T}^2 + \hat{b}\hat{T}, \tag{2}$$

which after incorporating the time or the population effect gives us the regression model as in Zhang *et al.* (2019)

$$\text{relvar}(\hat{p}) = \hat{v}_{tv} = \hat{a} + \hat{b} \cdot \frac{e_t}{\hat{T}} \tag{3}$$

where

$$\text{relvar}(\hat{p}) = \frac{\text{Var}(\hat{p})}{[\text{E}(\hat{p})]^2}, \tag{4}$$

$\hat{a}$ and $\hat{b}$ are estimated linear regression coefficients.

$e_t = M_t/\bar{M}$ takes into account the effect of change of population in NM for year t, where M_t is the population total for the state of NM reported by U.S Census Bureau for year t, $\bar{M}$ is the average population total over 15 years, $\hat{v}_{tv}$ is the response variable, and $e_t/\hat{T}$ forms the predictor variable. The need to incorporate the population effect can be explained using Figure 2 where the change in population of NM over 2003-2018 is shown. Equation (3) is a LGVF model. This model spans over multiple years of data. This is a generalization of GVFs over time. Interested reader can refer to Zhang *et al.* (2019) for more detailed description of this model.

To evaluate $\widehat{\text{var}}(\hat{T})$ in equation (2), we have made use of the estimator mentioned by Royall (1986),

$$\widehat{\text{var}}(\hat{T}_t) = \sum_h n_{th}(n_{th} - 1)^{-1} \sum_{S_{th}} \gamma_{thi}^2 r_{thi}^2, \tag{5}$$

where

$$\gamma_{thi} = M_{th}(n_{th}M_{thi})^{-1} \quad \text{and} \quad r_{thi} = \hat{T}_{thi} - \left(\sum_{S_{th}} \gamma_{thj} \cdot \hat{T}_{thj}/M_{th}\right) M_{thi}.$$

3. Longitudinal Adjusted Design Effect Model

In this section, we discuss incorporating design effects in LGVFs. We introduce the design effect d_{tv} and the adjusted design effect $f_{tv} = d_{tv}/\bar{d}_t$. $\bar{d}_t$ is the average of design effects

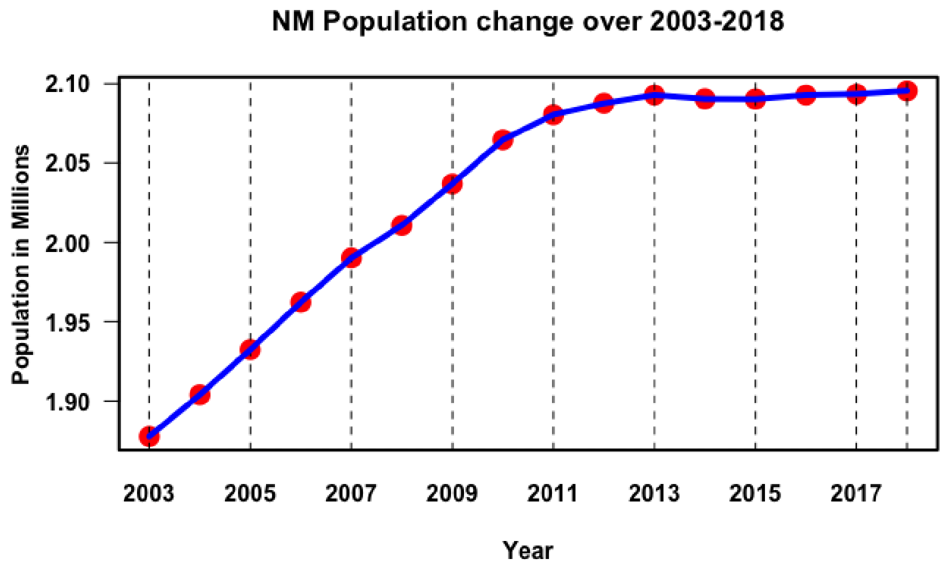

Figure 2: NM's population change over time.

for year t. $V = 18$ is the number of variables considered in the model and $\tau = 15$ is the number of years over which the model is spanned. $\boldsymbol{\theta} = (a, b)'$ be the LGVF coefficients which need to be estimated. $e_t = M_t/\bar{M}$ follows from previous section. Hence, for grouped variable case, we have $(V \times \tau)$ observations for regression to estimate a and b, whereas for individual variables case we have $(1 \times \tau)$ observations to estimate a and b. Let $a_{tv} = a = -\bar{d}_v/m$, and $b_{tv} = b = \bar{M}\bar{d}_v/m$ be the coefficients, from equation (2) we have

$$\begin{aligned}\widehat{\text{var}}(\hat{T}_{tv}) &= \frac{-\bar{d}_t}{m}\frac{d_{tv}}{\bar{d}_t}\hat{T}_{tv}^2 + \frac{\bar{M}\bar{d}_t}{m}\frac{M_t}{\bar{M}}\frac{d_{tv}}{\bar{d}_t}\hat{T}_{tv} \\ &= a_{tv}f_{tv}\hat{T}_{tv}^2 + b_{tv}e_t f_{tv}\hat{T}_{tv} \quad .\end{aligned}$$

The relative variance of $\hat{p}$ could be estimated by υ_{tv} for $t = 1, 2, \cdots, 15$ and $v = 1, 2, \cdots, 18$. We get

$$\upsilon_{tv}^* = \frac{\upsilon_{tv}}{f_{tv}} = a_{tv} + b_{tv} \cdot \frac{e_t}{\hat{T}_{tv}} \quad . \tag{6}$$

Equation (6), the LADE model, is applied for grouped variable variance estimation, and for individual variables estimation with $v = 1$. Properties of the estimators are reported in Zhang *et al.* (2019).

4. Results

We present the results for individual variables variance estimation and grouped variables variance estimation using LGVF and LADE models in this section.

4.1. Simulation

The data from 2003-2017 is considered to be the population for this study. Within each year, the households are assigned a USU after arranging the households in the increasing order of "h_seq". Then, 4 households are combined in that order to form a USU, 4 individuals are sampled from each USU. For years 2005, 2011 and 2014, one USU is dropped from each year because the USU contained 3 individuals. Table 2 on page 23 shows the number of USUs for each year. Following steps explain the simulation procedure in detail:

1. 100 USUs are picked based on PPS of the USUs which is about 50% sampling rate for each USU. 500 such random samples of size 100 USUs are picked for each year.

2. Estimates for relative variance, $\upsilon_{tv} = \widehat{\text{var}}(\hat{T}_{tv})/\hat{T}_{tv}$ are calculated using equations (1) and (5). Population totals are also calculated for year t. The population adjustment e_t is recorded as well. NM population totals (M_t) for years 2010-2017 are obtained from the US Census Bureau *factfinder.census.gov* (2020) and for years 2003-2009 from *countryeconomy.com* (2020). M_t is shown in Table 2 on page 23.

3. Ordinary least squares (OLS) regression model: LGVF1 model

$$\upsilon_{tv} = a_{tv} + b_{tv} \cdot \frac{e_t}{\hat{T}_{tv}} \tag{7}$$

   is applied and fits along with the coefficient estimates are recorded. Weighted least squares (WLS) regression – LGVF2 is applied with $weights = 1/\upsilon_{tv}$, and LGVF3 with *weights* estimated from regressing *residuals* from OLS (LGVF1) onto $e_t/\hat{T}_{tv}$ is also applied.

4. The Adjusted design effect $f_{tv} = d_{tv}/\bar{d}_t$ is recorded and $\upsilon^*_{tv} = \upsilon_{tv}/f_{tv}$ is calculated.

5. OLS regression: LADE1 model

$$\upsilon^*_{tv} = a_{tv} + b_{tv} \cdot \frac{e_t}{\hat{T}_{tv}} \tag{8}$$

   is applied and fits along with the estimated coefficients are recorded. WLS regression – LADE2 is applied with $weights = 1/\upsilon^*_{tv}$, and LADE3 with *weights* estimated from regressing *residuals* from OLS (LADE1) onto $e_t/\hat{T}_{tv}$ is also applied.

6. All the LGVF and LADE models are applied for grouped variables with $V \times \tau = 18 \times 15 = 270$ observations, and also for individual variables with $1 \times \tau = 1 \times 15 = 15$ observations.

7. This process is repeated for all the $R = 500$ samples picked in step (1).

8. Results along with the formulas used to calculate mean squared error (MSE), mean squared prediction error (MSPE), and Bias^2 are shown in Table 3 - 7 on page 24 - 30 in appendix.

5. Conclusion

Implications of obtained results are explained in this section. First, from grouped variable analysis – Table 3-5 on page 24-26, LADE models beat their LGVF counterparts when sum of MSEs, MSPEs or Bias^2 is considered but not necessarily for each variable, for example in Table 3 and 4, MSE and MSPE for variable 1 LADE1 model is higher than that of LGVF1 model. Similar conclusion can be drawn from Figure 3-4.

Second, from individual variable analysis – Table 6-7, LADE models outperform their LGVF counterparts when sum of MSEs or Bias^2 is considered and also for each variable when MSE is considered. Similar conclusion can be drawn from Figure 5.

Lastly, from overall analysis, LADE3 model – WLS regression where *weights* are estimated by using the *residuals* from OLS (LADE1), looks most promising out of all the models.

All in all, this paper has extended the results found in literature to individual variable variance estimation and proven that this application produces smaller error than the existing grouped variables variance estimation for longitudinal survey data. The idea of applying LGVFs for variance estimation in survey data is strengthened from the results obtained in this paper.

Acknowledgments

We thank the Chair Editor for his valuable advice and guidance.

References

Cho, M., Eltinge, J., Gershunskaya, J. and Huff, L. (2002). Evaluation of generalized variance function estimators for the U.S. current employment survey. *U.S. Bureau of Labor Statistics*.

countryeconomy.com. (2020). Retrieved from http://www.countryeconomy.com/

factfinder.census.gov. (2020). Retrieved from http://www.factfinder.census.gov/

Johnson, E. G. and King, B. F. (1987). Generalized variance functions for a complex sample survey. *Journal of Official Statistics*, **3**, 235-250.

McIllece, J. (2016). Calculating generalized variance functions with a single series model in the current population survey. *U.S. Bureau of Labor Statistics*.

Royall, R. M. (1986). The prediction approach to robust variance estimation in two-stage cluster sampling. *Journal of the American Statistical Association*, **81(393)**, 119-123.

Valliant, R. (1987). Generalized variance functions in stratified two-stage sampling. *Journal of the American Statistical Association*, **82(398)**, 499-508.

Wolter, K. M. (2007). *Introduction to Variance Estimation.* 2^{nd} *Edition.* New York: Spring-Verlag.

Zhang, G., Cheng, Y. and Lu, Y. (2019). Generalised variance functions for longitudinal survey data. *Statistical Theory and Related Fields*, **3(2)**, 150-157.

APPENDIX: Tables

Table 1: Eighteen variables studied – binary questions

SL. NO.	Variable	SL. NO.	Variable
1	Own business or self-employment	10	Survivors' payments
2	Unemployment compensation	11	Retirement payments
3	Social security benefits	12	Interest payments
4	Veterans' benefits	13	Dividend payments
5	Disability payments	14	Rental payments
6	Farm self-employment	15	Education benefits
7	Supplemental security benefits	16	Child support payments
8	Worker's compensation	17	Financial assistance payments
9	Public assistance/welfare benets	18	Other income payments

Table 2: Number of USUs for each year. One USU is dropped from 2005, 2011, and 2014.

Year	USUs	*Population total* $= M_t$
2003	253	1877574
2004	247	1903808
2005	232	1932274
2006	214	1962137
2007	220	1990070
2008	205	2010662
2009	208	2036802
2010	193	2064588
2011	184	2080395
2012	186	2087549
2013	184	2092792
2014	192	2090342
2015	318	2090211
2016	384	2092789
2017	371	2093395

In the following tables, second-last row is the sum, and last row is the mean of the respective column. $r = \{1, 2, \cdots, R = 500\}$, implies the number of iterations.

(1) Simulation-Grouped variables results:

Table 3: Grouped-MSE

Grouped-MSE						
SL. NO.	**LGVF1**	**LGVF2**	**LGVF3**	**LADE1**	**LADE2**	**LADE3**
1	0.00979473	0.00070947	0.00260306	0.01186915	0.00016289	0.00169219
2	0.02570878	0.06140138	0.037641	0.01771583	0.02574147	0.01973988
3	0.00512692	0.01052768	0.00302777	0.00677246	0.00292996	0.00213194
4	0.10071434	0.33833742	0.33316984	0.07326416	0.22924619	0.19022522
5	0.01381009	0.00008473	0.00389565	0.01558064	0.00001434	0.00228918
6	0.00340038	0.00601356	0.00163116	0.0040311	0.00125022	0.00092687
7	0.01309319	0.06647825	0.05722773	0.01129848	0.03332084	0.02568329
8	0.00646665	0.0149201	0.00374728	0.00466677	0.00234624	0.00162239
9	0.03551605	0.09590948	0.07396578	0.0310229	0.05307642	0.04046547
10	0.05594735	0.23740761	0.20623049	0.043161	0.16183356	0.11629299
11	0.0099287	0.0006615	0.00267133	0.0102696	0.0001471	0.00144438
12	0.01524628	0.00002822	0.00436762	0.02292429	0.00000854	0.00346307
13	0.01025905	0.00056497	0.00265315	0.01552496	0.00011266	0.00227115
14	0.00619912	0.00223324	0.00166028	0.00774532	0.00049188	0.0011493
15	0.00743651	0.00149724	0.00182384	0.01175148	0.00030338	0.00169697
16	0.00415483	0.00408964	0.00115219	0.00819017	0.00076848	0.00127823
17	0.06939598	0.40039895	0.45240352	0.05689342	0.29200676	0.24633193
18	0.10164471	0.53987857	0.61302082	0.07296997	0.39209662	0.32619876
$\sum_{v=1}^{18}\left\{\frac{\sum_{t=1}^{15}\left\{\frac{\sum_{r=1}^{500}(\hat{v}_{tv}^{(r)}-v_{tv}^{(r)})^2}{R}\right\}}{15}\right\}$	0.49384366	1.78114201	1.80289251	0.4256517	1.19585755	0.98490321
$\frac{\sum_{v=1}^{18}\left\{\frac{\sum_{t=1}^{15}\left\{\frac{\sum_{r=1}^{500}(\hat{v}_{tv}^{(r)}-v_{tv}^{(r)})^2}{R}\right\}}{15}\right\}}{18}$	0.02743576	0.09895233	0.1001607	0.02364732	0.06643653	0.05471685

Table 4: Grouped-MSPE

Grouped-MSPE						
SL. NO.	**LGVF1**	**LGVF2**	**LGVF3**	**LADE1**	**LADE2**	**LADE3**
1	0.14692099	0.010642	0.03904588	0.17803722	0.00244339	0.0253828
2	0.38563175	0.92102077	0.564615	0.26573739	0.38612198	0.29609827
3	0.07690383	0.15791527	0.04541662	0.10158697	0.0439494	0.03197908
4	1.51071516	5.07506124	4.99754763	1.09896245	3.43869281	2.85337825
5	0.20715132	0.00127092	0.05843473	0.23370953	0.00021511	0.03433766
6	0.05100575	0.09020342	0.02446746	0.06046653	0.01875324	0.01390303
7	0.19639785	0.99717382	0.85841597	0.16947713	0.49981262	0.38524942
8	0.09699981	0.22380151	0.05620918	0.0700015	0.03519358	0.0243359
9	0.53274068	1.43864213	1.10948665	0.46534351	0.79614627	0.60698199
10	0.83921018	3.56111412	3.09345728	0.64741494	2.42750345	1.74439489
11	0.14893043	0.0099225	0.04006995	0.15404396	0.00220643	0.02166576
12	0.22869413	0.00042323	0.06551435	0.34386428	0.00012817	0.05194609
13	0.15388573	0.00847451	0.03979731	0.23287445	0.00168997	0.03406732
14	0.09298682	0.03349853	0.02490425	0.11617975	0.00737819	0.0172395
15	0.11154768	0.02245857	0.02735761	0.17627213	0.00455069	0.02545455
16	0.06232243	0.06134462	0.01728288	0.12285255	0.01152714	0.01917341
17	1.04093967	6.00598423	6.78605287	0.85340126	4.38010134	3.69497891
18	1.52467068	8.09817858	9.19531234	1.09454957	5.88144932	4.89298145
$\sum_{v=1}^{18}\left\{\sum_{t=1}^{15}\left\{\frac{\sum_{r=1}^{500}(\hat{v}_{tv}^{(r)}-v_{tv}^{(r)})^2}{R}\right\}\right\}$	7.40765489	26.71713	27.043388	6.38477512	17.9378631	14.7735483
$\frac{\sum_{v=1}^{18}\left\{\sum_{t=1}^{15}\left\{\frac{\sum_{r=1}^{500}(\hat{v}_{tv}^{(r)}-v_{tv}^{(r)})^2}{R}\right\}\right\}}{18}$	0.41153638	1.484285	1.50241044	0.35470973	0.99654795	0.82075268

Table 5: Grouped-Bias^2

Grouped-Bias^2						
SL. NO.	**LGVF1**	**LGVF2**	**LGVF3**	**LADE1**	**LADE2**	**LADE3**
1	0.00932572	0.00043432	0.00197376	0.01064828	0.00004937	0.00101139
2	0.0062715	0.02416079	0.00228669	0.00334595	0.00323062	0.00079391
3	0.00081157	0.0059821	0.00000064	0.00229332	0.00113499	0.0000174
4	0.04073135	0.04129099	0.00143959	0.02801676	0.00076304	0.00020865
5	0.01337792	0.00002794	0.00321029	0.01409359	0.00000051	0.00147906
6	0.00186233	0.0039306	0.00020954	0.0024186	0.0005942	0.00010469
7	0.00058671	0.01136091	0.00012126	0.00003427	0.00089336	0.00001651
8	0.00003579	0.00792993	0.00001012	0.0001096	0.00115992	0.00000763
9	0.00900689	0.02377147	0.0010449	0.00887119	0.00356649	0.00150128
10	0.01459668	0.02012355	0.0000033	0.01174225	0.00015466	0.00000029
11	0.00948235	0.00039525	0.00207171	0.00913495	0.00005537	0.00084008
12	0.01478368	0.00000039	0.0036159	0.02086139	0.00000409	0.00227514
13	0.00981552	0.00037491	0.00202855	0.01382798	0.00003019	0.0013744
14	0.0055339	0.00149276	0.00094956	0.00660325	0.00020156	0.00052133
15	0.00690787	0.00105097	0.0012046	0.01021451	0.0001159	0.00091916
16	0.00321939	0.0029361	0.0003016	0.00660708	0.00034792	0.00049029
17	0.01583157	0.01323211	0.00255808	0.01319929	0.0005553	0.00105202
18	0.03075134	0.01860175	0.00295764	0.0225499	0.00221982	0.00244322
$\sum_{v=1}^{18} \frac{\sum_{t=1}^{15}\left\{\frac{\sum_{r=1}^{500}(\hat{v}_{tv}^{(r)})}{R}\right\}}{15} - \frac{\sum_{t=1}^{15}\left\{\frac{\sum_{r=1}^{500}(v_{tv}^{(r)})}{R}\right\}}{15}$	0.19293208	0.17709684	0.02598773	0.18457216	0.01507731	0.01505645
$\frac{\sum_{v=1}^{18} \frac{\sum_{t=1}^{15}\left\{\frac{\sum_{r=1}^{500}(\hat{v}_{tv}^{(r)})}{R}\right\}}{15} - \frac{\sum_{t=1}^{15}\left\{\frac{\sum_{r=1}^{500}(v_{tv}^{(r)})}{R}\right\}}{15}}{18}$	0.01071845	0.00983871	0.00144376	0.01025401	0.00083763	0.00083647

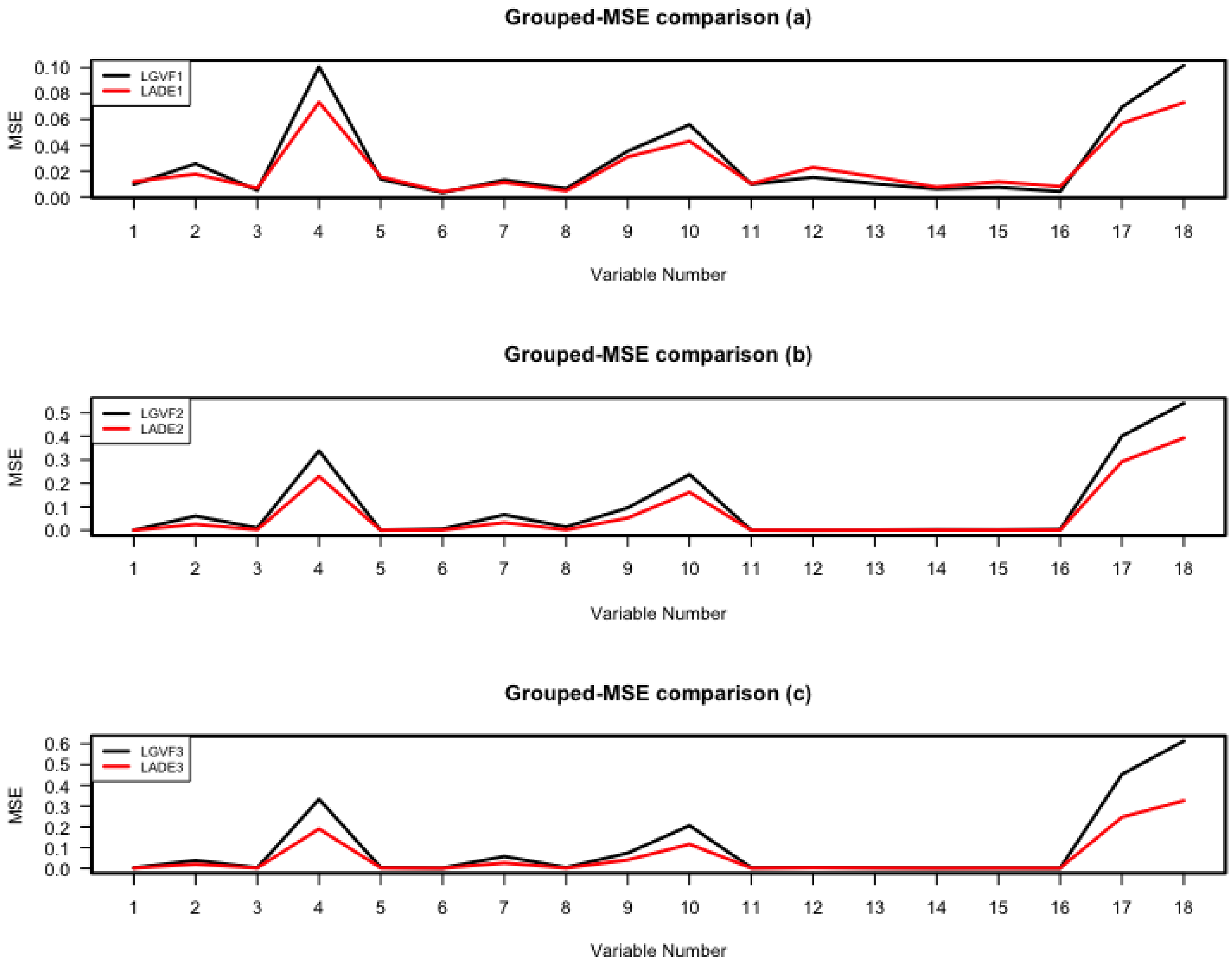

Figure 3: Grouped-MSEs for all variables – LADE models outperform LGVFs for most of the variables, but not all variables when each variable is compared.

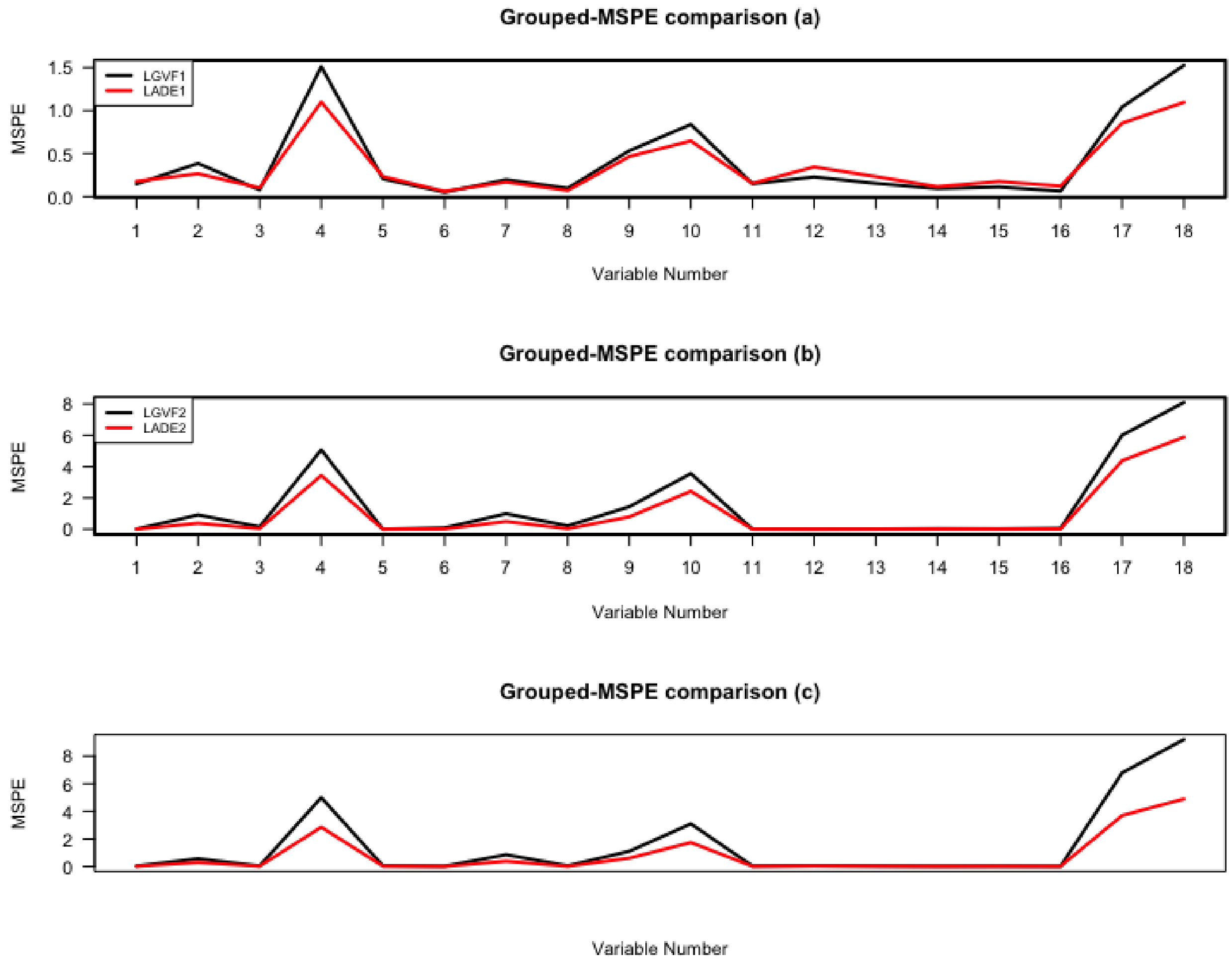

Figure 4: Grouped-MSPEs over all variables – LADE models outperform LGVFs for most of the variables, but not all the variables when each variable is compared.

(2) Simulation-Individual variables results:

Table 6: Individual-MSE

Individual-MSE						
SL. NO.	**LGVF1**	**LGVF2**	**LGVF3**	**LADE1**	**LADE2**	**LADE3**
1	0.00013186	0.00036702	0.00013588	7.5904E-05	8.5258E-05	7.6991E-05
2	0.00911577	0.03348467	0.00975517	0.00475778	0.00640601	0.00490773
3	0.00209517	0.00611522	0.00223322	0.00138177	0.00168737	0.00142106
4	0.03552953	0.10924189	0.03668525	0.03219963	0.03809822	0.03219542
5	1.1899E-05	6.5473E-05	1.2203E-05	6.9033E-06	7.3247E-06	6.9367E-06
6	0.00098323	0.00373517	0.00110428	0.00056473	0.00064215	0.00056823
7	0.00384158	0.02781115	0.00407711	0.00225242	0.00275966	0.00225011
8	0.00221485	0.01097709	0.00238354	0.00097313	0.00125433	0.00097575
9	0.01099218	0.04044535	0.01133325	0.00670696	0.00894584	0.00675416
10	0.02237646	0.0797136	0.02348702	0.01589717	0.02046451	0.01587004
11	0.0001145	0.00049466	0.00011709	7.3763E-05	8.5823E-05	7.4661E-05
12	2.9673E-06	1.9645E-05	3.0933E-06	1.0193E-06	1.058E-06	0.00000104
13	7.9252E-05	0.00029655	8.3895E-05	6.3977E-05	6.9632E-05	6.4492E-05
14	0.00036913	0.0014005	0.00038955	0.00023347	0.00026231	0.00023636
15	0.00020298	0.00083692	0.00020594	0.00016128	0.0001765	0.00016184
16	0.00052875	0.00221343	0.00054544	0.00036051	0.00040721	0.00036441
17	0.02863593	0.07480319	0.02905497	0.02817858	0.03538639	0.02789071
18	0.0402084	0.10769188	0.04082121	0.0377467	0.04157593	0.03749089
$\sum_{v=1}^{18}\left\{\frac{\sum_{t=1}^{15}\left\{\frac{\sum_{r=1}^{500}(\hat{v}_{tv}^{(r)}-v_{tv}^{(r)})^2}{R}\right\}}{15}\right\}$	0.15743444	0.49971342	0.16242812	0.13163568	0.15831552	0.13131083
$\frac{\sum_{v=1}^{18}\left\{\frac{\sum_{t=1}^{15}\left\{\frac{\sum_{r=1}^{500}(\hat{v}_{tv}^{(r)}-v_{tv}^{(r)})^2}{R}\right\}}{15}\right\}}{18}$	0.00874636	0.02776186	0.00902378	0.00731309	0.00879531	0.00729505

Table 7: Individual-Bias^2

Individual-Bias^2						
SL. NO.	**LGVF1**	**LGVF2**	**LGVF3**	**LADE1**	**LADE2**	**LADE3**
1	0.00000000	0.00007995	0.00000000	0.00000001	0.00000765	0.00000002
2	0.00000000	0.00781023	0.00000119	0.00000282	0.00062654	0.00000073
3	0.00000000	0.00148825	0.00000013	0.00000006	0.00022416	0.00000001
4	0.00000000	0.03136575	0.00000606	0.00039485	0.00357391	0.00032367
5	0.00000000	0.00001776	0.00000000	0.00000001	0.00000063	0.00000001
6	0.00000000	0.00080929	0.00000009	0.00000092	0.00006541	0.00000068
7	0.00000000	0.00396117	0.00000016	0.00000796	0.00027182	0.00000763
8	0.00000000	0.00252299	0.00000046	0.00000268	0.00016183	0.00000210
9	0.00000000	0.01072011	0.00000056	0.00000655	0.00110930	0.00000481
10	0.00000000	0.01717795	0.00000137	0.00008476	0.00138620	0.00007080
11	0.00000000	0.00015332	0.00000000	0.00000001	0.00001154	0.00000001
12	0.00000000	0.00000501	0.00000000	0.00000000	0.00000005	0.00000000
13	0.00000000	0.00006927	0.00000002	0.00000014	0.00000739	0.00000013
14	0.00000000	0.00037697	0.00000004	0.00000008	0.00002618	0.00000006
15	0.00000000	0.00022242	0.00000000	0.00000016	0.00002066	0.00000012
16	0.00000000	0.00057432	0.00000003	0.00000019	0.00004734	0.00000006
17	0.00000000	0.01985565	0.00000033	0.00023859	0.00379817	0.00021831
18	0.00000000	0.03093950	0.00000188	0.00042347	0.00433762	0.00034524
$\sum_{v=1}^{18} \frac{\sum_{t=1}^{15}\left\{\frac{\sum_{r=1}^{500}(\hat{v}_{tv}^{(r)})}{R}\right\}}{15} - \frac{\sum_{t=1}^{15}\left\{\frac{\sum_{r=1}^{500}(v_{tv}^{(r)})}{R}\right\}}{15}$	0.00000000	0.12814991	0.00001232	0.00116326	0.01567640	0.00097439
$\frac{\sum_{v=1}^{18} \frac{\sum_{t=1}^{15}\left\{\frac{\sum_{r=1}^{500}(\hat{v}_{tv}^{(r)})}{R}\right\}}{15} - \frac{\sum_{t=1}^{15}\left\{\frac{\sum_{r=1}^{500}(v_{tv}^{(r)})}{R}\right\}}{15}}{18}$	0.00000000	0.00711944	0.00000068	0.00006463	0.00087091	0.00005413

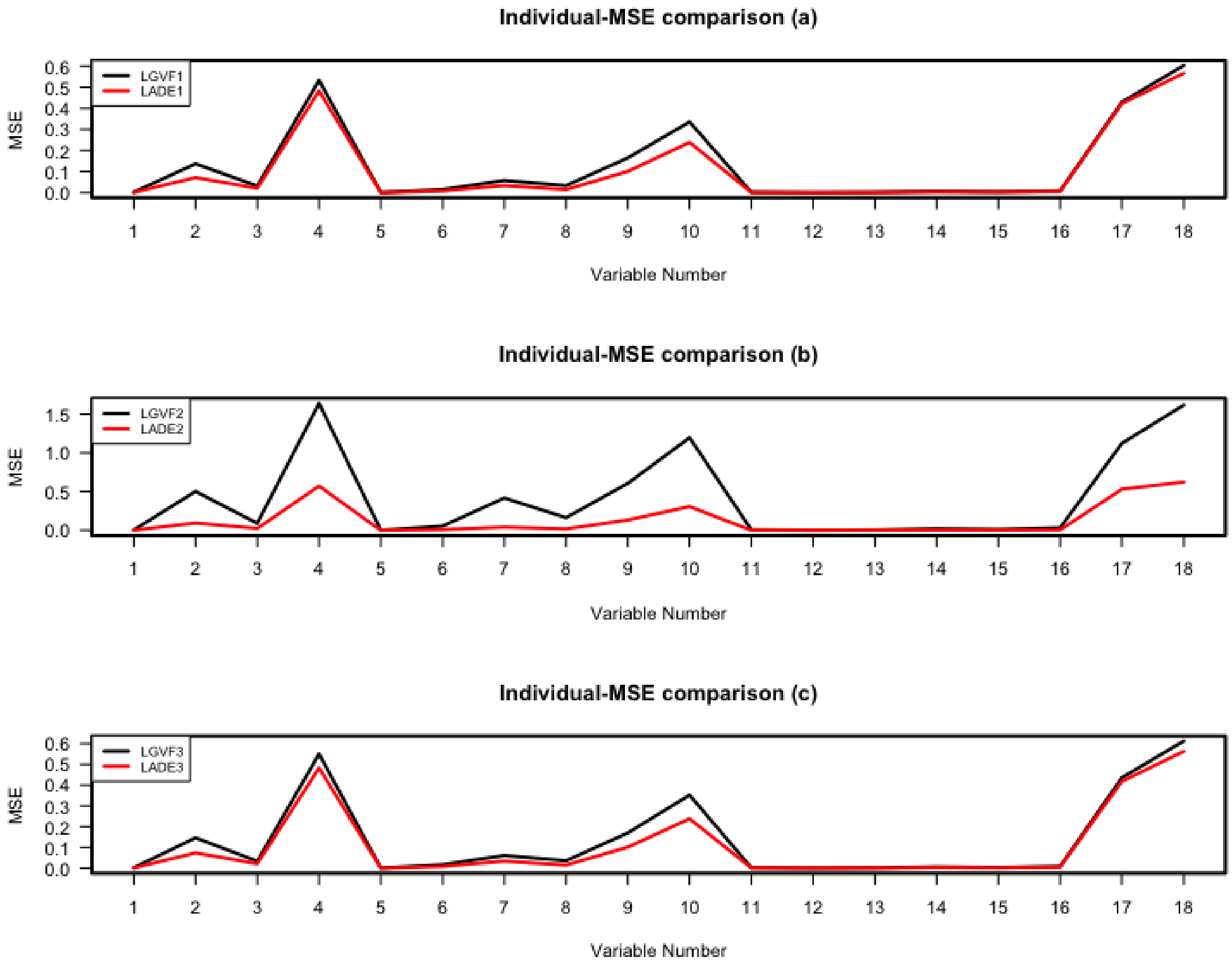

Figure 5: Individual-MSEs – LADE models outperform LGVFs for all of the variables when all the variables or each variable is compared.