Kashmir Economic Review, Volume 32, Issue 1, June 2023

Kashmir Economic Review
ISSN(P): 1011-081X, ISSN(E): 2706-9516
http://www.ker.org.pk

The Impact of Division-Level Production Outcomes upon Punjab Aggregate Wheat Production: An Application of Correlated Component Regression Approach

ABSTRACT

This study analyzes the relative importance of division-level production outcomes in predicting the Punjab outcome for wheat crops using the annual time-series data from 1982 to 2020. A newly developed regression analytic approach, called correlated component regression (CCR), was applied to overcome suppression effects and multicollinearity data problems. Standardized regression coefficients have been used to determine the relative importance of each division. Herfindahl-Hirschman Index (HHI) was applied to measure the geographic concentration of the division-level impacts. The empirical analysis was used at two time periods in which the first period includes the crop years from 1987 to 2003 and the second period covers 2004 to 2020 crop years. The regression results provide an HHI value of 1175 during 1987-2003 and 1351 during 2004-2020. A smaller value of HHI during the period 1987-2003 indicates that the technological change along with high-yield varieties, use of fertilizers, and pesticides have led to a greater concentration for wheat in this period, and the opposite was true for 2004-2020.

AUTHORS

Muhammad Naveed
Lecturer, Department of Economics, Government Allama Iqbal Graduate College Sialkot, Pakistan.
Author's Contributions: 1,2,3,4,5,6
m_naveed_skt@yahoo.com
https://orcid.org/0000-0003-3348-3830

Hafsa Hina *
Assistant Professor, Pakistan Institute of Development Economics, Islamabad, Pakistan.
Author's Contributions: 1,2,3,4,5,6
hafsahina@pide.org.pk
https://orcid.org/0000-0001-9931-1957

Keywords
Correlated Component Regression, Production Performance Index, Relative Significance in Prediction, Punjab, Geographical Concentration, Technological Change

JEL Classification
C19; Q10

Please cite this article as:
Naveed, M., & Hina, H. (2023). The impact of division-level production outcomes upon Punjab aggregate wheat production: An application of correlated component regression approach. *Kashmir Economic Review*, 32(1), 66-79.

*** Correspondence author**
Author's contribution in the article: 1- Conceived and designed the analysis, 2- Reviewed and compiled the literature, 3- Collected the data, 4- Contributed data or analysis tools, 5- Performed the analysis, 6- Wrote the paper, 7- Financial support for the conduct of the study, 8-Other

1. INTRODUCTION

Agriculture is considered the mainstay of Pakistan's Economy. Development in this sector reduces poverty and ensures sustainable food security in the country. According to the Economic Survey of Pakistan 2018-19, the agriculture sector has contributed 18.5 percent to Pakistan's gross domestic product (GDP) and provided 38.5 percent of the domestic labor force. Although the share of the agriculture sector has decreased in recent years, it is still considered a vital sector of the national economy as its good performance has a significant role in economic growth and poverty reduction. Important crops, namely cotton, rice, wheat, maize, and sugarcane, contribute predominantly to the agriculture sector of the national economy. Wheat is considered the leading staple food in Pakistan. In 2019-20 total wheat production of 24946 thousand tonnes, and a yield of 2827 Kgs per hectare was recorded in the country. Moreover, the share of wheat to value-added in agriculture and GDP of Pakistan was 8.7 percent and 1.7 percent respectively in 2019-20.

The largest production share in the total wheat production of Pakistan comes from Punjab. The economy of Punjab has a significant impact on the economy of Pakistan because the total percentage of the economy of Punjab in the GDP of Pakistan was 54.2 percent in 2017-18. Moreover, the contribution of agriculture in the economy of Punjab was 23 percent, 20.8 percent, and 20.20 percent in 2012-13, 2015-16, and 2017-18 respectively. During 2018-19, 38.94 percent of the total cropped area was cultivated wheat in Punjab. Punjab itself produced 19401.86 thousand tonnes of wheat in 2019-20, almost 78 percent of the total wheat produced in Pakistan. Punjab has nine divisions, and each division contributes differently to the whole wheat production of Punjab.

During the 1987 to 2003-time period, the highest average production share was held by the Multan division at 14.89 percent, followed by Bahawalpur (13.57 percent), Gujranwala (13.50 percent), Faisalabad (13.47 percent), D.G. Khan (11.46 percent), and Sahiwal (11.30 percent). While at the same period, the Rawalpindi division has contributed the lowest average production share at 4.66 percent. From 2004 to 2020, Bahawalpur had the most significant average production share at 15.29 percent, followed by Gujranwala (14.35 percent), Multan (13.87 percent), Faisalabad (13.65 percent), and D. G. Khan (12.76 percent). In contrast, the Rawalpindi division contributed the lowest average production share at 3.82 percent (Crop Reporting Service, Punjab Lahore).

As Punjab is a major producer of wheat crop in Pakistan, thus a significant change in wheat production in Punjab may affect the national wheat production. Therefore, it is crucial to know Punjab divisions' behavior and relative importance for Provincial wheat production. A careful review of the academic literature to date has turned up no evidence of any studies that have examined the relative implications of division level wheat contribution to Punjab provincial supply. To address this research gap, this study has the following objectives:

- The main objective of this study is to apply a recently developed regression analytic technique, called a correlated component regression (CCR), to measure the relative importance of division-level production outcomes in predicting the provincial production outcome for the wheat crop.
- A second important objective of this study is to analyze whether the technological change along with high-yield varieties, use of fertilizers, pesticides and policy factors would lead to a greater or lower geographic variation of the relative importance of each division upon the provincial aggregate.

To determine the relative importance of each division on Punjab wheat production, regression analyses were conducted using the Punjab production performance index as the dependent variable and each division's production performance index as independent variables. Standardized regression coefficients were calculated from these regressions, and these coefficients were used to rank the divisions based on their significance in influencing Punjab provincial wheat production. Due to issues with datasets such as its multicollinearity as well as sparsity, simple ordinary least squares (OLS) regression methods were not suitable as they resulted in large standard errors and unreliable insignificant coefficient estimates. Therefore, a regression technique called correlated component regression (CCR), developed in 2011, and was applied in this study.

CCR is specifically designed to handle sparse and multicollinear datasets, providing more reliable and stable coefficient estimates. By applying this methodology, the study aimed to provide a comprehensive understanding of the relative importance of each division's production outcomes in influencing Punjab wheat production. The use of CCR allowed for more robust and accurate estimations, overcoming the challenges posed by sparse and multicollinear data.

The second significant goal of this study is to find out the influence of technology and policy factors on the ranking of the 9 divisions and the geographical concentration of their relative importance. To achieve this, the time-series dataset was divided into two distinct periods: the years 1987 to 2003, and the years 2004 to 2020. The Herfindahl-Hirschman Index (HHI) was used in the study to determine the geographic concentration of importance. The HHI was computed by considering the percentage shares of the absolute values of the standardized regression coefficients. This index provides a measure of the concentration of importance across different divisions. By computing the HHI for both time periods in relation to wheat production, it became possible to assess the impact of technology and policy factors on the concentration of wheat production. The results from these calculations shed light on the degree to which technology and policy influences have contributed to the geographic concentration or dispersion of importance in the context of wheat production.

2. LITERATURE REVIEW

In literature, most of the studies have focused on the impact of climate change, input requirements such as plowing, seed rate, irrigation, fertilizer, farm size, and change in the conventional agriculture practices to the use of effective microorganisms upon wheat yields.

The studies that focused on the impact of climate-related variables, namely CO_2, rainfall, temperature, environmental degradation, on wheat production in Pakistan includes Janjua et al., 2013), Qureshi and Iglesias (1994), Arooj et al. (2018). Interestingly the finding of these studies does not align with each other such as Janjua et al. (2013) haven't found the impact of climate change on wheat production in Pakistan whereas fertilizers would be the only remedy to counter any deficiency of wheat production. Empirical results of Arooj et al. (2018) showed that environmental degradation has a significant impact on wheat production in Pakistan and a 7.2 percent increase in wheat production was caused by a one percent increase in wheat area. Likewise, Chandio et al. (2019) found fertilizer consumption, cultivated area, and support price have a positive and significant effect on wheat production. But on the other hand, Rauf et al. (2017) found that the use of fertilizer for the production of wheat and the yield of wheat does not maintain a significant results association.

Siddiqui et al. (2012) have studied the effect of climate change on wheat, cotton, rice, and sugarcane in Punjab, taking the annual time series data from 1980 to 2008. Estimated results of the fixed effect model show that climate change positively impacts wheat productivity. In contrast, the impact of climate change was negative for sugarcane, rice, and cotton.

Improvement in crop yields not only achieves food sustainability but also enhances its contribution to gross domestic product. The agriculture sector contributes 23% to GDP. Raza et al. (2012), Rehman et al. (2016), and Ali et al. (2020) studied the nexus of agriculture gross domestic product (AGDP) with wheat production, maize, and rice and confirmed the significant influence of these crops in AGDP. Rauf et al. (2017) also showed that there exists a powerful and highly significant relationship between area under cultivation for wheat and agriculture GDP. It is not necessary that all crops significantly affect the agricultural GDP as due to the higher cost of production of any crop may adversely affect the agricultural GDP such as Rehman and Jingdong (2017) analyzed the relationship between agricultural gross domestic product (GDP) and major food crops, including rice, wheat, cotton, sugarcane, corn, and tubers in China using annual time series data from 1980 to 2015. The study's empirical analysis shows that the production of wheat, cotton, corn, tubers, and sugarcane has a positive and significant impact on China's agricultural gross domestic product. However, rice production has a negative and insignificant impact on the agricultural GDP of China.

Pakistan cannot become a food-sufficient country until or unless it shifts from conventional agriculture practices to advanced ones. Hussain et al. (1999) intensively studied the use of effective microorganisms on rice and wheat. The results of their three-year study in Pakistan showed that effective microorganisms can improve soil quality; increase the growth, increase the yields and nutrient uptake of rice and wheat; provide plant protection against diseases and allow the farmers to increase their net profits.

From the review of literature, it is clear that most of the studies in the case of Pakistan have focused on the factors that affect the yield of production but a careful review of the academic literature to date has turned up no evidence of any studies that have examined the relative implications of division level wheat contribution to Punjab provincial supply. However, Bullock (2021) has analyzed the impact of state-level production outcomes upon U.S. aggregate soybean and corn production. In the study, the relative importance of each state was determined by the absolute value of the standardized regression coefficients. The degree of concentration was measured by applying the Herfindahl-Hirschman Index. The empirical results were found for two time periods: a pre-Genetic Modification cover 1975 to 1995 crop years and a post-Genetic Modification include 1996 to 2017 crop years. The study results show that U.S. corn production was geographically less concentrated in state-level importance, and the opposite was true for soybean production.

3. DATA AND METHODOLOGY

In the present research, the time series data at an annual frequency from 1982 to 2020 on wheat production is used for nine divisions, namely, Rawalpindi, Gujranwala, Lahore, Faisalabad, Sargodha, Sahiwal, Multan, Bahawalpur, and Bahawalnagar. The Punjab level aggregate data of wheat production is also used in this study. The data on wheat production is taken from the Directorate of Agriculture Crop Reporting Service, Punjab Lahore (crs.agripunjab.gov.pk) and Crops, Area an d Production (by districts) volume I issued by the Federal Bureau of Statistics. The wheat production is measured in tonnes. This study analyzes the data by applying the Correlated Component Regression (CCR) technique. The Production Performance Index (PPI) was constructed for two time periods in which the first period includes the crop years from 1987 to 2003 and the second period covers 2004 to 2020 crop years. Herfindahl-Hirschman Index (HHI) has also been applied for geographic concentration. A comprehensive introduction of all these measures is given below.

3.1 Production Performance Index (PPI)

In measuring the relative under-or over-performance of wheat production per year, a metric called production performance index (PPI), first developed by Bullock (2021), was constructed as a proxy for the divisionally and provincially production level relative to recent previous years. This study also follows the index constructed by Bullock (2021). This index is defined as follows:

$$PPI_t = P_t - OA(P_{t-1}, P_{t-2}, P_{t-3}, P_{t-4}, P_{t-5}) \tag{1}$$

Where P_t shows the current year production of wheat in Punjab and each of 9 divisions of the Punjab, and $OA(\cdot)$ indicates the Olympic Average function (calculated by ignoring the minimum and maximum values from the previous five-year values and taking the average of remaining three-year values). Hence, the PPI measures the degree by which the current year's production either decreased (under-performed) or exceeded (over-performed) the average production level from the previous five years. The PPI series is generally stationary and reduces the effects of autocorrelation because it was calculated by taking the lagged differencing. The Olympic average was used to reduce the influence of extremely bad or good production years when fixing the past benchmark of what could be taken as a "normal" level of production for a specific division of Punjab province at the time of comparison.

3.2 Correlated Component Regression (CCR)

The Correlated Component Regression technique (CCR) was developed by Magidson (2011) to address a significant issue encountered in traditional regression approaches, namely multicollinearity and suppression effects. Multicollinearity and suppression effects arise when explanatory variables exhibit moderate to high correlation with each other and have no direct impact on the dependent variable. These problems can lead to large variances and standard errors of estimated regression coefficients, rendering them unstable and statistically insignificant (Pandey & Elliot, 2010). In datasets like the one considered in this study, where the explanatory variables are highly correlated, a model's seemingly good predictive performance may be attributed to overfitting. Magidson (2013) demonstrated that the CCR approach is particularly effective for sparse and multicollinear datasets, as it provides more reliable and stable estimates of regression coefficients. A major advantage of CCR is its ability to mitigate confounding effects caused by high correlations among predictors and yield more robust parameter estimates. Bullock (2021) employed the CCR technique, which we have also adopted in this study. The CCR algorithm employs the ordinary least square (OLS) method to estimate a set of P single-variable regression equations. In the first step, we estimate the loadings for each explanatory variable by running Y on each explanatory variable through the Ordinary Least Square (OLS) technique as follows:

$$\hat{Y} = \hat{\gamma}_g^{(1)} + \hat{\lambda}_g^{(1)} X_g \tag{2}$$

Y indicates the dependent variable and X_g shows independent variables with $g = 1,2,3, \dots, P$; $\hat{\gamma}_g^{(1)}$ and $\hat{\lambda}_g^{(1)}$ are the constant and slope coefficients respectively for the specific independent variable g. The first correlated component variable, S_1, is then defined as the weighted sum of all 1-predictor effects using the slope coefficients obtained from (2) like weights, that is:

$$S_1 = \frac{1}{P} \sum_{g=1}^{P} \hat{\lambda}_g^{(1)} X_g \tag{3}$$

The predictions for Y in the 1-component CCR model are obtained by applying OLS regression of Y on S_1, that is:

$$\hat{Y} = \hat{\alpha}^{(1)} + \hat{\beta}_1^{(1)} S_1 \tag{4}$$

With the relevant CV metric, CV-R^2 or CV-MSE has been saved from the above regression for later determination of the optimal number of correlated components retained in the final model. The component variable S_1 is called the direct effects component because it shows the direct effect of each explanatory variable upon the outcome variable without capturing the suppressor effects.

Similarly, the 2-component variable S_2 is constructed by first estimating the following model for each predictor through the application of simple OLS, that is:

$$\hat{Y} = \hat{\gamma}_g^{(2)} + \hat{\lambda}_{1,g}^{(2)} S_1 + \hat{\lambda}_g^{(2)} X_g \tag{5}$$

The second correlated component variable S_2 is defined as the weighted sum of all 2-predictor effects using the slope coefficients of the explanatory variables estimated from (5) like weights, that is:

$$S_2 = \frac{1}{P} \sum_{g=1}^{P} \hat{\lambda}_g^{(2)} X_g \tag{6}$$

The predictions for Y in the 2-component CCR model are obtained by applying OLS regression of Y on S_1 and S_2, that is:

$$\hat{Y} = \hat{\alpha}^{(2)} + \hat{\beta}_1^{(2)} S_1 + \hat{\beta}_2^{(2)} S_2 \tag{7}$$

As the component variable S_2 and all the later derived component variables show the correlation with component variable S_1, component variable S_2 and all the later derived component variables capture the suppressor variables effects. It improves the prediction in the component variable model by removing the irrelevant variation from at least one of the explanatory variables with direct effects. This process for the derivations of component variables can continue until the optimal number of components is achieved.

In general, for any K (K is less than P)-component variables, S_k is defined by first estimating the following model through OLS for each of the independent variables, that is:

$$\hat{Y} = \hat{\gamma}_g^{(K)} + \hat{\lambda}_{1,g}^{(K)} S_1 + \hat{\lambda}_{2,g}^{(K)} S_2 + \cdots + \hat{\lambda}_{K-1,g}^{(K)} S_{K-1} + \hat{\lambda}_g^{(K)} X_g \tag{8}$$

After predicting regression equation (8), final correlated component variable S_k is constructed as follows:

$$S_k = \frac{1}{P} \sum_{g=1}^{P} \hat{\lambda}_g^{(k)} X_g \tag{9}$$

The predictions for Y in the k-component CCR model are obtained by applying OLS regression of Y on $S_1, S_2, \ldots, S_k$ that is:

$$\hat{Y} = \hat{\alpha}^{(K)} + \hat{\beta}_1^{(K)} S_1 + \hat{\beta}_2^{(K)} S_2 + \cdots + \hat{\beta}_k^{(K)} S_k \tag{10}$$

Putting (3), (6), and (9) in (10), we obtain the following:

$$\hat{Y} = \hat{\alpha}^{(K)} + \hat{\beta}_1^{(K)} \left(\frac{1}{P} \sum_{g=1}^{P} \hat{\lambda}_g^{(1)} X_g \right) + \hat{\beta}_2^{(K)} \left(\frac{1}{P} \sum_{g=1}^{P} \hat{\lambda}_g^{(2)} X_g \right) + \cdots + \hat{\beta}_k^{(K)} \left(\frac{1}{P} \sum_{g=1}^{P} \hat{\lambda}_g^{(k)} X_g \right) \tag{11}$$

From (11) we can get:

$$\hat{Y} = \hat{\alpha}^{(K)} + \hat{\beta}_k^{(K)} \left(\frac{1}{P} \sum_{g=1}^{P} \hat{\lambda}_g^{(k)} X_g \right) \tag{12}$$

$$\hat{Y} = \hat{\alpha}^{(K)} + \sum_{g=1}^{P} \hat{\beta}_g X_g \tag{13}$$

Where

$$\hat{\beta}_g = \frac{1}{P} \sum_{k=1}^{P} \hat{\beta}_k^K \hat{\lambda}_g^{(k)} \tag{14}$$

Equation (14) shows that the regression coefficient $\hat{\beta}_g$ for an explanatory variable X_g is simply the weighted sum of the loadings where the estimated coefficients obtained from the final component model (10) are used as weights. Estimated coefficients obtained from (14) are called non-standardized regression coefficients. The standard errors of the non-standardized regression coefficients are obtained by using the formula given below:

$$SE(\hat{\beta}_g) = \frac{1}{P}\sqrt{\sum_{k=1}^{K}\left(SE(\hat{\beta}_k^K)\right)^2(\hat{\lambda}_g^k)^2} \tag{15}$$

Where $\hat{\lambda}_g^k$ indicate the loadings and $SE(\hat{\beta}_k^K)$ represent the standard errors of the estimated coefficients of the final correlated component regression model. The divisions are ranked based on the absolute values of standardized regression coefficients. These are calculated as follows:

$$\hat{\beta}_g^* = \left(\frac{\hat{\sigma}_g}{\hat{\sigma}_y}\right) \times \hat{\beta}_g \tag{16}$$

Where $\hat{\sigma}_g$ represents the standard deviation of each independent variable and $\hat{\sigma}_y$the explained variable's standard deviation. Standardized regression coefficients are used to check which explanatory variables have a larger impact upon the explained variable. The standardized regression coefficient indicates the marginal effect (in standard deviations) upon the explained variable of a one standard deviation change in the explanatory variable. The absolute values of the standardized regression coefficients of each division and the percentage share of the total sum of absolute coefficients are used to determine the rank of each division. Divisions with a greater ranking represent that a one standard deviation change in their PPI measure has a greater effect (in standard deviations) upon the provincial PPI measure.

3.3 Herfindahl-Hirschman Index (HHI)

The Herfindahl-Hirschman Index (HHI) was initially developed by Herfindahl and Hirschman. Bullock (2021) applied the HHI as a measure of geographic concentration. This was determined by computing HHI using the absolute values of the percentage-shared standardized regression coefficients. Following the methodology proposed by Bullock (2021), this study also utilizes the HHI to assess the importance of geographic concentration for the division-level impacts during both time periods. The following equation provides a mathematical definition of the HHI:

$$HHI = \sum_{i=1}^{n}(s_i.100)^2 \tag{17}$$

Where n indicates, the number of divisions included in the regression and s_i is the percentage share of the total sum of absolute standardized coefficients for the i^{th} division (in decimal format). Changes in the value of HHI can be used to determine the degree of change in geographic concentration between the two time periods using the percentage share of standardized regression coefficients. The main hypothesis developed was whether the technological change along with high-yield varieties, use of fertilizers, and pesticides would lead to a greater or lower geographic variation of the relative importance of each division upon the provincial aggregate. A smaller value of HHI in any period would imply that the technological change along with high-yield varieties, use of fertilizers, and pesticides have led to greater concentration in that period.
In our analysis, we utilized the COR Express software developed by Statistical Innovations for estimation purposes. To determine the optimal values for two important parameters, namely the number of components (K) and the number of explanatory variables (P) in the CCR model, we employed relevant cross-validation techniques.

4. RESULTS AND DISCUSSION

Two critical parameters, namely the optimal number of correlated component variables K and the number of predictors P, have been determined using relevant cross-validation. To find the optimal value of component variables K, the eight maximum number of correlated components has been set, and cross-validation of 4-fold was used for both periods. Furthermore, we have taken all the nine divisions of Punjab in the estimation model, and the step-down procedure for variable selection has not been applied. During the period 1987-2003 CCR model with 4 component variables maximized the value of CV-R^2, whereas the 7 component variables maximized the value of CV-R^2 during the 2004-2020 period. Therefore, we have taken 4 and 7 component variables in our analysis from 1987 to 2003 and 2004 to 2020, respectively. The cross-validation R^2 for both the study periods are reported in Table 1.

Table 1: Cross-Validation R^2 for Both the Time Periods

Predictors	**Component Variables**	**CV-R^2**	
		1987-2003	**2004-2020**
9	1	0.967	0.977
9	2	0.993	0.990
9	3	0.995	0.993
9	4	0.996	0.995
9	5	0.996	0.995
9	6	0.995	0.996
9	**7**	0.995	0.997
9	8	0.994	0.997

(Source: Authors' calculation).

Table 2: Correlated components regression coefficients and standardized regression coefficients of wheat production for both the periods

Correlated Component	**Coefficient**	**Std. Error**	**T-Stat**	**Standardized Coefficient**	**Share of Standardized Total**
			1987 to 2003		
S1	0.23	0.01	42.75***	1.11	78.01
S2	0.29	0.03	8.39***	0.20	14.29
S3	0.83	0.25	3.27***	0.07	5.03
S4	0.51	0.27	1.87*	0.04	2.68
2004 to 2020					
S1	0.25	0.02	14.77***	0.86	38.17
S2	0.13	0.23	0.55	0.03	1.13
S3	1.73	0.31	5.57***	0.24	10.66
S4	3.07	0.80	3.84***	0.40	17.51
S5	3.08	0.90	3.42***	0.45	20.06
S6	0.61	0.20	3.02***	0.10	4.59
S7	0.43	0.22	1.93*	0.18	7.88

*, ** and *** indicate significantly different from zero at 10%, 5%, and 1% level of significance, respectively. (Source: Authors' calculation).

Table 2 shows the correlated components, non-standardized regression coefficients, and standardized regression coefficients for both periods. These results are obtained from (10). From 1987 to 2003, the first three components are statistically significant at a 1% significance level, whereas the last component is significant at a 10% significance level. However, the coefficient of the first, third, fourth, fifth, and sixth component variables is significantly different from zero at a 1% significance level from 2004 to 2020. At the same time, the coefficient of the last component variable is significant at a 10% level of significance. The correlated component variable S1 has a very high direct effect of 78.01 percent shares of standardized regression coefficients from 1987 to 2003. In contrast, at the same time, the remaining three components S2, S3, and S4, have only 21.99 percent indirect effect. Similarly, during the period 2004 to 2020, the correlated component variable S1 has a very low direct effect 38.17 percent shares of standardized regression coefficients while at the same time, the remaining six components, S2, S3, S4, S5, S6, and S7, have 68.13 percent indirect effect.

Table 3: Individual Division's non-standardized coefficients, standardized regression coefficients, and percentage shares of absolute standardized coefficients during the period of 1987 to 2003

Rank	Division	Coefficients	Std. Error	t-Statistic	Standardized Coefficient	Share
1	Gujranwala	1.1065	0.06	18.62***	0.22	15.54
2	Bahawalpur	1.31	0.07	19.02***	0.21	14.71
3	Multan	0.85	0.04	19.03***	0.19	13.57
4	Lahore	1.05	0.12	8.99***	0.16	11.20
5	Sargodha	2.42	0.28	8.49***	0.14	9.79
6	D.G. Khan	0.99	0.12	8.50***	0.14	9.57
7	Rawalpindi	0.62	0.10	6.25***	0.13	9.22
8	Faisalabad	0.71	0.07	10.51***	0.12	8.52
9	Sahiwal	0.86	0.12	7.20***	0.11	7.87
	Constant	-73763.48	33814.10	-2.18*		
	Herfindahl-Hirschman Index				1175	

*, ** and *** indicate significantly different from zero at 10%, 5%, and 1% level of significance, respectively. (Source: Authors' calculation).

The individual division's non-standardized regression coefficients, standardized regression coefficients along with their statistical significance, and percentage shares ranked by the absolute values of standardized coefficients during the period 1987 to 2003 are shown in Table 3. The estimated results show that the coefficients of all divisions have a positive and statically significant effect on wheat production of Punjab at a 1% level of significance from 1987 to 2003. In addition, the estimated results also specify that during this period, production outcomes in Division Gujranwala, Bahawalpur, and Multan had a major share of 43.82 percent upon the Punjab wheat production outcome. Lahore, Sargodha, D.G. Khan, Rawalpindi, and Faisalabad have 11.20 percent, 9.79 percent, 9.57 percent, 9.22 percent, and 8.52 percent shares respectively in Punjab wheat production. Sahiwal has the lowest share of 7.87 percent in Punjab wheat production. The Herfindahl-Hirschman Index (HHI), based on the standardized regression coefficients shares, has a value of 1175, which is below the 1500 threshold that the U. S. Department of Justice and the Federal Trade Commission consider a low level of concentration.

Table 4: Individual Division's non-standardized coefficients, standardized regression coefficients, and percentage shares of absolute standardized coefficients during the period of 2004 to 2020

Rank	Division	Coefficients	Std. Error	t-Statistic	Standardized Coefficient	Share
1	Gujranwala	1.15	0.11	10.44***	0.32	18.57
2	Multan	1.19	0.17	7.12***	0.28	16.13
3	Faisalabad	1.22	0.85	1.42	0.27	15.50
4	D.G. Khan	1.23	0.26	4.64***	0.23	13.47
5	Rawalpindi	1.02	0.30	3.43***	0.21	12.00
6	Bahawalpur	0.97	0.17	5.62***	0.19	10.97
7	Sahiwal	0.59	0.33	1.81	0.09	5.15
8	Lahore	0.74	0.27	2.72**	0.07	4.29
9	Sargodha	0.48	0.19	2.59**	0.07	3.92
	Constant	1635.37	42545.44	0.04		
	Herfindahl-Hirschman Index				1351	

*, ** and *** indicate significantly different from zero at 10%, 5%, and 1% level of significance, respectively (Source: Authors' calculation).

Table 4 provides the information regarding individual division's non-standardized regression coefficients, standardized regression coefficients, and their statistical significance and percentage shares ranked by the absolute values of standardized coefficients from 2004 to 2020. The estimated results show that the coefficients of all divisions have a positive and statistically significant effect on wheat production of Punjab at 1% and 5% level of significance during the period of 2004-2020 except Faisalabad and Sahiwal divisions which have a positive but statistically insignificant effect on wheat production of Punjab. In addition, the estimated results also specify that during this period, production outcomes in Division Gujranwala and Multan had a major share of 34.70 percent upon the Punjab wheat production outcome. Moreover, Division Faisalabad, D.G. Khan, Rawalpindi, Bahawalpur, Sahiwal, and Lahore have 15.50 percent, 13.47 percent, 12.00 percent, 10.97 percent, 5.15 percent, and 4.29 percent shares respectively in Punjab wheat production. Sargodha has the lowest share of 3.92 percent in Punjab wheat production during this period. The Herfindahl-Hirschman Index (HHI), based on the standardized regression coefficients shares, has a value of 1351, which is below the 1500 threshold that the U.S. Department of Justice and the Federal Trade Commission consider a low level of concentration.

The estimated results of Division-level change in rankings and shares during two time periods (1987-2003) and (2004-2020) for nine divisions of Punjab are reported in Table 5. The level of geographic concentration, based on the absolute shares of standardized regression coefficients, has improved moderately, as shown by the 176-point increase in the HHI value. From 1987 to 2003, the top three divisions (Gujranwala, Bahawalpur, and Multan) hold 43.82 percent share, with Gujranwala holding 15.54 percent Bahawalpur holding 14.71 percent, and Multan holding 13.57 percent share. During the period 2004 to 2020, a total of 45.67 percent share has been held by these three divisions with 18.57 percent share was held by Gujranwala, Bahawalpur held 10.97 percent share, and Multan held 16.13 percent share. Faisalabad Division has improved its ranking by +5, whereas D. G. Khan, Rawalpindi, and Sahiwal have improved their rankings by +2 places between the two study periods under analysis. However, the Bahawalpur, Lahore, and Sargodha divisions have decreased their ranking by -4 places during the two study periods. The Faisalabad division enjoyed the largest improvement in share +6.98 percent, followed by D. G. Khan +3.89 percent and Gujranwala +3.03 percent. Lahore and Sargodha have the largest decrease in the share of -6.91 percent and -5.87 percent, respectively, during two time periods.

Table 5: Division-Level Change in Rankings and Shares between Two Time Periods

Division	**Rank**			**Shares**		
	1987-2003	**2004-2020**	**Change**	**1987-2003**	**2004-2020**	**Change**
Gujranwala	1	1	0	15.54	18.57	+3.03
Bahawalpur	2	6	-4	14.71	10.97	-3.74
Multan	3	2	+1	13.57	16.13	+2.56
Lahore	4	8	-4	11.20	4.29	-6.91
Sargodha	5	9	-4	9.79	3.92	-5.87
D.G. Khan	6	4	+2	9.57	13.47	+3.89
Rawalpindi	7	5	+2	9.22	12.00	+2.78
Faisalabad	8	3	+5	8.52	15.50	+6.98
Sahiwal	9	7	+2	7.87	5.15	-2.72
HHI	**1175**	**1351**	**176**			

(Source: Authors' calculation).

To check the validity of the model in both periods, certain diagnostic and stability tests are applied, and the estimated results of these tests are shown in Table 6. The empirical analysis of the diagnostic tests indicates that the model has not shown any problem of non-normality, autocorrelation, heteroscedasticity, and instability of the estimated parameters during both periods. Moreover, Figure 1 and Figure 2 represent that the cumulative sum of square residuals (CUSUMSQ) during both periods does not cross the straight line at a 5% level of significance. It indicates the non-presence of structural instability in the estimated parameters of the model during both periods.

Table 6: Diagnostic Tests for both the Time Periods

Test	Value of Test Statistic	Probability Value	Critical Value
Diagnostic tests of the model during 1987-2003			
Normality Test (Jarque Bera)	0.13	0.94	$\chi^2_{0.05(2)} = 5.99$
Serial Correlation LM Test	0.72	0.39	$\chi^2_{0.05(1)} = 3.84$
ARCH Test	0.08	0.77	$\chi^2_{0.05(1)} = 3.84$
Ramsey Reset Test	0.04	0.84	$\chi^2_{0.05(1)} = 3.84$
Diagnostic tests of the model during 2004-2020			
Normality Test (Jarque Bera)	0.09	0.96	$\chi^2_{0.05(2)} = 5.99$
Serial Correlation LM Test	2.65	0.10	$\chi^2_{0.05(1)} = 3.84$
ARCH Test	0.10	0.76	$\chi^2_{0.05(1)} = 3.84$
Ramsey Reset Test	0.05	0.82	$\chi^2_{0.05(1)} = 3.84$

(Source: Authors' calculation).

Figure 1: Cumulative Sum of Square Residuals of the Model during 1987-2003

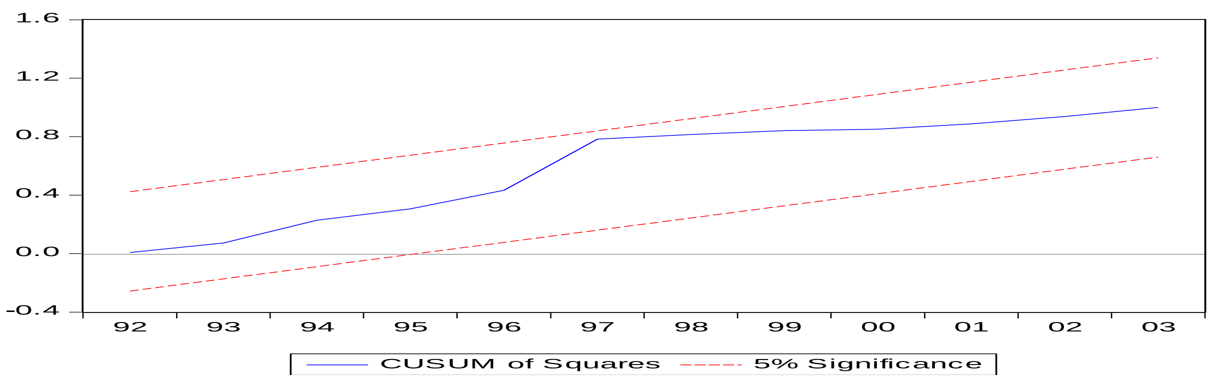
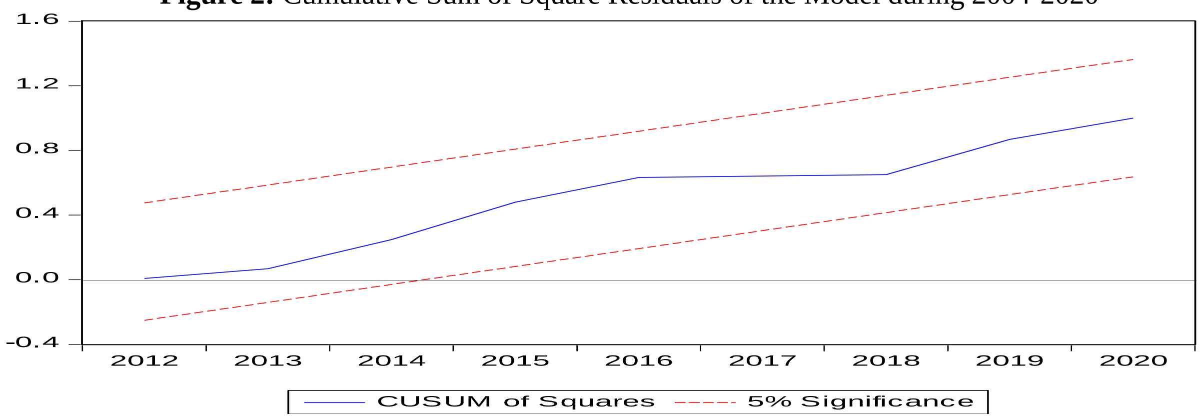


Figure 2: Cumulative Sum of Square Residuals of the Model during 2004-2020



5. CONCLUSION AND POLICY IMPLICATIONS

This study analyzes the division level geographic importance for Punjab wheat production using a regression methodology called correlated component regression. The important feature of this technique is that it can be best applicable for the data sets such as those used in the present study in which there exists a problem of sparse and/or multicollinearity. A PPI metric called a production performance index had been constructed to measure the production performance of each of 9 divisions as well as Punjab aggregate by excluding the Olympic average (calculated by ignoring the minimum and maximum values from the previous five values and taking the average of remaining three values) from the value of the current year. This index was calculated for wheat production during two time periods in which the first period includes the crop years from 1987 to 2003 and the second period covers 2004 to 2020 crop years.

A correlated component regression model was estimated for each period with Punjab aggregate PPI as the explained variable and nine divisions PPI as the predictor variables. The percent shares of the absolute value of the standardized regression coefficients from the estimated regression models were used to determine each division of Punjab's rank and relative importance. To determine the concentration in each period, an HHI was found using the percent shares of the standardized regression coefficients for each period under analysis and compared across the periods. This study's major hypothesis was whether the technological change along with high-yield varieties, use of fertilizers, and pesticides had shown a greater level of geographic concentration of production importance in the first period (1987-2003) or the second period (2004-2020). The estimated value of HHI checked this.

Our results indicate that the level of geographic concentration, based on the absolute shares of standardized regression coefficients, has improved moderately, as shown by the 176-point increase in the HHI from 2004 to 2020. From 1987 to 2003, the top three divisions (Gujranwala, Bahawalpur, and Multan) held 43.82 percent share, with Gujranwala holding 15.54 percent, Bahawalpur holding 14.71 percent, and Multan holding 13.57 percent share. During the 2004 to 2020 period, the top three divisions (Gujranwala, Multan, and Faisalabad) held 50.20 percent share, with Gujranwala holding 18.57 percent, Multan holding 16.13 percent, and Faisalabad holding 15.50 percent share. Faisalabad Division has improved its ranking by +5 places, whereas D. G. Khan, Rawalpindi, and Sahiwal have improved their ranking by +2 places between the two study periods under analysis.

However, the Bahawalpur, Lahore, and Sargodha divisions have decreased their ranking by -4 places during the two study periods. The Faisalabad division enjoyed the largest share improvement +6.98 percent, followed by D. G. Khan +3.89 percent and Gujranwala +3.03 percent during two time periods. Lahore and Sargodha have the largest decrease in the share of -6.91 percent and -5.87 percent, respectively, during two time periods. Gujranwala division has the highest share in the Punjab aggregate during both the study periods. Diagnostic tests indicate that the model has not shown any problem of non-normality, autocorrelation, heteroscedasticity, and instability of the parameters. The cumulative sum of square residuals (CUSUMSQ) during both periods indicates no structural instability in the estimated parameters of the model.

From the above results, the following conclusions can be drawn.

- **i.** It is very interesting to note that the rankings determined by the estimated correlated component regressions do not imitate each division's simple average production shares. The rankings determined by the correlated components regressions and simple average production share may differ because the correlated components regressions measure both the direct (prime variable) and indirect (suppressor variable) effects of each division's PPP metric.
- **ii.** Correlated component regression results from 1987 to 2003 provide a smaller value of the standardized coefficient HHI. This demonstrates that the technological change along with high-

yield varieties, use of fertilizers, and pesticides have shown a greater geographic dispersion of the relative importance of each division upon the Punjab aggregate during the time span of 1987 to 2003. The opposite was true for the timeframe of 2004 to 2020.

iii. A third notable change in the regression results during the time period of 2004 to 2020 was that the number of indirect effects components had increased from four to seven. The share of standardized coefficient (S1) called direct effect has decreased from 78.01 percent in the first period (1987-2003) to 38.17 percent in the second period (2004-2020).

Acknowledgment

The authors acknowledge the comments made by the reviewers and members of the editorial board on the earlier version of this manuscript.

Funding Source:

The author(s) received no specific funding for this work.

Conflict of Interests:

The authors have declared that no competing interests exist.

REFERENCES

Ali, I., Khan, I., Ali, H., Baz, K., Zhang, Q., Khan, A., & Huo, X. (2020). Does Cereal crops asymmetrically affect Agriculture gross domestic product in Pakistan? Using NARDL model approach. *Ciência Rural*, *50*.

Bullock, D. W. (2021). The Influence of State-Level Production Outcomes upon US National Corn and Soybean Production: A Novel Application of Correlated Component Regression. *Journal of Agricultural and Applied Economics*, 53(1), 55-74.

Chandio, A. A., Jiang, Y., & Rehman, A. (2019). Using the ARDL-ECM approach to investigate the nexus between support price and wheat production: An empirical evidence from Pakistan. *Journal of Asian Business and Economic Studies*, 26, 139-152.

Hussain, T., Javaid, T., Parr, J., Jilani, G., & Haq, M. (1999). Rice and wheat production in Pakistan with Effective Microorganisms. *American Journal of Alternative Agriculture*, 14(1), 30-36.

Janjua, P. Z., Samad, G., & Khan, N. (2014). Climate Change and Wheat Production in Pakistan: An Autoregressive Distributed Lag Approach, NJAS-Wageningen Journal of Life Sciences, 68, 13-19.

Magidson, J. (2013). Correlated component regression: Rethinking regression in the presence of near collinearity. In New perspectives in partial least squares and related methods Springer, New York, 65-78.

Pandey, S., & Elliott, W. (2010). Suppressor variables in social work research: Ways to identify in multiple regression models. *Journal of the Society for Social Work and Research*, *1*(1), 28-40.

Qureshi, A. & Iglesias, A. (1994). Implication of Global Climate Change for Agriculture in Pakistan: Impacts on Simulated Wheat Production, Climate Institute Washington DC, 1-14.

Rauf, A., Liu, X., Sarfraz, M., Shahzad, K., & Amin, W. (2017). Economic stance of wheat crop yield in Pakistan: application of ARDL bound testing model. *J. Glob. Innov. Agric. Soc. Sci*, *5*, 175-180.

Raza, S. A., Ali, Y., & Mehboob, F. (2012). Role of agriculture in the economic growth of Pakistan. *International Research Journal of Finance and Economic*, 83, 181-185.

Rehman, A., & Jingdong, L. (2017). An econometric analysis of major Chinese food crops: An empirical study. *Cogent Economics & Finance*, *5*(1), 1323-1372.

Rehman, A., Jingdong, L., Shahzad, B., Chandio, A. A., Hussain, I., Nabi, G., & Iqbal, M. S. (2015). Economic perspectives of major field crops of Pakistan: An empirical study. *Pacific Science Review B: Humanities and Social Sciences*, *1*(3), 145-158.

Siddiqui, R., Samad, G., Nasir, M., & Jalil, H. H. (2012). The impact of climate change on major agricultural crops: evidence from Punjab, Pakistan. *The Pakistan Development Review*, 261-274.