1 *Article*

2 Heterogeneity of Cooperative Membership:
3 Implications for Cooperative Sustainability

4 **Matthew Elliott [1], Lisa Elliott [2] and Evert Van der Sluis[2,*]**

5 [1] South Dakota State University; Matthew.Elliott@sdstate.edu
6 [2] South Dakota State University; Lisa.Elliott@sdstate.edu
7 [3] South Dakota State University; Evert.VanDerSluis@sdstate.edu
8
9 * Correspondence: Matthew.Elliott@sdstate.edu; Tel.: +01-605-688-4855

10

11 **Abstract:** The effects of heterogeneity of cooperative membership on cooperative and collective
12 action sustainability has been previously discussed. However, despite the importance of membership
13 heterogeneity in recent theoretical frameworks, empirical examinations have been limited. We
14 determine the effect of changes to cooperative member heterogeneity on cooperative sustainability
15 and discuss changes to heterogeneity overtime that can advance our understanding to cooperative
16 sustainability long-term. This study uses USDA Agricultural Management Resource Survey data,
17 coupled with USDA-Rural Development cooperative financial data at the state level, to quantify
18 effects of cooperative member heterogeneity to sustainability of U.S. farmer cooperatives. We use
19 random forest regression to interpret the significance of heterogeneity with cooperative sustainability
20 at an aggregate level. The findings of this empirical study narrowly reconciles the theoretical
21 understanding of the emergence of intra-cooperative issues while providing consistent empirical
22 evidence and expectations for the sustainability of cooperatives in the near term.

23

24 **Keywords:** Cooperatives, Membership Heterogeneity, Random Forest, Collective Action
25

26 1. Introduction

27 Recent cooperative theoretical literature has focused on membership heterogeneity in
28 understanding cooperative sustainability long-term [1]. Despite the significance of membership
29 heterogeneity in advancing the theoretical understanding of cooperative sustainability, the empirical
30 attention has not been comparable. This is in part due to lack of data on cooperative membership
31 heterogeneity and/or limited empirical methods that can advance our understanding. An ideal
32 dataset to conduct an empirical examination of cooperative membership heterogeneity on
33 cooperative performance is not presently available. However, demographic and financial
34 information on farm producers who reported they received a cooperative patronage or had equity in
35 a cooperative has been collected in the USDA Agricultural Resource Management Survey (ARMS)
36 since 1996. Additionally, the ARMS survey collects a number of other variables that have been
37 largely posited to be associated with cooperative membership heterogeneity that would affect
38 cooperative sustainability [2]. Moreover, recent developments in machine learning methods have
39 enabled researchers to examine high dimensional data, and make accurate statistical inferences
40 concerning variables with low explanatory power due to aggregation issues [3-4]. Specifically,
41 random forest ensemble methods have been found to be substantially more powerful in
42 understanding treatment effect heterogeneity compared to classical methods [5-6]. We aim to use
43 random forest methods to understand the treatment effect to cooperative membership heterogeneity

44   on cooperative sustainability if membership heterogeneity was more or less.   Using random forests,
45   we can better account for the interactions, non-linearities, and hidden effects that membership
46   heterogeneity may have in predicting cooperative sustainability long-term -- particularly in the
47   presence of irrelevant and strongly relevant covariates.
48   We used annual aggregated data at the state level due to the limited availability of more detailed
49   data on cooperatives and their membership makeup.   We combined multiple farm and
50   socioeconomic variables of cooperative membership and non-cooperative membership heterogeneity
51   at the state level from the ARMS data with the long historical series of farmer cooperative
52   membership, gross business volume, and number of cooperatives headquartered in a state that is
53   maintained by the USDA-Rural Development [7].   We employed the random forest method to draw
54   statistical inferences of variables associated with membership heterogeneity in order to advance our
55   understanding of the effects to cooperative sustainability at an aggregated level. We aim to
56   understand: 1) the effects to cooperative sustainability due to heterogeneity of cooperative
57   membership, 2) changes to cooperative members and participation due to membership
58   heterogeneity, and 3) assess the effect of membership heterogeneity on long-term cooperative
59   sustainability.   We included other variables that may have an equal impact on cooperative
60   sustainability long-term.   These variables include changes in prices for farm products, food and feed
61   products, consumer products, and data on the amount of value added by industry derived by the
62   Bureau of Economic Analysis (BEA) in estimating U.S. state GDP [8].
63   The findings of our study in general show that membership heterogeneity plays a lesser role
64   compared to regional variables in understanding cooperative gross business volume at the state level.
65   This may be in part due to cooperatives have established themselves in industries and regions for
66   providing goods and services at competitive prices.   Furthermore, we find that value addition at the
67   farm is correlated with greater cooperative business volume at the state level.   However, we draw
68   no conclusions on causality in this analysis.   The correlation is consistent with the theoretical notion
69   that cooperatives are able to return greater value to the farm than would be expected in oligopsony
70   or oligopoly markets.   The finding is also consistent with the notion that cooperatives are sustainable
71   long-term despite the potential for intra-cooperative issues.
72   We do find stronger evidence that membership heterogeneity plays an important role in
73   understanding the future makeup and number of cooperatives long-term.   For example, the model
74   that was generated indicates that farmer cooperative member mean socioeconomic (SES) status is
75   inversely related to the number of cooperative members and the number of cooperatives
76   headquartered in a state.   Indeed, states where cooperative members have greater mean
77   socioeconomic status we observe greater degrees of cooperative consolidation and cooperative
78   member declines. Moreover, cooperative members SES diversity was found to be inversely related to
79   the number of cooperative members and cooperatives headquartered in the state. Additionally, the
80   coefficient of variation in cooperative membership age class was found to also be inversely related to
81   the number of cooperatives headquartered in the state and the number of cooperative members.
82   Contrary to the notion that diversity makes collective action more difficult, the random forest model
83   that was generated showed expectations that there would be more cooperatives and cooperative
84   members as age and socioeconomic diversity of cooperative members increased.   This finding may
85   be counterintuitive to the theoretical notion that cooperative membership heterogeneity decreases
86   the likelihood of cooperative action.   However, the finding may be consistent in explaining that in
87   the presence of greater cooperative membership heterogeneity, consolidation and acquisition of
88   cooperatives has been less rapid.
89   The findings of this study seem to suggest that intra-cooperative issues associated with
90   membership heterogeneity expectedly play a role in cooperative consolidation and cooperative
91   extent.   At the same time, intra-cooperative issues may not be as important in understanding the
92   long-term sustainability of cooperatives business volume and market share of cooperatives in the
93   agri-food industry. The findings of this study can (i) further our understanding of emerging
94   cooperative issues, (ii) draw implications on the long-term sustainability of cooperatives due to
95   changes to cooperative member heterogeneity, and (iii) reconcile the recent theoretical focus on intra-

96   cooperative issues that is consistent with observations of cooperative survival over long periods of
97   time. Our empirical study supports a continued focus in understanding intra-cooperative issues for
98   advancing our understanding of cooperative sustainability, membership makeup, and cooperative
99   extent at a micro level.   At the same time, the study provides a reason to why a focus on intra-
100   cooperative issues due to member heterogeneity does not necessarily imply that long-term
101   cooperative sustainability is in peril.   Indeed, the results of our analysis indicate that in the near-
102   term the cooperative gross business volume will remain stable, if not increase, regardless of
103   membership heterogeneity and associated intra-cooperative issues.
104
105   **2. Conceptual Framework: Defining Cooperative Sustainability in the Presence of Membership**
106   **Heterogeneity**
107   Cook, Chaddad and Illopoulos [1], discuss three theoretical approaches that have advanced
108   cooperative understanding.   The first approach addressed how membership heterogeneity may not
109   necessarily lead to the instability of the cooperative.   The supposition was that the cooperative can
110   be sustainable by providing socially desirable efficiency outcomes in imperfect markets.   The
111   theoretical works in this approach view the cooperative as extensions of the firm in a neo-classical
112   framework.   Theoretical deductions provide proof that cooperatives could successfully provide
113   heterogeneous cooperative members surplus in multipurpose cooperatives [9-10], and over wider
114   spatial dimensions, and in the presence of competition from investor owned firms (IOFs) [11].   These
115   frameworks largely assume an oligopsony or oligopoly market, and/or degrees of asset specificity
116   (site, temporal, physical, human) in segments of the supply chain. The frameworks largely view
117   cooperative sustainability as dependent on providing socially desirable efficiency improvements by
118   correcting market failures resulting from IOFs exerting market power.
119   The second approach describes cooperative sustainability as maintaining a coalition with a
120   common interest, yet potentially diverse incentives.   Much of the literature in this approach
121   implicitly concedes the need for the cooperative, and/or collective action institution in general, to
122   address market inefficiencies from externalities. The second approach is largely framed in the *tragedy*
123   *of the commons* context--where cooperatives can provide a second-best contractual solution to market
124   failures [12].   However, the second approach highlights the inevitable influence problem and
125   horizon problems that can result, and lead to a similar inefficiencies that exists without a collective
126   action solution.   Horizon and influence problems can be exacerbated by membership heterogeneity
127   that would diminish the common interest of the coalition and create greater instability of cooperative
128   equilibriums long-term.   The frameworks rely primarily on game theory and draw largely from
129   public choice [13] and collective action frameworks [14].   The second approach defines sustainability
130   of the cooperative as making optimal governance decisions despite membership heterogeneity.
131   Cooperatives can be sustainable by altering cooperative size and governance rules to more effectively
132   economize on the bargaining costs and decision-making costs that result in sustaining collective
133   action.
134   The third approach defines cooperative sustainability as controlling agency issues by designing
135   optimal incentive contracts and reducing exhaustive bargaining in contractual holdup.   This
136   approach largely uses incomplete contract theory and ex-ante/ex-post asset specificity in
137   understanding the sustainability of cooperatives.   Cooperatives are unique in that the typical
138   principal-agent frameworks used for understanding optimal contracts, and incentive mechanisms in
139   incomplete contract theory, do accurately represent most cooperative relationships.   This is because
140   cooperative members can be both the principal and agent in contractual relationships with the
141   managers of the cooperative.   For example, a cooperative member/owner is a principal that is
142   allowed to ratify manager (agent) proposals, but cooperative members can also be an agent when
143   supplying products or services to the cooperative. The dual objectives of the managers and
144   cooperative member owners complicates the modeling for optimal incentive schemes to maintain
145   desirable agent action when action is hidden.   Thus, the design of optimal cooperative contracts are
146   mired in complexities that are not well understood or are expected to remain stable at an optimal
147   equilibrium.   Thus, sustainability of cooperative in the third approach is defined at how effective

148 cooperatives can control agency issues in exhaustive bargaining in incomplete contract relationships,
149 particularly as member and manager heterogeneity increases where incentives may not be aligned
150 and complexity of strategic interaction increases.
151   In all three approaches, membership heterogeneity plays an important role in determining the
152 sustainability of cooperatives overtime.

153 **3. Materials and Methods**

154   To analyze the relationship of cooperative membership heterogeneity on cooperative
155 sustainability we obtained annual cross sectional data from the USDA-ARMS survey on U.S. farm
156 producers and data from the USDA-Rural Business Development on cooperatives. We combined
157 these datasets by aggregating the ARMS data to the state level and joining the data with USDA-Rural
158 Business development data on farmer cooperative members that is made publically available at the
159 state level. Since 1996, the USDA-ARMS data has annually surveyed a sub sample of farm producers
160 (approximately 10,000 to 30,000 a year) in each state, across all types of farms, to collect financial and
161 demographic information on farm businesses and households to maintain suitable farm benchmarks.
162 In each survey, they ask the respondent to indicate the amount of cooperative patronage and
163 cooperative equity they have received in that year or possess as a part of their financial and asset
164 information. We coded each producer who reported receiving patronage or possessing equity as a
165 cooperative member and ones who didn't as a non-cooperative member. We then estimated the
166 mean and variance of cooperative members in each state compared to non-cooperative members over
167 a number of other variables that are in the ARMS dataset. The variables chosen have all been
168 previously identified in the cooperative literature as being important for understanding membership
169 heterogeneity issues with cooperatives.
170   Hohler and Kuhl [2] quantified the number of publications that identify different dimensions of
171 membership heterogeneity as being important in understanding impact to cooperative sustainability.
172 The most cited dimensions they found were farm size, type of product, age, location, and education
173 in that order. In our study, we estimate farm size and diversity by the mean farm asset value
174 (C1_ATOT_mean) and coefficient of variation of asset value by cooperative members
175 (C1_CV_ATOT). We also estimate farm size by the mean (C1_Acres_mean) and coefficient of
176 variation acres (C1_CV_Acres) that is operated. We estimate diversity of farm type by two survey
177 questions, one whether the farm is reported to be primarily grain or livestock (C1_Farmtype_stddev),
178 and then a more detailed question concerning what type of product represents the largest portion of
179 the operation's gross income. In this question there are 16 enterprise types that range from grains
180 and oilseed production to equine and aquaculture (C1_typefarm_stddev). We estimate location of
181 the cooperative member's state by the sub region (Sub_region), where there is 9 sub regions in the
182 U.S., and by the longitude (Centroid_x) and latitude (Centroid_y) of the center of the state. Finally,
183 we score the diversity and mean level of cooperative member education (C1_cv_education), as well
184 as income (C1_dv_finci), age (C1_cv_agecls_stddev), and amount of other business income
185 (C1_cv_busi) in a socioeconomic variable using factor analysis. The factor incorporates the means
186 (SES) and coefficients of variations (SES_diversity) of age, income, and education variables. We also
187 include producer and consumer price indexes that are derived by the Bureau of Labor and Statistics
188 (BLS). Values we include are the farm producer price index (farm_ppi), the food and feed producer
189 price index (food_feed_ppi), the urban consumer price index for the U.S. (cpi), and the commodity
190 producer price index (comdty_ppi).   All producer price indexes have a base year of 1982. In addition,
191 we joined value added estimates and quantity indexes at the state level for industries that are derived
192 by the BEA to estimate U.S. and State GDP. The BEA's industry identificaiton for the Agricultural,
193 Forestry, Farming and Fishing Industry is 3, thus our coded variable for value added in this industry
194 in the dataset is "Ind_va_3". Values for the farms alone (Ind_va_4), and food and kindered products
195 (Ind_va_20) are available as well. We included the quantity indexes estimated by BEA as "Ind_q_"
196 and value added per quantity as "Ind_va_q_".
197   To calculate the variables means and variances, and to measure the socioeconomic factor we
198 used SAS 9.4. We then used the *randomForest* package in R (see Appendix for R code) (Liaw and

199 Wiener, 2002) to generate the random forest model, assess variable importance, and plot the marginal
200 effects of membership heterogeneity on the response variables that indicate cooperative
201 sustainability.   We derived variable importance plots and main and interaction plots using Welling
202 et al.'s [4] R package *forestFloor*.
203   The target variables (dependent) that we associate with cooperative sustainability and we
204 predict in the random forest regressions consists is the data available from the USDA Rural business
205 development on cooperative businesses at the state level.   The data includes the sum of cooperative
206 gross business volume by cooperatives headquartered in a state (gbv), we deflate using the
207 commodity producer price index (gbv_dfl), cooperative net business volume in the state (net_gbv),
208 deflated (net_gbv_dfl), the sum of the number of cooperative members (members), and sum of
209 cooperatives that are headquartered in the state (coops_num).
210   The decision trees that are randomly generated that make up a random forest model are distinct
211 rule sets that comprehensively describe how to interpret attributes of inputs to make a prediction for
212 a target value.   In this case, an ensemble of decision trees create rules to examine attributes of
213 cooperative members in a state, industries in the state, the state and/or region itself, and predict the
214 amount of cooperative business volume, number of cooperatives, and the number of cooperative
215 members.   Each decision-tree is expected to be biased, but when included in an ensemble of decision
216 trees and aggregated is expected to enhance predictive accuracy.   The training data to develop the
217 rule sets are the observed patterns given the input values. Accuracy assessments are done by cross
218 validating against observations not included in the training set.   The decision trees are uncorrelated
219 because of preselected parameters that constrain what inputs variables can be used at each decision
220 node split, and through a random selection of the training data to allow different slices of the data to
221 inform the prediction.   For example, a hypothetical decision tree for cooperatives headquartered in
222 a state, for one random draw of the training data, may be a simple, single decision rule such as: "if
223 the state is in the East North Central Region then then the number of cooperatives will be x, else the
224 number of cooperatives will be y".   Another random sample of the training data could result in a
225 distinct decision tree with a similar rule but with a different cutoff for region (e.g. if state is in the
226 West North Central Region or East North Central Region then x, else y), or in combination with
227 another variable (e.g. if in the East North Central Region and if the number of farms in the state is
228 greater than 20,000 then x, else y).   A third tree may not have randomly selected the region variable
229 to be used in a node split, thus the tree may consist of a split based on the number of farms alone.   In
230 random forest prediction, the final prediction is the mode of the predictions of many distinct, decision
231 trees generated, and the confidence in the prediction is indicated by the frequency the set of decision
232 trees result in the same prediction relative to the total number of decision trees generated.
233   After we omitted observations from our dataset with missing data, the number of observations
234 we were left in the analysis was approximately 580 cross sectional observations of cooperative
235 numbers, business volume, and attributes related to cooperative membership heterogeneity, and
236 other control variables.   To generate a number of uncorrelated decision trees, but also maintain
237 stability in our importance and main effects of variable rankings, we generated 800 trees to make up
238 the random forest model.   The number of observations from our data that we randomly drew from
239 at each tree to train the regression trees was 200.   We restricted the number of variables that were
240 randomly selected to optimally split a node in the tree to be 4.   The selections of the number of trees,
241 training data, and number of variables to try at a node split is somewhat arbitrary, though there are
242 general rules of thumb given the number of variables and number of observations being analyzed.
243 As the number of trees that are generated increase, the predictions and accuracy is expected to
244 converge.   Thus importance rankings and marginal effects are not expected to be sensitive to the
245 preliminary selections, but may affect values of the measures used to rank the variables including the
246 percent of accuracy increase in the mean square error and the gini node purity index [4-5].
247
248
249
250

251   **3. Results**

252   *3.1. Prediction of Number of Cooperatives Headquarted in the State*

253   The number of farmer cooperatives headquarted in each state has been in decline across all
254   regions of the U.S.   This is in part due to mergers, consolidations, acquisitions, and exits.
255   However, some regions have expierienced greater rates of decline (See Figure 1).   For example, the
256   West North Central region and the East North Central Region have historically had the most
257   cooperatives headquartered in the state, but have also observed the greatest rates of declines since
258   1979 as indicated by the loess slope in Figure 1.
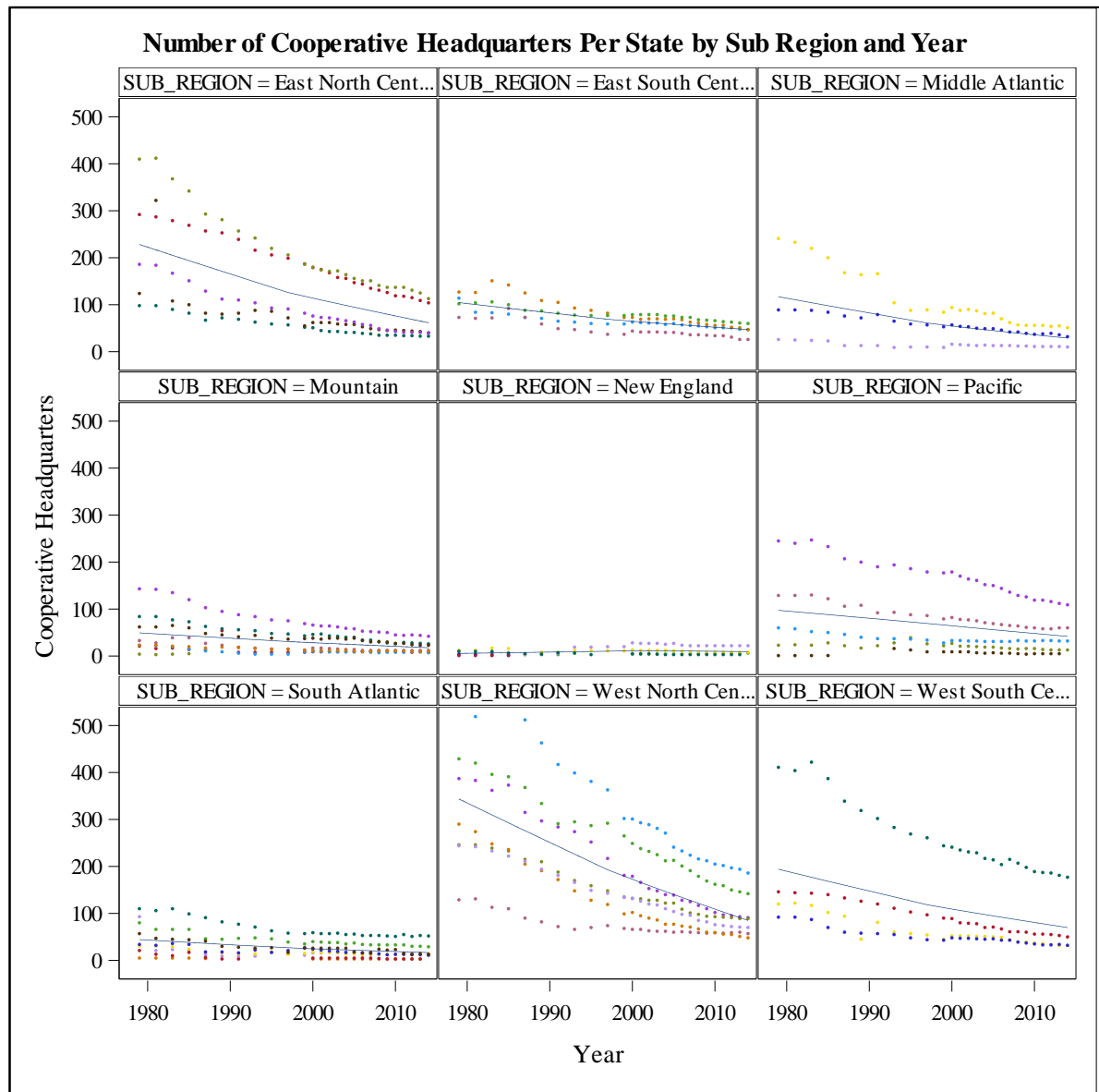
259

260   *Figure 1.   Source: USDA Rural Business Cooperative Statistics*

261
262   When we analyze the variables that were most important in explaining the number of cooperatives
263   headquartered in the state over this period, regional variables, as well as the number of farms and
264   the number of acres, or milk production tended to be ranked the highest (See Figure 2).   What is of
265   interest in this study, however, is the expected effect of the lesser variables related to membership
266   heterogeneity that help in providing explanatory power unrelated to the effects of the more important

267   variables. Specifically, do indicators of membership heterogeneity provide a significant main or
268   interaction effect to cooperative numbers in a state? In the case of number of cooperatives
269   headquartered in a state we find that socioeconomic status (SES) of cooperative members and the
270   coefficient of variation of the age class of cooperative members appears to be relatively important.
271   Perhaps even more important than the diversity of farm size—measured by acres—of cooperative
272   members (c1_cv_acres), for example.
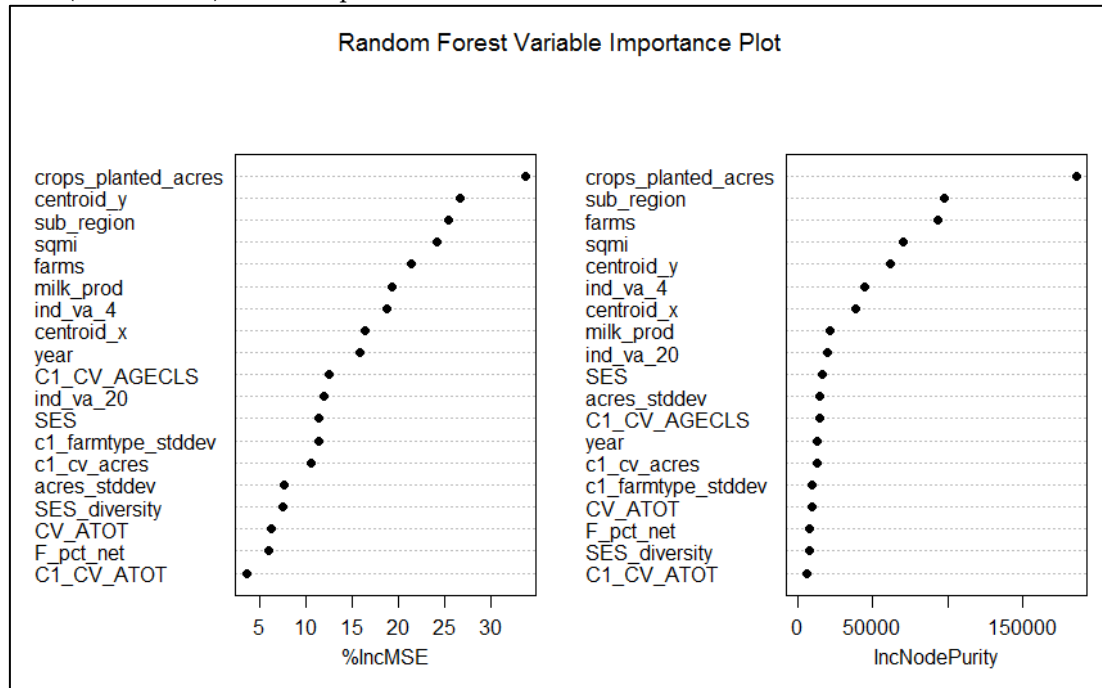
273
274   *Figure 2. Random Forest Variable Importance Plot for Number of Cooperatives Headquartered in a State*

275   To further examine the main effects of membership heterogeneity we can plot the main effects of the
276   number of cooperatives headquartered in a state using the *forestFloor* package. With the forestFloor
277   package, we can also observe what interactions effect may occur with other variables in the analysis
278   by examining the color gradient. For example, the main effect of cooperative member's age class
279   coefficient of variation on number of cooperatives indicates that states where cooperative members
280   have greater age diversity (.3 to .4 on the x axis), the random forest model predicts slightly more
281   cooperatives (y-axis) *cetaris paribus*. Specifically, the forest predicts up to 5 more cooperatives in a
282   state when the member coefficient of variation of age class is .35 (See Figure 3). The number of
283   cooperatives predicted is indicated on the y-axis in the model. This value is additive to other
284   cooperative number predictions in the model. Thus the total number of cooperatives predicted by
285   the random forest model would include a sum of all the marginal effect values. For example, when
286   cooperative member coefficient of variation score is .2, indicating less age diversity, the random forest
287   model expected no change in the number of headquarters in a state compared to the other variables
288   in the model. This is illustrated by gray main effects line for certain values of the coefficient of
289   variation of age class in Figure 3 that is near 0.
290   The color gradient in Figures 3 and 4 was determined by what sub region the observation was
291   in. Blue in Figure 3, for example, indicates the observation was in the South Atlantic, West North
292   Central, or West South Central regions. It appears in these regions the diversity of age effect was
293   more pronounced than in other regions of the U.S. Thus when the age coefficient of variation was
294   increased, coupled with being in the South Atlantic, West North Central, or West South Central
295   Regions, the model predicted more cooperatives headquartered in the state. The 12 most important
296   main effects are plotted in the same fashion as Figure 3 in Figure 4. In the lower right panel of Figure
297   4 (4c), the plot on the x-axis is the standard deviation of cooperative member's farm type, for example.
298   On the y-axis is the number of cooperative headquarters predicted in the state. The random forest
299   model predicts that states where cooperative members have greater variations in farm types there

300   was a slight expectation of more cooperatives headquartered in the state *cetaris paribus*.   Though this
301   effect was not very pronounced and could be largely disregarded as nearly meaningless.
302
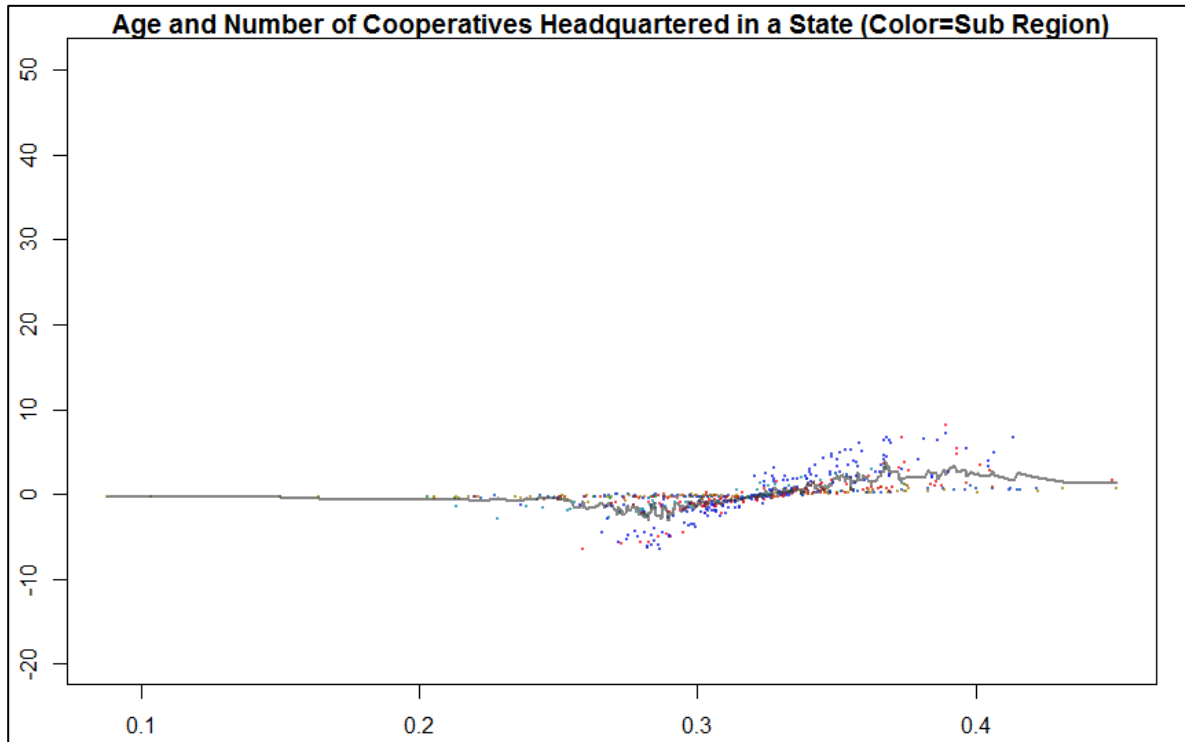
303
304   *Figure 3. Coefficient of Variation of Farmer Cooperative Member Age Class (x-axis) and the Predicted Number of*
305   *Cooperatives Headquartered in a State (y-axis).*

306
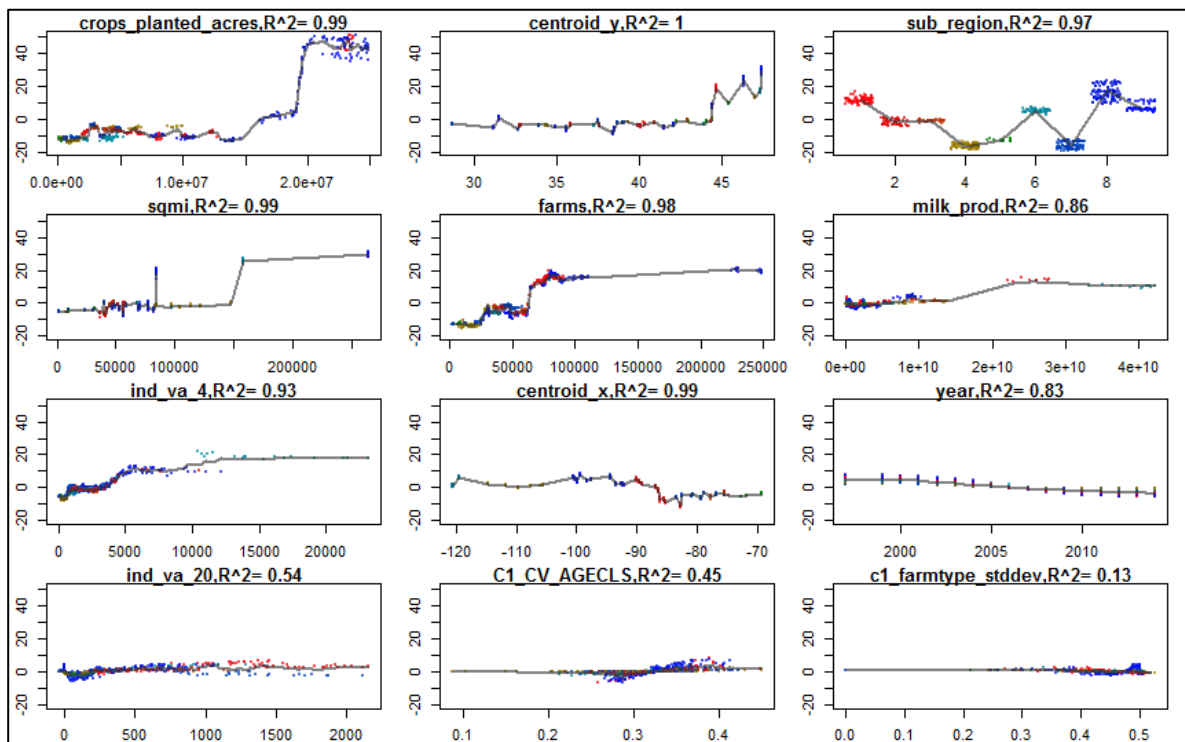
307
308   *Figure 4.*

309   *Panel (1a). The number of acres planted to field crops in a state (x-axis) and cooperative headquarters predicted (y-axis).*

310   *Panel (1b). The latitude of the center of a state (x-axis) and cooperative headquarters predicted (y-axis)*

311

312   *Panel (1c).   The sub region the state is in and the number of cooperative headquarters predicted.   X-axis values*

313   *correspond to `1=East North Central, 2=East South Central, 3=Middle Atlantic, 4=Mountain, 5=New England,*

314   *6=Pacific, 7=South Atlantic, 8=West North Central, and 9= West South Central.*

315   *Panel (2a).   The square miles of a state (x-axis) and the number of cooperative headquarters predicted (y-axis).*

316   *Panel (2b).   The number of farms (x-axis) in a state and the number of cooperative headquarters predicted (y-axis).*

317   *Panel (2c).   The amount of fluid milk produced (x-axis) in a state and the number of cooperative headquarters predicted*

318   *(y-axis).*

319   *Panel (3a).   The amount of value added (millions of US Dollars) by farms (x-axis) in a state and the number of*

320   *cooperative headquarters predicted (y-axis). Source: U.S. Bureau of Economic Analysis.*

321   *Panel (3b).   The longitude of the center of state (x-axis) and the number of cooperative headquarters predicted (y-axis).*

322   *Panel (3c).   The year (x-axis) and the number of cooperative headquarters predicted (y-axis).*

323   *Panel (4a).   The value added by the food and kindered products industry (millions of U.S. Dollars) (x-axis) and the*

324   *number of cooperative headquarters predicted in a state (y-axis). Source: U.S. Bureau of Economic Analysis.*

325   *Panel (4b).   The coefficient of variation of cooperative member age class (x-axis) and the number of cooperative*

326   *headquarters predicted (y-axis).*

327   *Panel (4c).   Farm type diversity (x-axis) and the Predicted Number of Cooperatives Headquartered in a State (y-axis).*

328   *3.2. Prediction of the Number of Cooperatives Members in a State*

329   The number of farmer cooperative members have also decreased in a similar fasion to the number
330   of coopertives that are headquaretered in a state (See Figure 5).   The decreasing trends in
331   cooperative members have closely followed trends of larger, more consolidated farms and greater
332   labor productivity from technical advances in agricultural production.   However, in some regions
333   cooperative members have been more resistent to the declining trend.   For example, in the New
334   England region, cooperative membersship has not declined at the rapid rate observed in the
335   Mountain region.

336   When we predicted cooperative membership by state in the random forest model we found the
337   most important variables were similar to the variable importance rankings when we predicted
338   cooperative headquarters.   Theese included included regional variables and the number of farms
339   in the state overall (See Figure 6).   Again, the variables associated with membership heterogeneity
340   that improved prediciton performance the most in the random forest model was the mean
341   socioeconomic (SES) status of the cooperative members in the state, and coefficient of variation of
342   age class (C1_cv_agecls).   One variable that was was more predictive in cooperative membership
343   than was predictive in the headquarters model was the standard deviation of acres of non-
344   cooperative members (acres_stddev).

345   Similar to the number of cooperatives headquartered in a state, when socioeconomic status of
346   cooperative members was below average, the model expected more cooperative members.
347   Specifically, a mean cooperative member socioeconomic status of negative one was expected to

348 increase state cooperative membership by approximately 8,000 members *cetaris paribus* (See Figure
349 7).   Other important variables for predicitng cooperative membership in the state can be observed
350 in Figure 8.   Notably, cooperative membership is expected to particularly high in the North Central
351 regions of the U.S., and when ratio of net income to gross cash sales is particularly low.
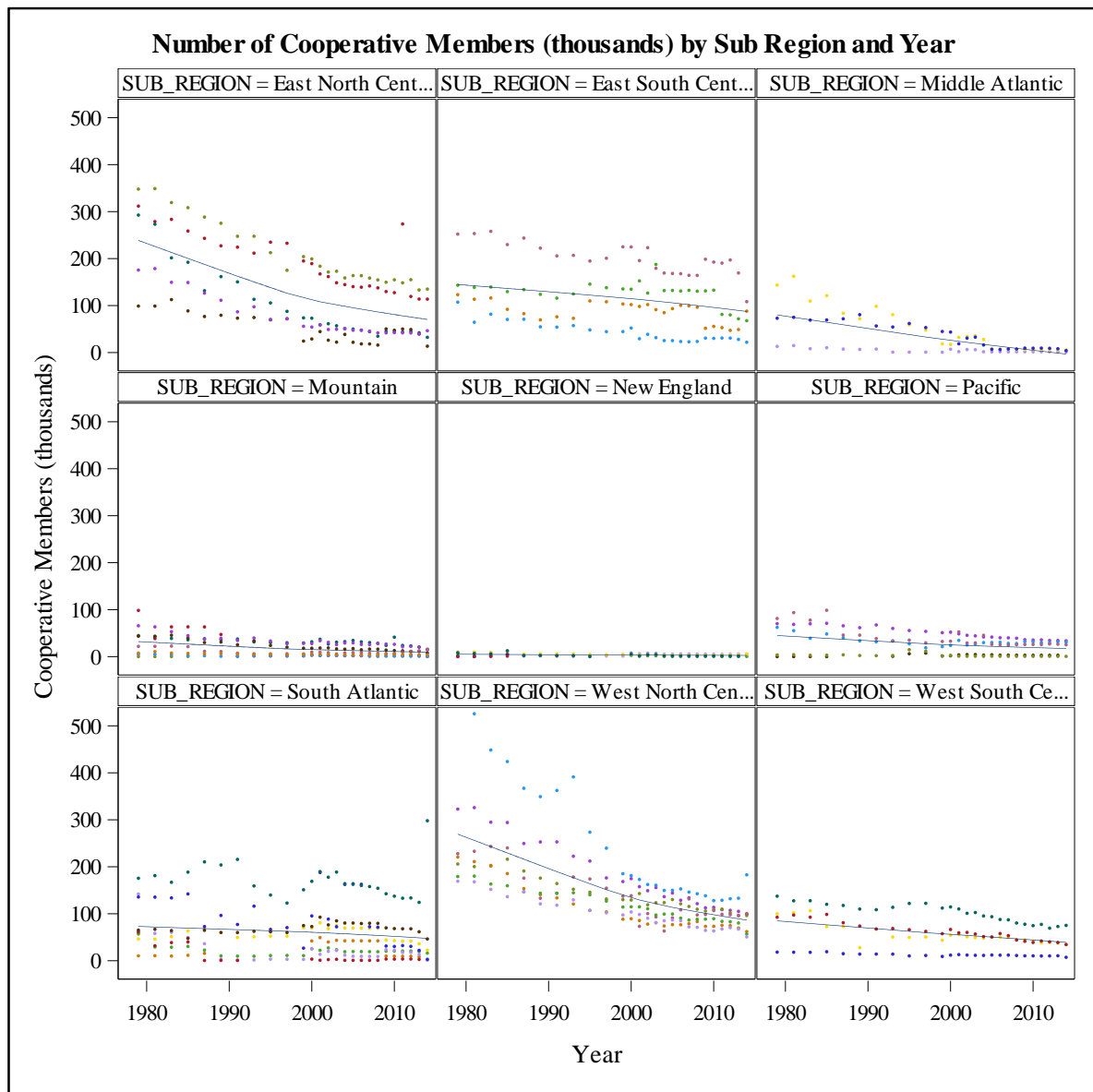
352

353 *Figure 5.   Number of Farmer Cooperative Members (thousands) by Sub Region and Year.   Source: USDA Rural*
354 *Business Cooperative Statistics*

355
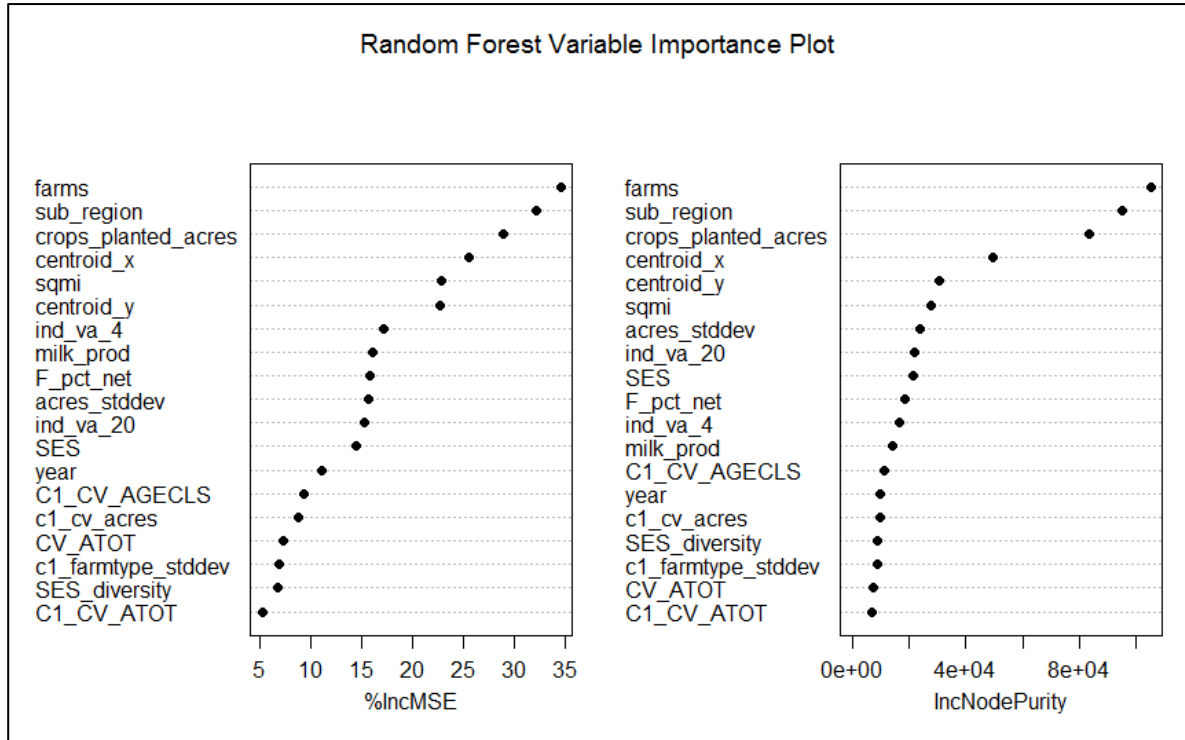
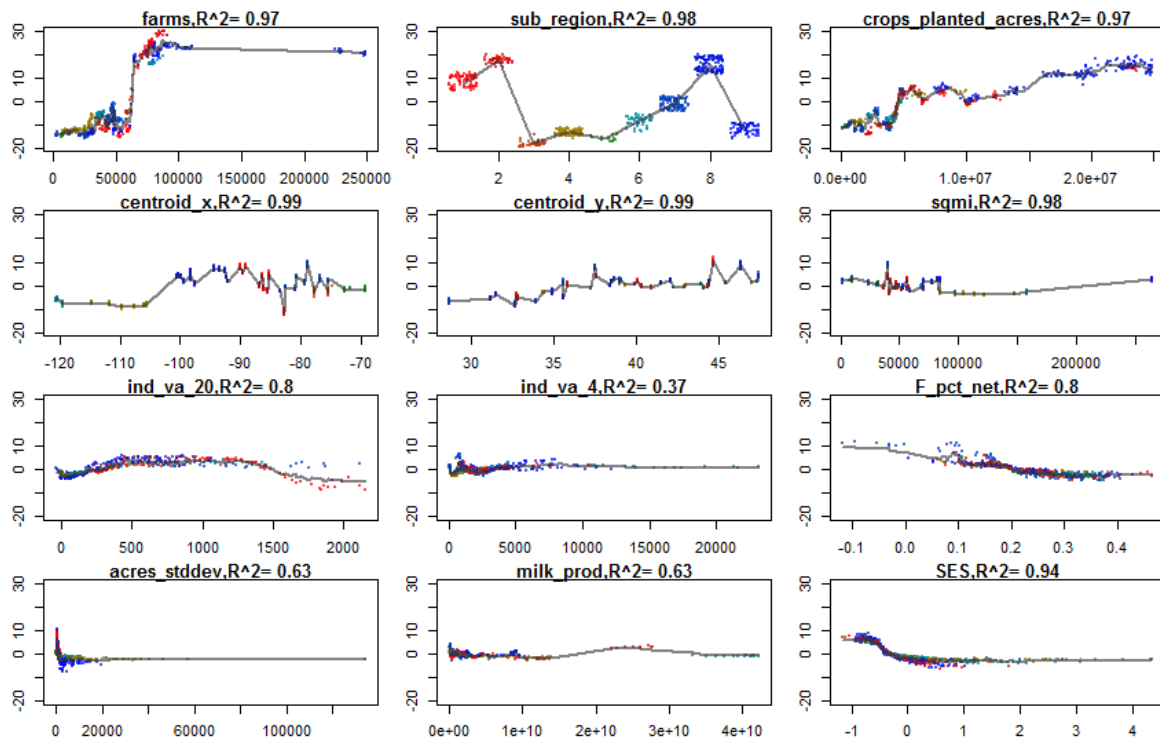*Figure 6.   Random Forest Variable Importance Plot for Number of Cooperatives Members in a State*

356

357

358
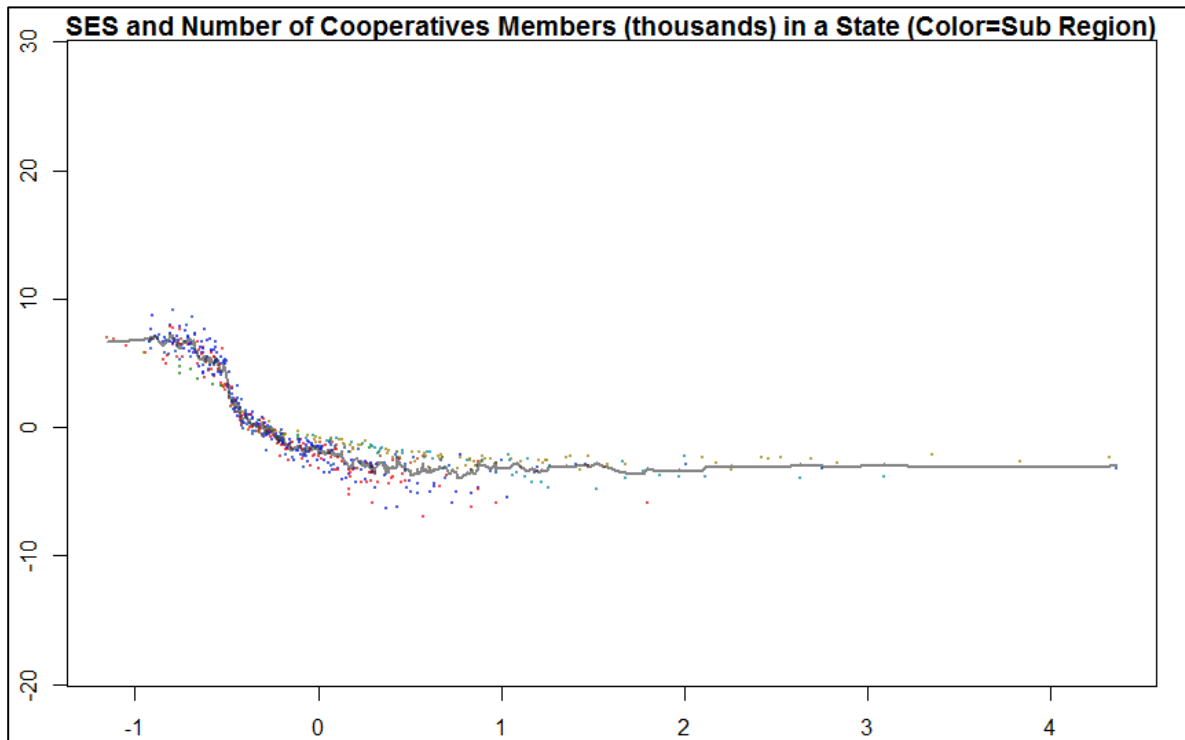
Figure 7. Socioeconomic status of Farmer Cooperative Members (x-axis) and the Predicted Number of Cooperatives Members in a State (y-axis).

359

360

361

362

363
364   *Figure 8.*

365   *Panel (1a). The number of farms in a state (x-axis) and cooperative members predicted (y-axis).*

366   *Panel (1b).   The sub region the state is in and the number of cooperative members predicted.   X-axis values correspond*
367   *to,1=East North Central, 2=East South Central, 3=Middle Atlantic, 4=Mountain, 5=New England, 6=Pacific, 7=South*
368   *Atlantic, 8=West North Central, and 9= West South Central.*

369   *Panel (1c). The number of field crop acres planted (x axis) and predicted cooperative membership (y-axis).*

370   *Panel (2a).   The longitude of the center of state (x-axis) and the number of cooperative members predicted (y-axis).*

371   *Panel (2b).   The latitude of the center of a state (x-axis) and cooperative members predicted (y-axis)*

372   *Panel (2c).   The square miles of a state (x-axis) and the number of cooperative members predicted (y-axis).*

373   *Panel (3a).   The amount of value added (millions of US Dollars) by the food and kindered products industry (x-axis) in a*
374   *state and the number of cooperative members predicted (y-axis). Source: U.S. Bureau of Economic Analysis.*

375   *Panel (3b).   The amount of value added (millions of US Dollars) by farms (x-axis) in a state and the number of*
376   *cooperative members predicted (y-axis). Source: U.S. Bureau of Economic Analysis.*

377   *Panel (3c).   The ratio of net farm income to gross cash farm sales in a state and the number of cooperative members*
378   *predicted (y-axis). Source: Bureau of Economic Analysis.*

379   *Panel (4a).   The standard deviation of farm acres of non-cooperative members (x-axis) and the number of cooperative*
380   *members predicted (y-axis).*

381   *Panel (4b).   The amount of fluid milk produced (x-axis) in a state and the number of cooperative members predicted (y-*
382   *axis).*

383   *Panel (4c).   Socioeconomic status of Farmer Cooperative Members (x-axis) and the Number of Cooperatives members*
384   *predicted in a State (y-axis).*

385   *3.3. Predicted Deflated Cooperative Gross Business Volume in a State*

386   Unlike the decreasing trends seen in cooperative headquarters and cooperative members per
387   state, the deflated cooperative gross business volume has seen a revival in recent years.   Specifically,
388   in the East North Central Region, the West North Central Region, the West South Central Region, the
389   Mountain Region, and the Pacific region have shown a "U" shaped pattern to gross business volume
390   (See Figure 9).

391
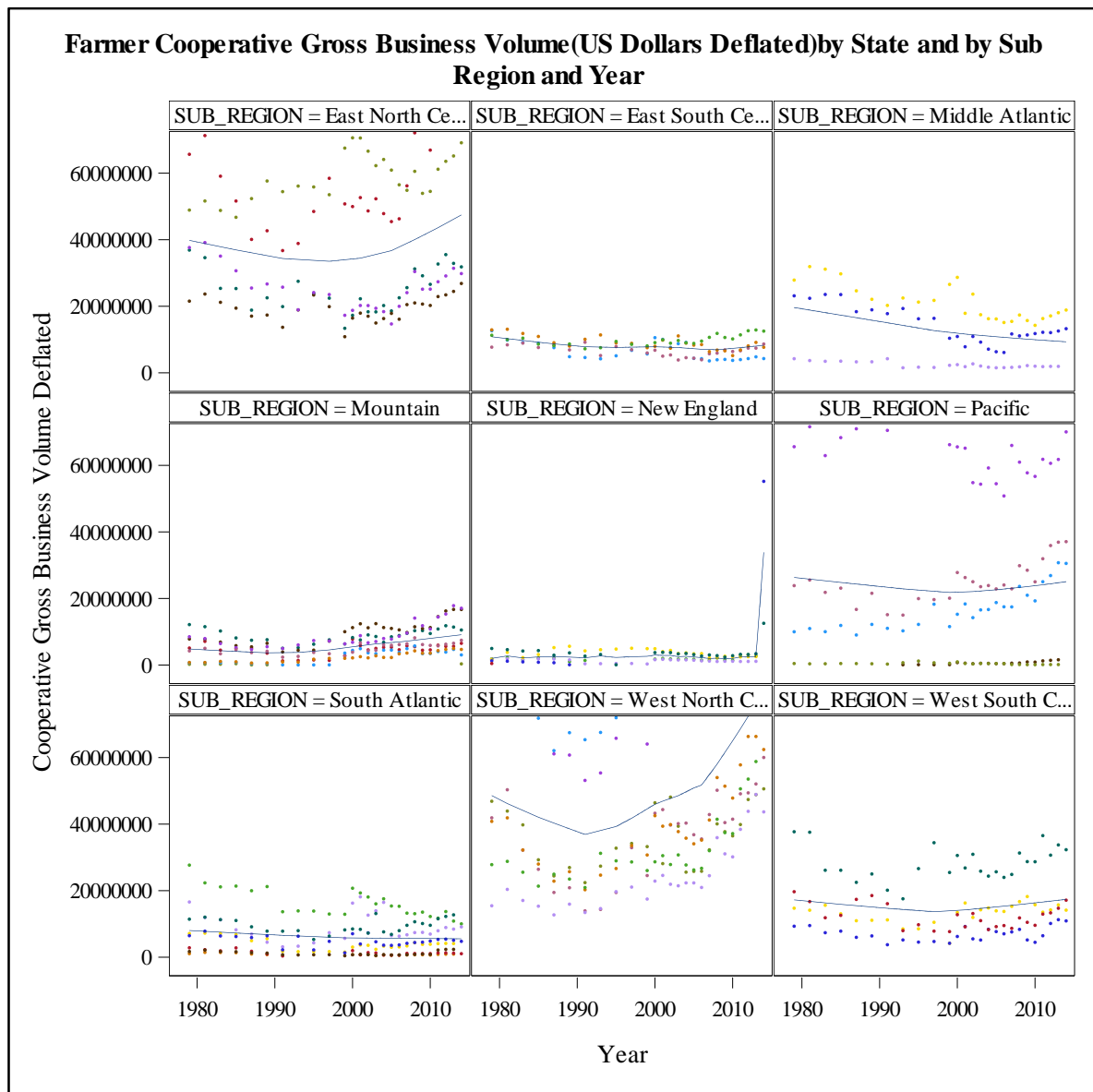
393   *Figure 9.   Deflated Cooperative Gross Business Volume (Base Year=1982).   Source: USDA Rural Business Cooperative*
394   *Statistics*

395

396   Accordingly, the random forest model that was generated found the sub region to be the most
397   important variable in predicting deflated gross business volume at the state level.   Variables of

398 membership heterogeneity were found to be less important in predicting deflated gross business
399 volume in the state relative to their importance in predicting the cooperative numbers and members
400 (See Figure 10).
401   Perhaps the most interesting finding in predicting gross business volume was the importance of
402 the amount of value added dollars by the farm industry (Ind_va_4). Specifically, value added by
403 farms was ranked third in importance under both measures (mse and node purity) in improving the
404 understanding of cooperative gross business volume at the state level (See Figure 10).   Indeed, the
405 amount of value added the farm industry contributes can have a strongly significant impact on the
406 amount of cooperative business volume that exists according to the random forest model.
407 Specifically, as the farm industry expects to add value to the state GDP in the order of 5 billion current
408 U.S. dollars, the corresponding expected value of deflated (1982 based) cooperative gross volume at
409 the state level increases to 10 million dollars (See Figure 11).
410   The variables mostly associated with membership heterogeneity appeared to play a lesser role
411 in understanding cooperative gross business volume at the state level overall. Though in certain
412 regions, the mean level of socioeconomic status (SES) of cooperative members appears to play a slight
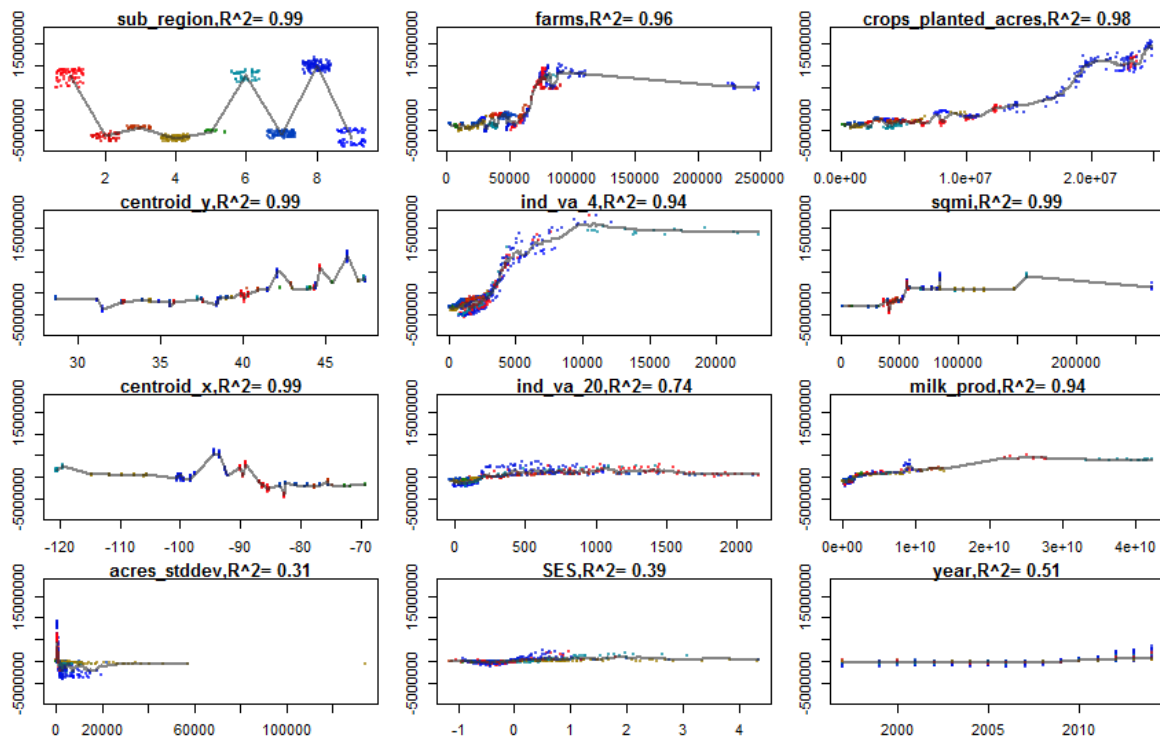413 role in increasing gross business volume (See Figure 12) .
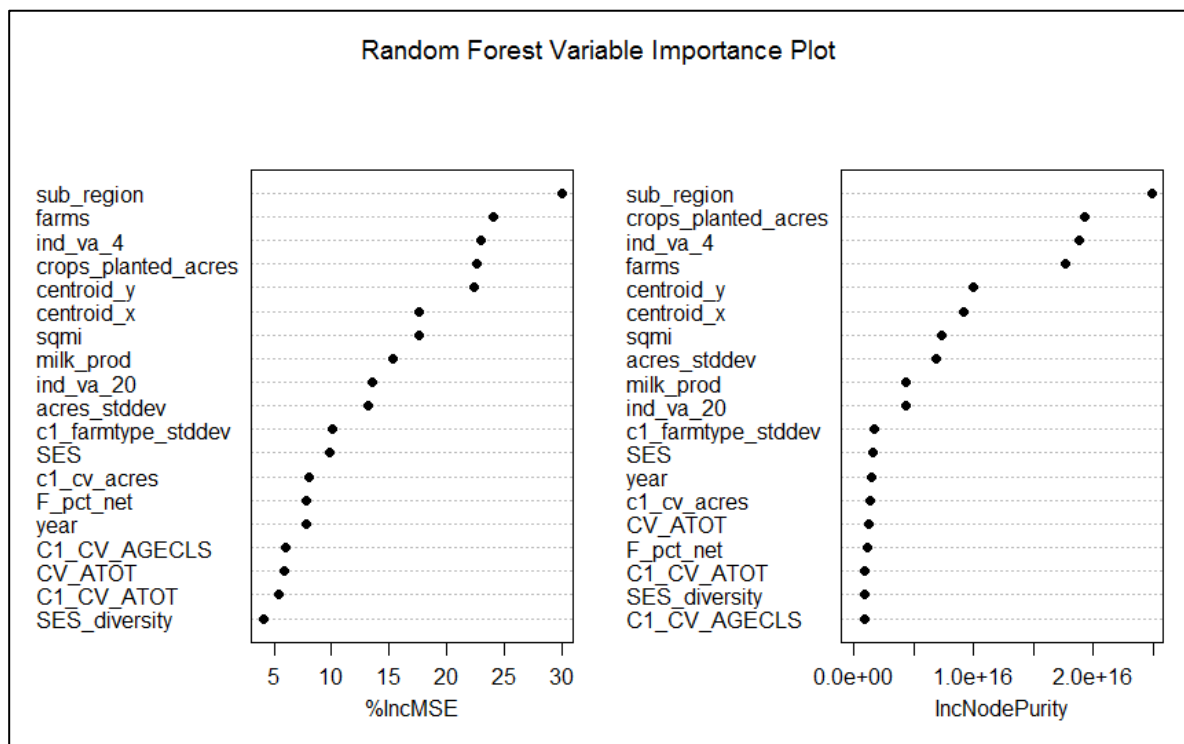
414

415 *Figure 10. Random Forest Variable Importance Plot for Deflated Gross Business Volume of Cooperatives in a State*

416
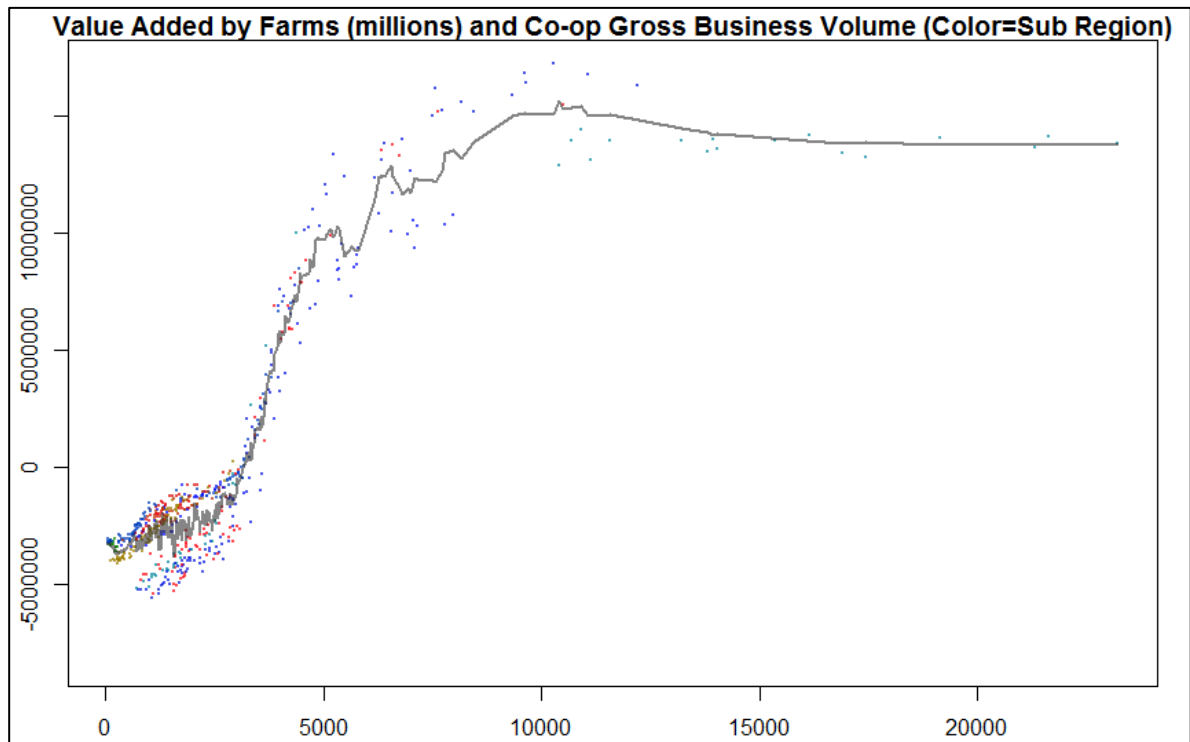
Figure 11.   State value added by Farms in millions of current U.S. dollars (x-axis) and deflated cooperative business

volume in 1982 dollars.

419

Figure 12.

Panel (1a). The sub region the state is in and the deflated   cooperative business volume predicted.   X-axis values

correspond to,1=East North Central, 2=East South Central, 3=Middle Atlantic, 4=Mountain, 5=New England, 6=Pacific,

7=South Atlantic, 8=West North Central, and 9= West South Central.

424

425   *Panel (1b). The number of farms in a state (x-axis) and cooperative gross business volume predicted (y-axis).*

426   *Panel (1c). The number of field crop acres planted (x axis) and the cooperative gross business volume predicted (y-axis).*

427   *Panel (2a).   The latitude of the center of a state (x-axis) and the cooperative gross business volume predicted (y-axis)*

428   *Panel (2b).   The amount of value added (millions of US Dollars) by farms (x-axis) in a state and the cooperative gross*
429   *business volume predicted (y-axis). Source: U.S. Bureau of Economic Analysis.*

430   *Panel (2c).   The square miles of a state (x-axis) and cooperative gross business volume predicted (y-axis).*

431   *Panel (3a).   The longitude of the center of state (x-axis) and cooperative gross business volume predicted ((y-axis).*

432   *Panel (3b).   The amount of value added (millions of US Dollars) by the food and kindered products industry (x-axis) in a*
433   *state and the cooperative gross business volume predicted (y-axis). Source: U.S. Bureau of Economic Analysis.*

434   *Panel (4a).   The standard deviation of farm acres of non-cooperative members (x-axis) and cooperative gross business*
435   *volume predicted (y-axis).*

436   *Panel (4b).   Socioeconomic status of Farmer Cooperative Members (x-axis) and the cooperative gross business volume*
437   *predicted in a State (y-axis).*

438   *Panel (4c).   The year (x-axis and cooperative gross business volume predicted in a State (y-axis).*

439
440

441   **4. Discussion**

442   In this study, we empirically examined the effects of cooperative membership heterogeneity on cooperative
443   sustainability in U.S. farmer cooperatives. We found that membership heterogeneity was expected to affect the
444   number of cooperatives headquartered in a state and the number of cooperative members. However,
445   membership heterogeneity was found to be less important in understanding cooperative gross business volume
446   at the state level.   The findings of this empirical work reconciles the notion that cooperatives can be sustainable
447   in the long-term despite much of the recent focus of cooperative literature on intra-cooperative issues that may
448   arise due to cooperative membership heterogeneity. Moreover, we find that cooperative member heterogeneity
449   may play an important role in decreasing the rate of consolidation and acquisition of cooperatives in the U.S.
450   This finding raises new questions about what is efficient, and more sustainable long-term, and whether
451   cooperative membership numbers and number of cooperatives should be the objective measure of cooperative
452   sustainability.

453   Future research should continue to pursue a greater theoretical and empirical understanding to intra-
454   cooperative issues related to membership heterogeneity.   More detailed data sets and new empirical methods
455   may allow us to parse the effect of cooperative membership heterogeneity on cooperative sustainability more
456   precisely.   This study provides an initial expectation and understanding of how membership heterogeneity may
457   affect cooperative sustainability long-term.

458

460

461

462

References

1. Cook, Michael Lee, Fabio Ribas Chaddad, and Constantine Iliopoulos. "Advances in cooperative theory since 1990: A review of agricultural economics literature." Erasmus University Rotterdam, Rotterdam School of Management, 2004.

2. Höhler, Julia, and Rainer KÜHL. "POSITION AND PERFORMANCE OF FARMER COOPERATIVES IN THE FOOD SUPPLY CHAIN OF THE EU-27." *Annals of Public and Cooperative Economics* 85.4 (2014): 579-595.

3. Breiman, Leo. "Random forests." *Machine learning* 45.1 (2001): 5-32.

4. Liaw, Andy, and Matthew Wiener. "Classification and regression by randomForest." *R news* 2.3 (2002): 18-22.

5. Caruana, Rich, and Alexandru Niculescu-Mizil. "An empirical comparison of supervised learning algorithms." *Proceedings of the 23rd international conference on Machine learning*. ACM, 2006.

6. Wager, Stefan, and Susan Athey. "Estimation and inference of heterogeneous treatment effects using random forests." *Journal of the American Statistical Association* just-accepted (2017).

7. United States Department of Agriculture, Rural Development. Annual Farmer, Rancher, and Fishery Cooperative Statistical Data. United States Department of Agriculture: Washington, DC, USA, 2017; Available online: <u>https://www.rd.usda.gov/programs-services/all-programs/cooperative-programs</u> (accessed on 29 March 2018).

8. U.S. Bureau of Economic Analysis, "NAICS Gross Operating Surplus by State" https://www.bea.gov/regional/downloadzip.cfm (accessed March 21st, 2018).

9. Feinerman, Eli, and Meira S. Falkovitz. "An agricultural multipurpose service cooperative: Pareto optimality, price-tax solution, and stability." *Journal of Comparative Economics*15.1 (1991): 95-114.

10. Choi, E. Kwan, and Eli Feinerman. "Regulation of nitrogen pollution: taxes versus quotas." *Journal of Agricultural and Resource Economics* (1995): 122-134.

11. Sexton, Richard J. "Imperfect competition in agricultural markets and the role of cooperatives: a spatial analysis." *American Journal of Agricultural Economics* 72.3 (1990): 709-720.

12. Zusman, Pinhas. "Constitutional selection of collective-choice rules in a cooperative enterprise." *Journal of Economic Behavior & Organization* 17.3 (1992): 353-362.

13. Buchanan, James M., and Gordon Tullock. *The calculus of consent*. Vol. 3. Ann Arbor: University of Michigan Press, 1962

14. Olson, Mancur. *Logic of collective action: Public goods and the theory of groups (Harvard economic studies. v. 124)*. Harvard University Press, 1965.