Journal of Agricultural Studies
ISSN 2166-0379
2013, Vol. 1, No. 2

Structuring State Intervention Policies to Boost Rice Production by Multinomial Logistic and Ordinal Regression Application and Multicollinearity Cautiousness

Shahin Shadfar

Department of Agricultural Development, Science and Research Branch

Islamic Azad University, Tehran, Iran

E-mail: shahin.shadfar@gmail.com

Iraj Malekmohammadi

Department of Agricultural Development, Science and Research Branch

Islamic Azad University, Tehran, Iran

E-mail: irajmalek@yahoo.com

Received: June 19, 2013 Accepted: July 4, 2013

doi:10.5296/jas.v1i2.3869 URL: http://dx.doi.org/10.5296/jas.v1i2.3869

Abstract

To structure state interventions policies to develop production of rice in Iran; developing two indexes to measure level of rice production development in dichotomous and categorical level; ordinal and multinomial logistic regression application are implicated to test the model by predictor variables in proposed policy structure. Taking extra care on Multicollinearity (MC), appropriate treatment by calculating Tolerance and Variance Inflation Factor (VIF) is performed. This is to test the fitness of the model by real data from the field, and to evaluate state intervention policies and plans, given this fact if the model fits at this stage, then it merits for further analysis to light up casual relationships among the effective factors on rice production development in Iran.

Keywords: Ordinal & multinomial logistic regression, Rice production development, State interventions

1. Introduction

The researches on the role of the state in the economy is generated many debates and countless pages of writings. The new millennium also poses new challenges for policy makers; government, private sectors and social society to set the development agenda of tomorrow. Specifically the proper role of government in the new millennium appears to be an interesting and challenging one. Therefore, if the future of modern economies and societies needs to be very different from the past, it requires a much sharper focus on radical development policy agendas (Karagiannis and Madjd-Sadjadi, 2007) In country like Iran where state involvement in economy has long legacy, it is expected that government who has absolute controlling power, makes rules and set the policies; drives re-structural reforms (Sinayiee, 2005). On the other hand, like many other Asian countries, rice in Iran is a staple food for the majority of the population, and ultimately is a food security concern. Therefore, substantial state intervention in terms of both regulation and support is being carried on. However, studying the other rice-consuming countries in the world shows constant rice production to feed the growing population is strategic agenda for any government. Therefore state interventions policies are indispensable. Obanil & Dano (2005) showed under the framework of continuing state interventions, options for developing rice production to meet domestic requirements are not very much different. Having said that, finding the appropriate formula comprising production related and market-based interventions determine whether self-sufficiency [strategic goal for many Asian countries as well as Iran] is feasible? Global rice data shows Iran was 4th rice importer country in the world (2010) by over 985,000 T imported rice, which is accounted for 3.4% of total global exported rice (Rice International Conference & Exhibition, 2011). Studies also shows rice production in Iran has serious issues like, natural resources degradations, low pace in rural growth and development, low contribution by rice farmers to set the policies and make decisions, existing powerful yet effective traditional & local structures, high risk and cost of production, deficiency in rice industry and lands leveling, fragmented farms, and change of land usage into real estate projects (Fallah, 2007). Though, state agencies in Iran actively plan and execute projects to increase the rice outputs; most of the Iranian state plans and programs in this section had been developed without feasibility studies and scientific background (Najafi, 2000). Consequently, government cannot properly analyze challenges in rice business and build up practical & effective solutions. Every year, Iran's government spends millions of dollars to achieve self-sufficiency in rice, but organizational structure and internal conflict of state plans and projects cannot address the challenges in this sector (Think-Tank of Ministry of Agriculture - Iran, 2009). Therefore, it worth it to study the state interventions policies in rice sector, in order to propose alternative plan in a way that brings the highest benefits and return-on-investment (ROI) in line with the sate intended development path. To address these challenges and increase the productivity and growth in business of rice production in Iran; this paper, as part of the PhD dissertation entitled "Designing the Structure of State Interventions for Developing Rice Production in Northern Region of Iran Based on Framers' Preferences" is aimed to analyze the state intervention policies in rice production development and re-structure them in a more effective & competent form. The study has reviewed state policies in regards to rice production development in major rice producing countries, to propose state intervention model (Malekmohammadi et al, 2011) and accordingly has identified principal areas that the state should take serious steps to develop the rice production. Regression applications have implicated to test the model by predictor variables in proposed intervention structure. Taking extra care on Multicollinearity, appropriate treatment by calculating Tolerance and VIF also has performed.

2. Theoretical Model

To define a theoretical model of this study, commonalities among state intervention policies in major rice producing countries (table 1) have studied. The output was globalized structure model (fig 1) which re-structure policies states have been taken to intervene into rice business. As initially assumed, in absence of any analytical model that can simplify the complexity of rice production involving factors and serve as an alternative analytical model, the successful government interventions in studied countries (i.e. major rice producing countries) can be duplicated as role model (see Malekmohammadi et al, 2011). Such a model can be used to understand the intricacies of the system and to study in advance the effects of changes in various internal and external variables.

Table 1. Major Rice Producing Countries

Country	Rice Production (Million Ton)	Global Production Share (%)
China	182.0	28.80
India	136.5	21.60
Indonesia	54.4	8.60
Vietnam	35.8	5.70
Thailand	29.6	4.60
Philippines	15.3	2.40
United States	8.8	1.40
South Korea	6.3	1.00
Malaysia	2.2	0.30
Source: Workman, 2008		

Another assumption of developing this analytical model was this fact that positive effects of these policies already have been endorsed by enormous amount of rice these countries had been producing. Therefore, following the same path might help to form same structure to ensure desired result; which is boost in rice output and ultimately developing rice production in Iran. Having said that, wide range of policies had exercised in these countries clearly pointed to state intervention as crucial factor for the success of rice production increase; the type of intervention was just important, though – if not more important. Nevertheless, common policy areas in state interventions, which major rice producing countries had been implemented can be summarized & re-structured as below:

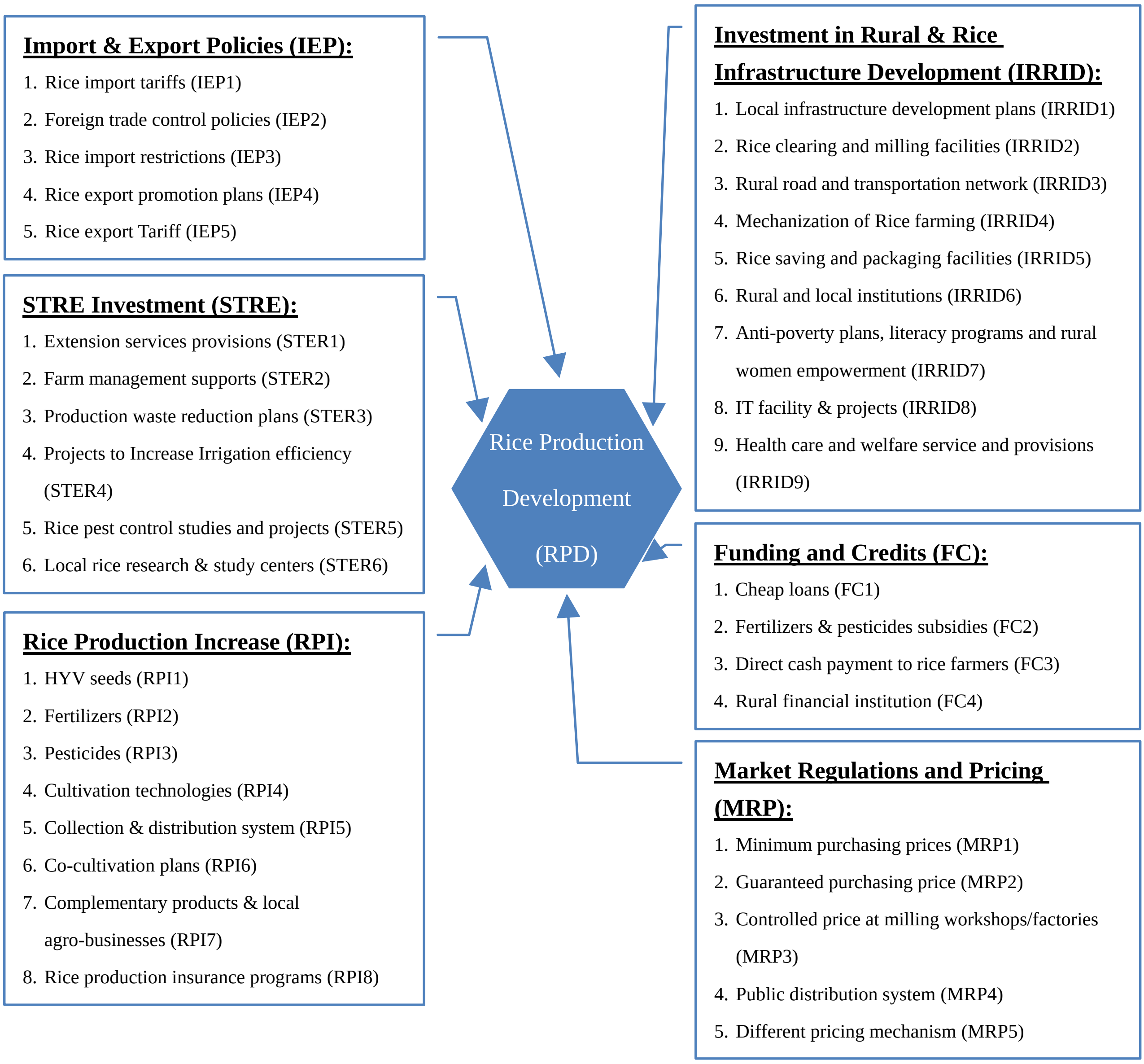

Figure 1. Theoretical Model to Structure State Interventions Policies in order to Develop Rice Production

1. Investment in Rural and Rice Infrastructure Development (IRRID)
2. Rice Production Increase Policies (RPI)
3. Investment in Science, Technology, Research and Extension (STRE)
4. Funding and Credits Policies(FC)
5. Market Regulations and Pricing Policies (MRP)
6. Import and Export Policies (IEP)

This structure describes the policy environment that had helped shaping the viability of the rice sector and the affordability and reliability of rice supply, specifying the institutional details of state interventions as well as the strategic policies that drive them.

3. Analyzing Method

To measure effects of the super-variables of the model on rice production, all rice farmers in state of Mazandaran (N = 176,792, n = 385) studied. To collect the data, questionnaire with different type of statements (in total 92 statements) developed in Likert scale. Validity and internal reliability of questionnaire measured by Cronbach's alpha coefficient (0.90), Theta coefficient (0.96) and Average Variance Extracted [AVE] (0.93) indices. By using innovative variable refinery technique (Malekmohammadi, 2008) some of statements and variables which could create bias omitted as well. From previous studies learnt that the factors that affect rice production include soils, geographic, water, climatic and biotic. Even further, under each, there are some sub-factors which have to be considered if rice production development is going to be measured. Technology changes also affect these variables. However, FAO (1996) took the average annual yield as measurement tool to compare grain production development in different countries (e.g. Bangladesh, China, India, Indonesia and Viet Nam). In addition, FAO has measured the country's productivity by variation in production per hectare (tones/ha) to show the situation (increase or decrease) in production. Therefore, in order to find an index and measure the level of rice production development, regardless of environmental, local and economical involving factors; following FAO technique; Dependent Variable (DV) of this study (level of rice production development) is measured by distance ratio of annual average production (kg/ha), calculated from reported average annual production by sample data and its distance to Iran annual average rice production = 3,910 kg/ha (SCI, 2010). This index provided a parameter in which higher than the country average production considered “developed production” and lower value index regarded as “not-developed” production. Rice production development index reported by sample of this study is shown in chart 1 in blue. By using this dichotomous index, we can map the rice production in conjunction to national average production which is being monitored by FAO in cross-country comparisons (FAO, 1996). As shown in figure 2, by looking deep inside the collected data, it’s revealed that 61% of sample produced lower than the national average output and 39% reported higher index values; means they have developed rice production business. To have categorical measure on level of rice production development as well, having annual average production data reported by sample and based on Percentiles statistics given in descriptive data, in distance to national average production; five levels of development slabs have been defined for rice production development (table 2).

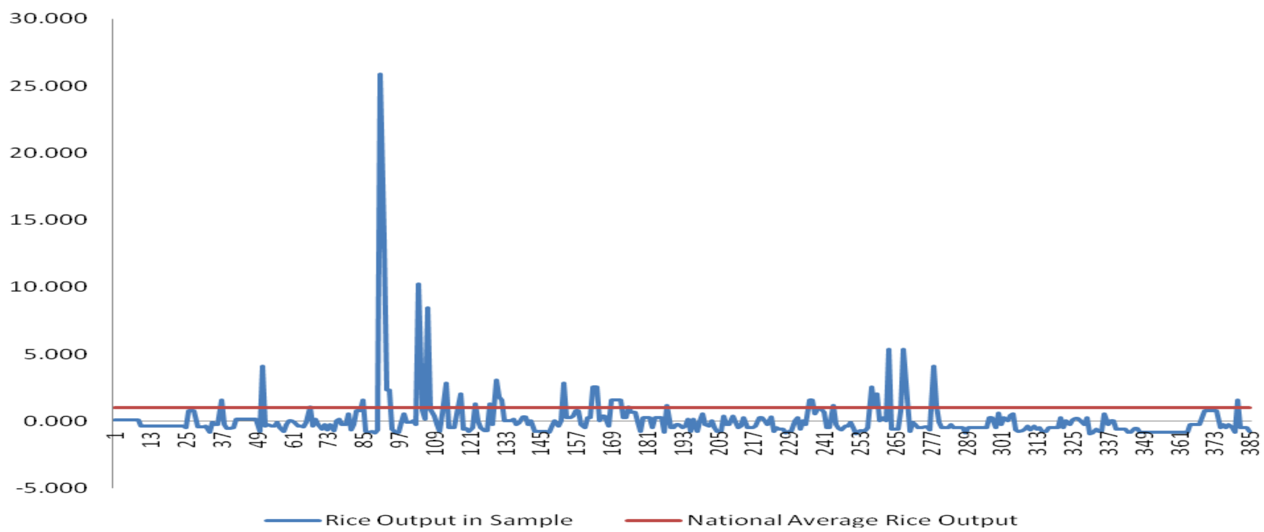

Figure 2. Level of Rice Production Development

Table 2. Level of Rice Production Development

Annual Average Production (T)	Level of Development
>1500	Not-Developed
1500 – 3500	Less-Developed
3500 – 5500	Semi-Developed
5500 – 7500	Developed
<7500	Highly-Developed

Table 3. Rice Production Development Classification

Level	Freq.	%	Valid (%)	Cumulative (%)
Not-Developed	87	22.6	22.6	22.6
Less-Developed	147	38.2	38.2	60.8
Semi-Developed	83	21.6	21.6	82.3
Developed	29	7.5	7.5	89.9
Highly-Developed	39	10.1	10.1	100.0
Total	385	100.0	100.0	

Applying this classification has made rice production development (DV) in ordinal scale. Now the rice production development in this study can be ranked and analyzes in a spectrum from “not-developed” to “highly-developed”. Table 3 shows the situation of rice production reported by sample ranked in ordinal scale. Following dichotomous measurement results, this table also shows more than 60% of the rice production in the sample was at “less-developed” level.

3.1 Ordinal & Multinomial Logistic Regression

The next step is to cross the model Independent Variables (IVs) by Defined Dependent

Variable (DV) and investigate the relationships between & among them. For this purpose, regression applications were carried out. Regression methods such as linear, logistic, and ordinal regression are useful tools to analyze the relationship between multiple Independent Variables (IVs) and Dependent Variable (DV). The regression methods are capable of allowing researchers to identify Independent Variables (IVs) related to Dependent Variable (DV). These methods also permit researchers to estimate the magnitude of the effect of the independent variables on the dependent variable. Therefore, regression methods seem to be superior in studying the relationship between the independent and dependent variables (Chen and Hughes, 2004). However, in despite of popularity of linear and logistic regression analyses, researchers are experiencing the challenge of using ordinal regression analysis to study the ordinal dependent because in part, they have not been fully exposed to the mathematical theory and the application software. Therefore, the ordinal regression model becomes a preferable modeling tool that does not assume the normality and constant variance, but require the assumption of parallel lines across all levels of the categorical dependent, were essentially assessed for selecting the best model (Plubin and Techapunratanakul, 2006). Basically, binary, ordinal (or binomial) logistic regression is a form of regression which is used when the dependent variable is a dichotomy and the independents variables are of any type. Multinomial logistic regression also exists to handle the case of dependent variable with more classes than two, though it is sometimes used for binary dependents also as it generates somewhat different output described below. Continuous variables are not used as dependents in logistic regression. However logistic regression can be used to predict a categorical dependent variable on the basis of continuous and/or categorical independents; to determine the effect size of the independent variables on the dependent; to rank the relative importance of independents; to assess interaction effects; and to understand the impact of covariate control variables. The impact of predictor variables is usually explained in terms of odds ratios (Garson, 2011). Similarly, in this study also, ordinal regressions have applied to check the fitness of the model by calculated dichotomous criterion and multinomial logistic regression have used by given categorical index of rice production development.

4. Results and Discussions

4.1 Ordinal Regression

Results of application ordinal regression on binary index of rice production (chart 1) are given as followings. In table 5 the -2Likelihood of the model with only intercept is 510.643 while the -2Likelihood of the model with intercept and independent variables is 31.310. That is the difference (Chi-square statistics) is 510.643 – 31.310 = 479.333 which is significant at 0.05 (p > .0001). Therefore, we can conclude that there is the association between the dependent and independent variable(s).

Table 5. Model Fitting Information

Model	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	510.643			
Final	31.310	479.333	258	.000
Link function: Logit.				

That means the null hypothesis that states the model without predictors is as good as the model with the predictors is rejected; means the model is valid. As Tabachnick & Fidell (2000) indicated, if the model fits well, the observed and expected cell counts are similar, the value of each Pearson & Deviance statistics is small, and the observed significant value is large. Good models have large observed significant level. As shown in table 6, p = 1.000, and model is not fit. The model-fitting statistic, named the pseudo R-square, measured the success of the model in explaining the variations in the data. It also measures the strength of association between the dependent variable (DV) and predictor variables (PV) (Tabachnick & Fidell, 2000).

Table 6. Goodness-of-Fit Statistics

Statistics	Chi-Square	df	Sig.
Pearson	16.450	119	1.000
Deviance	28.198	119	1.000
Link function: Logit.			

The pseudo R-square is calculating depend upon the likelihood ratio. For example, the McFadden's R-square compared the likelihood for the intercept only model to the likelihood for the model with the independent variables in order to assess the model goodness of fit. The interpretation of pseudo R-square in the ordinal regression model is similar to that of the R-square (e.g., coefficient of the determination) in the linear regression model. The pseudo R-square indicated that the proportion of variations in the dependent variable was accounted for by the independent variables. The larger the pseudo R-square is, the better the model fitting is (Plubin and Techapunratanakul, 2006). The pseudo R-squares of model in this study for McFadden was 0.931, for Cox and Snell was 0.712, and for Nagelkerke was 0.966, with the Logit link. Therefore, it can be said that the relationships between dependent variable and independent variables (predictor variables) overall is strong. The ordinal regression output shows only one threshold indicates where the latent variable is cut to make the five groups that we observe in our data. Note that this latent variable is continuous. In general, these are not used in the interpretation of the results. However, we expect that for a one unit increase in latent variable (i.e., going from 0 to 1), 97.4 increase in the ordered log odds of being in a higher level of dependent variable can be expected; given all of the other variables in the model are held constant. In addition, the Wald statistics in which is the square of the ratio of the coefficient to its standard error, tells based on the small observed values in the output, the null hypothesis that β is zero also can be rejected; means there appears to be a relationship between dependent variable (DV) and predictor variables (PV).

4.2 Multinomial Logistic Regression

Results of multinomial logistic regression application on categorical index of rice production are given in this section. Logistic Q-Q plot of the index also is given (Chart 2). As chart shows actual values lining up along the diagonal that goes from lower left to upper right. This plot also shows that index is normally distributed. The Case Processing Summary (table 7) simply shows how many cases or observations were in each category of the outcome variable (as well as their percentages). Results of multinomial logistic regression require that the minimum ratio of valid cases to independent variables be at least 10 to 1 (Schwab, 2006). The ratio of valid cases (385) to number of independent variables (5) was 77 to 1, which was greater than the minimum ratio needed. Therefore requirement for a minimum ratio of cases to independent variables was satisfied. It also shows there was no missing data. The Model Fitting Information (table 8) shows various indices to assess the intercept model only (sometimes referred to as the null model) and the final model which includes all the predictors and the intercept (sometimes called the full model). Both the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) are information theory based model fit statistics. Lower values of indicate better model fit and both can be below zero (i.e. larger negative values indicate better fit than values closer to zero). However, the BIC tends to be more conservative (Starkweather and Herrington, 2011). Similarly, the -2 Log Likelihood (-2LL) should be lower for the full model than it is for the null model; lower values indicate better fit. The -2LL is a likelihood ratio and represents the unexplained variance in the outcome variable, the smaller the value, the better the fit. The Likelihood Ratio chi-square test is alternative test of goodness-of-fit. As with most chi-square based tests however, it is prone to inflation as sample size increases. However, the model fit for this study was not significant $\chi^2 = 16.430$, $p = .690$, which indicates full model does not predict significantly better, or more accurately, than the null model. To be clear, p-value should be less than established cutoff (0.05) to indicate good fit.

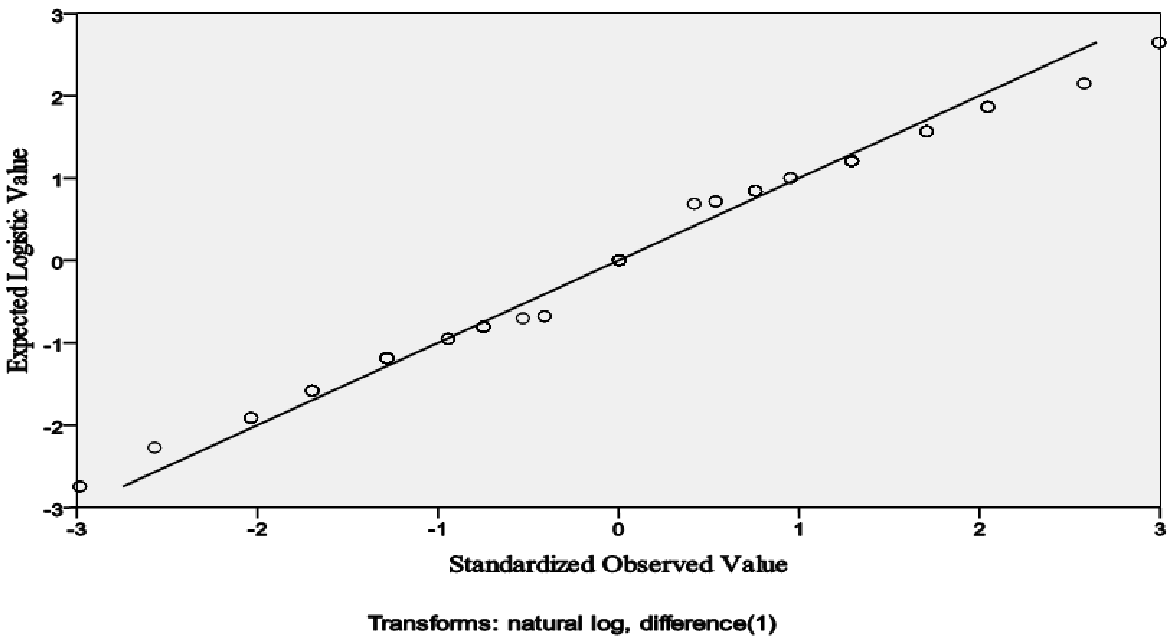

Figure 3. Logistic Q-Q Plot of RPD S-Index

Journal of Agricultural Studies
ISSN 2166-0379
2013, Vol. 1, No. 2

Table 7. Case Processing Summary by Multinomial Logistic Regression

Level of RPD	N	Marginal Percentage
Not-Developed	87	22.6%
Less-Developed	147	38.2%
Semi-Developed	83	21.6%
Developed	29	7.5%
Highly-Developed	39	10.1%
Valid	385	100.0%
Missing	0	
Total	385	

Table 8. Model Fitting Information

Model	Model Fitting Criteria			Likelihood Ratio Tests		
	AIC	BIC	-2LL	Chi-Square	df	Sig.
Intercept Only	1.127	1.143	1.119			
Final	1.150	1.245	1.102	16.430	20	.690

A more useful measure to assess the utility of a multinomial logistic regression model is classification accuracy, which compares predicted group membership based on the logistic model to the actual known group membership, which is the value for the dependent variable. The suggested benchmark to characterize a multinomial logistic regression model as useful is a 25% improvement over the rate of accuracy achievable by chance alone. Even if the independent variables had no relationship to the groups defined by the dependent variable, still it's expected to be correct in predictions of group membership some percentage of the time. This is referred to as "by chance accuracy" (Schwab, 2006). As indicated by Schwab (2006) the estimate of "by chance accuracy" that have used in this study, is the proportional by chance accuracy rate, computed by summing the squared percentage of cases in each group. The proportional by chance accuracy rate was computed by calculating the proportion of cases for each group based on the number of cases in each group in the Case Processing Summary table (table 7), and then squaring and summing the proportion of cases in each group ($0.226^2 + 0.382^2 + 0.216^2 + 0.075^2 + 0.101^2 = 0.257$). To characterize model as useful, the overall percentage accuracy rate produced by SPSS at the last step in Classification table (table 9) [37.7%] in which variables are entered to 25% more than the proportional by chance accuracy (calculated from Case Processing Summery) should be compared.

Journal of Agricultural Studies
ISSN 2166-0379
2013, Vol. 1, No. 2

Table 9. Classification

Observed	Predicted					
	Not-Developed	Less-Developed	Semi-Developed	Developed	Highly-Developed	Percent Correct
Not-Developed	0	82	5	0	0	0.0%
Less-Developed	2	143	2	0	0	97.3%
Semi-Developed	5	76	2	0	0	2.4%
Developed	0	28	1	0	0	0.0%
Highly-Developed	0	37	2	0	0	0.0%
Overall Percentage	1.8%	95.1%	3.1%	0.0%	0.0%	37.7%

The proportional by chance accuracy criteria is 32.1% (1.25 X 25.7% = 32.1%). Since, the classification accuracy rate from Classification table in SPSS output was 37.7%, and is greater than the proportional by chance accuracy criteria of 32.1%; therefore the criterion for classification accuracy is satisfied and model is useful. The Goodness-of-Fit (table 10) provides further evidence of good fit for the model. Both the Pearson and Deviance statistics are chi-square based methods and subject to inflation with large samples. Here, the lack of significance as indicator of good fit is being interpreted. To be clear, p-value should be greater than established cutoff (0.05) to indicate good fit of model (Schwab, 2006). As shown in table 10, Deviance and Pearson values are greater than 0.05, showing model is fit. The Pseudo R-square in table 11, displays three metrics which have been developed to provide a number familiar to those who have used traditional and standard multiple regression. They are treated as measures of effect size, similar to how R-square is treated in standard multiple regressions.

Table 10. Goodness-of-Fit

Statistics	Chi-Square	df	Sig.
Pearson	1512.275	1488	.325
Deviance	1097.634	1488	1.000

Table 11. Pseudo R-Square

Cox and Snell	.042
Nagelkerke	.044
McFadden	.015

Table 12. Likelihood Ratio Tests

Effect	Model Fitting Criteria			Likelihood Ratio Tests		
	AIC of Reduced Model	BIC of Reduced Model	-2 LL of Reduced Model	Chi-Square	df	Sig.
Intercept	1.151	1.230	1.111	8.898	4	.064
STREF	1.143	1.222	1.103	.774	4	.942

TM	1.145	1.224	1.105	3.049	4	.550
ID	1.145	1.224	1.105	2.421	4	.659
FT	1.149	1.228	1.109	6.242	4	.182
MR	1.145	1.224	1.105	2.930	4	.570

Table 13. Parameter Estimates

RPD S-Indexa		B	Std. Error	Wald	df	Sig.	Exp(B)	99% Confidence Interval for Exp(B)	
								Lower Bound	Upper Bound
Not-Developed	Intercept	2.668	.959	7.736	1	.005			
	STREF	.003	.018	.028	1	.868	1.003	.957	1.051
	TM	.017	.049	.125	1	.724	1.017	.897	1.154
	ID	-.013	.038	.117	1	.732	.987	.895	1.089
	FT	-.032	.017	3.513	1	.061	.969	.928	1.012
	MR	.058	.085	.468	1	.494	1.060	.851	1.321
Less-Developed	Intercept	1.903	.887	4.600	1	.032			
	STREF	-.002	.017	.021	1	.885	.998	.954	1.043
	TM	.019	.046	.183	1	.669	1.020	.907	1.147
	ID	-.004	.035	.012	1	.912	.996	.910	1.090
	FT	-.021	.016	1.861	1	.173	.979	.940	1.019
	MR	.102	.079	1.676	1	.195	1.108	.904	1.358
Semi-Developed	Intercept	2.409	.964	6.248	1	.012			
	STREF	-.007	.019	.123	1	.725	.993	.947	1.042
	TM	.066	.049	1.829	1	.176	1.069	.942	1.213
	ID	-.037	.038	.948	1	.330	.963	.873	1.063
	FT	-.020	.017	1.414	1	.234	.980	.938	1.024
	MR	.016	.085	.035	1	.852	1.016	.816	1.265
Developed	Intercept	1.672	1.221	1.876	1	.171			
	STREF	.007	.022	.092	1	.761	1.007	.952	1.064
	TM	.006	.061	.011	1	.915	1.006	.861	1.176
	ID	.021	.050	.183	1	.669	1.022	.898	1.162
	FT	-.047	.021	5.063	1	.024	.954	.905	1.007
	MR	.059	.110	.287	1	.592	1.061	.799	1.408

a. The reference category is: Highly Developed.

However, these metrics do not represent the amount of variance in the outcome variable accounted for by the predictor variables. Higher values indicate better fit, but they should be interpreted with caution (Starkweather and Herrington, 2011). The statistics in the Likelihood Ratio Tests have shown in table 12; are the same types as those reported for the null and full models above in the Model Fitting Information (table 8). However, each element of the model is being compared to the full model in such a way as to allow the researcher to determine if it (each element) should be included in the full model. In other words, does each element (predictor) is contributed meaningfully to the full effect? To be clear, if the p-value is less than established cutoff (0.05) for a predictor; that predictor contributes significantly to the full (final) model. Here none of the predictor variables p-values are significant. The maximum likelihood method used to calculate multinomial logistic regression is an iterative fitting process that attempts to cycle through repetitions to find an answer. Sometimes, the method will break down and not be able to converge or find an answer. Sometimes the method will produce wildly improbable results, reporting that a one-unit change in an independent variable increases the odds of the modeled event by hundreds of thousands or millions. These implausible results can be produced by “multicollinearity”, categories of predictors having no cases or zero cells, and complete separation whereby the two groups are perfectly separated by the scores on one or more independent variables [this is the case of this study as per SPSS output warning] (Schwab, 2006). Therefore, as shown in table 12, when none of p-values are significant, that means proper treatment has to be done on multicollinearity and some of independent variables have to be dropped to make model fit. In addition, the Parameter Estimates (table 13) shows the logistic coefficient (B) of each predictor variable for each alternative category of the outcome variable. The logistic coefficient is the expected amount of change in the Logit for every one unit change in the predictor variables. The Logit is what is being predicted; it is the odds of membership in the category of the outcome variable which has been specified. The closer a logistic coefficient is to zero, the less influence the predictor has in predicting the Logit. For instance, in first category (Not-developed) the least effective predictor independent variable is STREF, following by TM & MR. ID & FT has negative affect (for abbreviation reference, please see the fig. 1). Similarly, in each category the least and most effective predictor independent variables can be identified. The table also displays the standard error, Wald statistic, df, Sig. (p-value); as well as the Exp(B) and confidence interval for the Exp(B). The Wald test (and associated p-value) is used to evaluate whether or not the logistic coefficient is different than zero. The Exp(B) is the odds ratio associated with each predictor. It is expected predictors which increase the Logit to display Exp(B) greater than 1.0, those predictors which do not have an effect on the Logit will display an Exp(B) of 1.0 and predictors which decease the Logit will have Exp(B) values less than 1.0 (Starkweather and Herrington, 2011). For example, in fourth category, STREF, TM, ID & MR, which all have Exp(B) greater than 1.0; are increase the Logit, means they have higher probability to have positive effect on rice production at Developed level. As seen in chart 3, in categories of rice development, depend upon the level of development, each time, one of the predictor variables has the least effect to increase the odd. That means, for instance, in Less-Developed category, ID is expected to be the least effective one (based on its Exp(B) value); while in Developed category, STEF is expected to be the least effective one. However, based on the B, Wald and Exp (B) values in each category, even at this stage the effectiveness of any of predictor variables (independent variables) of the proposed model can be predicted.

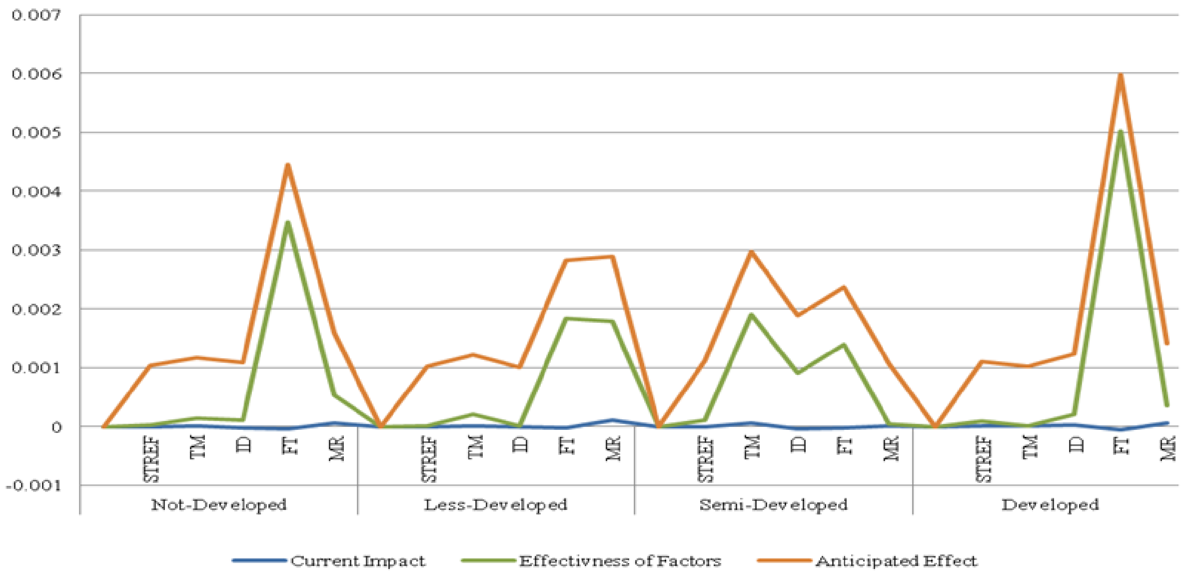

Figure 4. Predictor Variables Estimates and Expected B Values on DV

5. Treating Multicollinearity (MC)

As Belsley et al. (1980) and Greene (1993) indicated even when pseudo R-squares is large (like what has seen in ordinal regression), there might be collinearity which can increase estimates of parameter variance; yield models in which no variable is statistically significant; produce parameter estimates of the “incorrect sign” and of implausible magnitude; create situations in which small changes in the data produce wide swings in parameter estimates; and, in truly extreme cases, prevent the numerical solution of a model. These problems can be severe and sometimes crippling (O’brien, 2007). Having said that multicollinearity (MC) is a multivariate problem, not a bivariate problem, means a simple perusal of the bivariate correlation matrix is not sufficient to eliminate consideration of the problem of multicollinearity. The problem is not only those two independent variables are highly correlated, but that one independent variable is highly correlated with all of the other independent variables. That means it is needed to examine the R-square of each independent variable regressed on the other independent variables (Ethington, 2011). To detect the multicollinearity, examining the correlations (continuous and ordinal variables) and associations (nominal variables) between independent variables can be a solution. However, in some situation, when no pair of variables is highly correlated, but several variables are involved in interdependencies, it may not be sufficient. Then, it is better to use multicollinearity diagnostic statistics produced by linear regression analysis (Variance Inflation Factor [VIF] & Tolerance) (Technical Academic Support, 2011). The collinearity diagnostic statistics are based on the independent variables only, so the choice of the dependent variable does not matter. Therefore, Tolerance and Variance Inflation Factor (VIF) for each variable can be examined. Since for each independent variable, Tolerance = 1 – R-square, where R-square is the coefficient of determination for the regression of that variable on all remaining independent variables, low values indicate high multivariate correlation. Variance Inflation Factor is 1/Tolerance, it is always >= 1 and it is the number of times the variance of the corresponding parameter estimate is increased due to multicollinearity as compared to as it would be if there were no multicollinearity (Ibid). There is no formal cutoff value to use with Variance Inflation Factor for determining presence of multicollinearity. Values of Variance Inflation Factor exceeding 10 are often regarded as indicating multicollinearity, but in weaker models, which is often the case in logistic regression, values above 2.5 may be a cause for concern (see, P.D. Allison, Logistic Regression Using the SAS System, SAS Institute, 1999).

Table 14. Linear Regression Coefficients for Model I

Model	Un-standardized Coefficients		Standardized Coefficients	t	Sig.	Correlations			Collinearity Statistics	
	B	Std. Error	Beta			Zero-order	Partial	Part	Tolerance	VIF
(Constant)	1.45	0.121		12.01	.000					
Market Regulations	-0.017	0.011	-.110	-1.51	.131	-.06	-.07	-.07	.49	2.0
Farming Technologies	0.001	0.002	.042	0.52	.600	-.00	.02	.02	.40	2.4
Infrastructure Development	*-0.002*	*0.005*	*-.049*	*-0.48*	*.628*	*-.02*	*-.02*	*-.02*	*.26*	*3.8*
Trade & Marketing	*0.005*	*0.006*	*.093*	*0.83*	*.402*	*-.00*	*.04*	*.04*	*.21*	*4.6*
STRE & Finance	*0.000*	*0.002*	*-.015*	*-0.13*	*.890*	*-.01*	*-.00*	*-.00*	*.22*	*4.4*
Dependent Variable: RPD Index										

Table 14 shows, all the regressions expect the last three independent variables that have low Tolerance and high value for Variance Inflation Factor indicating a high degree of multicollinearity. It therefore, makes sense to omit variables with insignificant coefficient, but one at a time (Ramanathan, 1993). By looking into Standardized Coefficient column, Infrastructure Development has the least significant coefficient value (-0.49) and has to be omitted. Regression by the left independent variables again shows STREF & Finance has low Tolerance (0.234) and high Variance Inflation Factor (4.269). Therefore, in the next step, this variable also omitted and after that regression by the remaining independent variables (table 15) shows there is no more multicollinearity. However, the problem with this solution (eliminating one or more of the independent variables that are highly correlated with the other independent variables) is that dropping Xj from the equation means that the ith regression coefficient no longer represents the relationship between the Y and Xi controlling for Xj and any other independent variables in the model. The model being tested has shifted, and this often means that the theory being tested by the model has changed. Simply dropping Xj because it is highly correlated with the Xi or using step-wise regression to select variables (typically the selected variables are not "too highly correlated" with each other) leads to a model that is not theoretically well motivated. At times, however, it may be reasonable to eliminate or combine highly correlated independent variables, but doing this should be

Journal of Agricultural Studies
ISSN 2166-0379
2013, Vol. 1, No. 2

theoretically motivated (O'brien, 2007). However, combining independent variables can be done when they are conceptually similar. Since in this model, omitted variables are not measuring the same concept, therefore, the only solution was to eliminate both of them stepwise.

Table 15. Linear Regression Coefficients for Model II

Model	Un-standardized Coefficients		Standardized Coefficients	t	Sig.	Correlations			Collinearity Statistics	
	B	Std. Error	Beta			Zero-order	Partial	Part	Tolerance	VIF
(Constant)	1.44	.117		12.41	.000					
Market Regulations	-.017	.011	-.115	-1.61	.107	-.063	-.082	-.082	.510	1.96
Farming Technologies	.001	.002	.027	.390	.697	-.004	.020	.020	.529	1.89
Trade & Marketing	.003	.004	.054	.690	.490	-.006	.035	.035	.429	2.32
Dependent Variable: RPD Index										

6. Conclusion

Comparing two regressed models resulted in two findings. First, all model selection statistics in Model II (table 14) are better than Model I (table 15). Second, remaining independent variables in Model II has higher Tolerance and lower Variance Inflation Factor values. Therefore, overall, the Model II is superior and now can be taken to Structural Equation Modeling for final analysis and exploring more relationships among the variables. However, due to importance of the variables, despite diagnosing multicollinearity, having results of ordinal and multinomial regression (model is fit and there is at least on latent variable), the Model I will be run into Structural Equation Modeling for further analysis. It is quite possible for an independent variable to be non-significant in an ordinal or logistic regression model, yet its total direct and indirect effects to be significant in a structural equation modeling or path model. The primary reason is that in logistic regression, any given independent variable is controlled for all other independent variables in the model, whereas in a path model independent variables are controlled only for incoming arrows. Other independents variables not in the prior causal chain for the independent variable under consideration are not controlled. The outcome is that a given independent variable is less likely to be found significant in a logistic or OLS model than in a path or structural equation modeling model. Put another way, in a path model, where there is no direct or indirect arrow from one independent variable to a second independent variable, the researcher has posited that there is no control effect, whereas in a logistic or OLS model no such assumptions are made by the researcher who, in fact, is unable a priori to rule out any independent variable as a control for any other independent variable (Garson, 2011). Therefore, the results of structural equation modeling application on this model are due to be reported in next forthcoming paper.

References

Allison, P. D. (1999). *Logistic Regression Using the SAS System*, SAS Institute,

Chen, Ch. K., & J. Hughes Jr. (2004). *Using ordinal Regression Model to Analyze Student Satisfaction Questionnaires*, Association for Institutional Research, Vol. 1.

Ethington, B., (2011), *Multicollinearity*, University of Memphis.

Fallah, A., (2007), *Challenges and Solutions in Rice production in Iran*, Institute of Rice Researches.

Garson, G. D., (2011), *Logistic Regression*, Statnotes, retrieved from portal (http://faculty.chass.ncsu.edu/garson/PA765/logistic.htm).

Karagiannis, N., & Z. Madjd-Sadjadi, (2007), *Modern state intervention in the era of globalization; New directions in modern economics*, Edward Elgar Publishing.

Anonymous, (2010), *Market Info*, Retrieved from Rice International Conference & Exhibition portal, www.riceexhibition.com.

Malekmohammadi, I., S. J. Farajollah Hosseini, S. Yazdani, & Sh. Shadfar, (2011), Analytical Globalized State Intervention Model in Rice Production Development, *African Journal of Agricultural Research, 6*(9), 2015-2025.

Malekmohammadi, I., (2008), Variable Refinery Process to Ensure Research Unbiasedness (Validity) and Invariance (Reliability) in Agricultural Extension and Education, *American Journal of Agricultural and Biological Sciences, 3*(1), 342-347. http://dx.doi.org/10.3844/ajabssp.2008.342.347

Anonymous, (2011), Technical Academic Support, *Multicollinearity in Logistic Regression*, retrieved from Center for Statistical Computing Support of University of Kentucky,

Najafi, B., (2000), *Review State Policies in Rice, Issues and Solutions*, Agricultural Economy and Development, Vol. 8, Issue 31, pp 7.

Obanil, R., & E. Dano, (2005), *State Intervention in the Rice Sector in Selected Countries, Lessons from Major Rice production Countries*, Oxfam.

O'brien, R. M., (2007), A Caution Regarding Rules of Thumb for Variance Inflation Factor, *Quality & Quantity, 41*, 673-690. http://dx.doi.org/10.1007/s11135-006-9018-6

Plubin, B. & N. Techapunratanakul, (2006), *Ordinal Regression Analysis in Factors Related to Sensorial Hearing Loss of the Employees in Industrial Factory in Lampang Thailand*, Proceeding of the 2nd IMT-GT Regional Conference on Mathematics, Statistics and Application, University Sains Malaysia, Malaysia.

Ramanathan, R.. (1993). *Statistical Methods on Econometrics,* Academic Press.

Schwab, J. (2006). *Data Analysis and Computers II, Multinomial Logistic Regression Basic Relationships*, Course Handout Number: SW 388R7, retrieved from University of Texas at

Austin, School of Social Work portal (http://www.utexas.edu/courses/schwab/)

Starkweather, J., & R. Herrington. (2011). *Logistic Regression (Multinomial),* SPSS Short Course, Model 9, Computing and Information Technology, University of North Texas, retrieved from Center, Research and Statistical Support portal (http://citc.unt.edu/)

Statistics Centre of Iran. (2010). *Journal of Annual Statistics of Iran*, retrieved form portal of Statistics Center of Iran (www.isc.org).

Sinayiee, S., (2006), *Analyzing Affects of Monetary and Financial Policies on Key Economic Factors in Agriculture Sector*, Shahid Bahonar University of Kerman.

Anonymous, (2009), *Structural Adjustment and Necessity of Reconsideration in Ministry of Agriculture Mission and Strategy*, Ministry of Agriculture think-thank, Published on Ministry of Agriculture portal.

Anonymous, (1996), *Trends of yield and productivity of modern rice in irrigated rice systems in Asia*, retrieved from FAO portal.

Workman, D., (2008), *Top Rice Producing Countries, Asian Nations Generate over 90% of World Rice Production*, Retrieved from www.suit101.com.

Anonymous, (2010), *World Rice Production 2009-2010*, retrieved from www.rice-trade.com.