

Modeling the contribution of district-level cotton production to aggregate cotton production in Punjab (Pakistan): An empirical evidence using correlated component regression approach

 Muhammad Naveed¹

 Muhammad Furqan Maqsood²

 Ali Raza Cheema³

How to cite this article:

Naveed, M., Maqsood, M. F., & Cheema, A. R. (2024). Modeling the contribution of district-level cotton production to aggregate cotton production in Punjab (Pakistan): An empirical evidence using correlated component regression approach. *Journal of Excellence in Social Sciences*, 3(3), 147–164.

Received: 5 April 2024 / Accepted: 7 June 2024 / Published online: 5 August 2024

Abstract

This study aims to determine the relative importance of major cotton-producing districts with reference to their aggregate cotton production in Punjab using the time series data from 1982-2021. The empirical analysis is conducted with a correlated component regression approach, which is comparatively more suitable to multicollinear and high-dimensional data sets like those examined in this study. The standardized regression coefficients' absolute value determines each district's rank and relative importance. The degree of geographic concentration of the district-level impact is examined by the Herfindahl-Hirschman Index (HHI). The empirical analysis is analyzed for two-time windows. The first-time window includes crop years from 1982 to 2001, and the second window covers crop years from 2002 to 2021. The Findings of the study indicate that all the included districts had a positive and statistically significant impact on Punjab provincial aggregate cotton production in both the study periods, with the exception of districts Sargodha and Mianwali, which had a negative impact on Punjab-aggregate-cotton-production from 1982 to 2001. Furthermore, the estimated results of the correlated component regression show a decline in the degree of geographic concentration during 2002-2021, as indicated by a 36-point decrease in the value of standardized coefficient HHI. This indicates that technological advancements and policy factors have contributed to greater geographic variation in the relative importance of each district during 2002-2021 compared to 1982-2001. The cumulative sum of residuals (CUSUM) and cumulative sum of square residuals (CUSUMSQ) for both time periods show that the estimated parameters

¹Govt. Allama Iqbal Graduate College, Paris Road Sialkot, Pakistan

Corresponding Author: m_naveed_skt@yahoo.com

²Govt. Allama Iqbal Graduate College, Paris Road Sialkot, Pakistan

³Agriculture Department (Crop Reporting Service) Punjab, Sialkot-Pakistan



of the correlated component models are structurally stable.

Keywords: Correlated Component Regression, Cotton Production, Relative Importance in Prediction, Punjab, Districts, Geographic Concentration.

1 Introduction

Agriculture is the backbone of a country's economy, particularly in developing economies, and cotton is the crucial crop among agricultural crops (Ahmad et al. [2014](#)). Cotton is white gold for some countries because it produces huge revenue for these countries (Ali et al. [2020](#)). In many countries, cotton is a key contributor to revenue creation and economic development. It is a lifeline for millions of small-scale farmers, laborers, and their families. It provides employment and income and attracts export revenues for some of the low-income countries in the world. Moreover, it helps some poor countries in the world to pay their import bills (Food and Agriculture Organization of the United Nations [FAO], [2021](#)).

Agriculture holds immense significance in the Pakistani economy, playing a significant role across multiple dimensions (Qazi et al., [2021](#)). It is indispensable in the country for economic growth, food availability, employment opportunities, and poverty alleviation. It accounts for approximately 19.2 percent to the national Gross Domestic Product (GDP) and serves as the primary source of employment for around 38.5 percent of the workforce. While its share in the GDP has experienced a decline in recent decades, the agricultural sector remains crucial due to its potential for leveraging modern farming technologies to enhance productivity (Sarfraz et al., [2023](#)).

The success of the agriculture sector is not only pivotal for overall economic growth but also for poverty reduction, as it has the capacity to generate additional job opportunities for the country's labor force. In developing nations, agriculture is widely recognized as an engine of economic growth (Tiffin and Irz [2006](#)). Therefore, prioritizing the development and advancement of the agriculture sector is essential for Pakistan. By embracing innovative farming techniques and technologies, the agricultural industry can increase its economic impact and alleviate poverty by creating more employment opportunities.

Agriculture is pivotal in Punjab's economic progress and has significantly contributed to the domestic agricultural economy (Qazi et al., [2021](#)). The advancement of the agriculture sector leads to higher earnings for low-income households, making it the most advantageous sector for pro-poor growth. According to the Punjab Economic Report, issued by the Asian Development Bank ([2005](#)), the agriculture sector employs about 40 percent of the total labor force in the province. Furthermore, the agriculture sector in Punjab contributes 27 percent to Pakistan's Gross Domestic Product (GDP), whereas Pakistan's agriculture sector constitutes only 23.6 percent of the GDP (Rasheed & Nosheen, [2024](#)).

In real terms, the GDP of the province of Punjab has gone up to about 4.5 percent a year, which was 0.8 percent larger than the annual growth recorded in the rest of the country between the period of 1991-92 and 2002-03. The contribution of agriculture to the overall economy of Punjab was 23 percent in 2012-13, which declined to 20.8 percent in 2015-16 and further decreased to 20.20 percent in 2017-18 (Planning & Development Board, [2023](#)). These statistics indicate that Punjab's economy, with its diverse sectors and agricultural strength, substantially impacts Pakistan's overall economic growth and development.

In Pakistan, the cultivation of various important crops, including wheat, cotton, rice, maize, and sugarcane, plays a predominant role in the agriculture sector. These crops collectively contribute 23.60 percent to the agricultural GDP and 5.45 percent to the overall GDP of the country (Ali et al., [2020](#)). Cotton production is paramount to Pakistan's economy (Baig et al. [2023](#)). Cotton is a

vital cash crop in Pakistan, and of our total foreign exchange earnings, cotton products export account for 55 percent. Almost 26 percent of growers grow cotton primarily across Pakistan, and 15 percent of land is devoted to this crop. Cotton crop is concentrated region wise. Almost 65 percent of cotton is cultivated in Punjab under dry climate conditions, and the rest is cultivated in Sindh under humid conditions. However, Khyber Pakhtunkhwa and Baluchistan have very small areas under cultivation. Cotton production contributes around 0.8 percent to the GDP and 4.5 percent of the value added to agricultural GDP. (Rana et al. [2020](#)).

Regardless of the significance of cotton, it serves as an important raw material for the national textile industry, and its productivity in Pakistan has been underwhelming. Pakistan ranked 4th in terms of cotton area under cultivation (Shuli et al. [2018](#)) but, unfortunately, ranked 39th in the case of its productivity per hectare (Rana et al., [2020](#)). On the one end, productivity is very low; on the other, the area under cultivation is declining, resulting in low production. Cotton crop in the country was cultivated on 2079 thousand hectares, reflecting a contraction of 17.4 percent compared to last year's sown area of 2517 thousand hectares. Production declined by 22.8 percent to 7064 thousand bales against production of 9148 thousand bales last year (Government of Pakistan Finance Division [GOP], [2021](#)).

In Pakistan's economy, all the provinces have a substantial share of GDP; however, Punjab is the major contributor. The growth rate of Punjab, particularly, determines the overall growth rate of the economy (Planning & Development Board, [2023](#)). Its share in national economic accounts was 54.2 percent in 2017-18 but 60 percent in overall agricultural production. The economy of Punjab is not much different from our national economy because it covers 62.4 percent of services, 17.5 percent of industry, and 20.2 percent of agriculture. In agriculture, the Punjab province is playing a principal role. It covers almost 68 percent of our national grain needs. Rice and cotton are central crops of Punjab. These two cash crops have contributed noticeably to the state treasury.

During 2020-21, 5044 thousand bales of cotton were produced in Punjab, which was 71.40 percent of Pakistan's total cotton production (Khurshid et al., 2021). At the same time, Bahawalnagar contributed the largest share of the aggregate cotton production in Punjab, 19.70 percent, followed by Bahawalpur, 17.07 percent, and R. Y. Khan, 12.24 percent. Table 1 indicates the average relative production shares in major cotton-producing districts of Punjab during two time periods. During the time window 1982-2001, the highest average relative production share was contributed by district R. Y. Khan at 13.34 percent followed by Vehari (11.96 percent), Bahawalpur (11.06 percent), Lodhran (9.20 percent), Khanewal (8.33 percent) and Multan (8.10 percent) while during the same time window, the lowest average relative production share was contributed by Mianwali, Bhakkar, Kasur, Sargodha and Layyah at the rate of 0.23 percent, 0.32 percent, 0.49 percent, 0.54 percent and 0.83 percent respectively. However, during the time window 2002-2021, the largest average relative production share in aggregate cotton production of Punjab was contributed by Bahawalpur 13.15 percent, followed by R. Y. Khan (11.71 percent), Bahawalnagar (10.48 percent), Vehari (9.23 percent) and Lodhran (9.01 percent) while during the same time period the lowest average relative production share was contributed by Sargodha at the rate of 0.15 percent.

Table 1: District-Wise Averages Relative Production Shares (%) with Reference to Aggregate Cotton Production in Punjab during Two Time Periods

District	1982-2001	2002-2021
R. Y. Khan	13.34	11.71
Vehari	11.96	9.23
Bahawalpur	11.06	13.15
Lodhran	9.2	9.01
Khanewal	8.33	8.58
Multan	8.1	7.59
Bahawalnagar	7.25	10.48
Sahiwal	5.77	2.91
Muzaffargarh	5.45	6.55
Rajanpur	4.4	6.83
Jhang	2.74	1.38
D.G.khan	2.74	4.33
Okara	2.39	0.75
Pakpattan	1.5	1.6
Faisalabad	1.36	1.12
Layyah	0.83	1.48
Sargodha	0.54	0.15
Kasur	0.49	0.25
Bhakkar	0.32	0.54
Mianwali	0.23	0.89

Based on the aforementioned discussion, it is clear that Punjab plays an important role as a major cotton producer in Pakistan. Consequently, any significant change in cotton production within the districts of Punjab can have substantial implications for cotton production in Pakistan. Therefore, understanding the behavior and relative importance of major cotton-producing districts with reference to their aggregate cotton production in Punjab is of utmost importance. A careful review of the academic literature to date has turned up no evidence of any studies that have examined the relative importance of major cotton-producing districts with respect to aggregate cotton production in Punjab. To address this research gap, this study examines the relative importance of major cotton-producing districts in relation to aggregate cotton production in Punjab.

1.1 Research Objectives

This study aims

- To determine the relative importance of major cotton-producing districts regarding their aggregate cotton production in Punjab using the correlated component regression technique.
- To measure and compare the degree of geographic concentration between the two time periods (1982-2001 and 2002-2021) across the districts using the Herfindahl-Hirschman Index (HHI). This objective helps us to determine the influence of technology and policy factors on the ranking of major cotton-producing districts and the geographical concentration of their relative importance.
- To measure and compare the positions of the districts determined by the average relative production shares with rankings and shares calculated using correlated component regression.

1.2 Research Questions

Based on the research objectives, the study aims to address the following research questions:

- How much does each district among the major cotton-producing districts contribute to aggregate cotton production in Punjab, and with what rank?
- During which period does each district's relative importance in aggregate cotton production in Punjab indicate greater geographic variation?

- Do the rankings of cotton-producing districts calculated using average relative production share match those calculated using correlated component regression?

This study employs correlated component regression (CCR) methodology, developed by Magidson (2011), as used by Bullock (2021), Naveed and Hina (2023) in agricultural economics. Data for this study was collected from 1982 to 2021 from the Directorate of Agriculture Crop Reporting Service, Punjab Lahore (crs.agripunjab.gov.pk), and crops, area, and production (by districts) volume I issued by the Federal Bureau of Statistics. To determine the relative importance of each district on aggregate production, regression analyses are conducted using the aggregate cotton production in Punjab as the dependent variable and each district's cotton production as the independent variable. Standardized regression coefficients are calculated from these regressions, and these coefficients are used to rank the districts based on their importance in affecting aggregate cotton production.

Due to issues with datasets such as their multicollinearity and sparsity, simple ordinary least squares (OLS) regression methods were unsuitable as they resulted in large standard errors with unreliable and insignificant coefficient estimates. Therefore, correlated component regression (CCR) is 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, this study aimed to provide a comprehensive understanding of the relative importance of each district's cotton production with reference to aggregate cotton production in Punjab. Using CCR allowed for more robust and accurate estimations, overcoming the challenges of sparse and multicollinear data.

This study's second significant goal is to determine the influence of technology and policy factors on the ranking of major cotton-producing districts and the geographical concentration of their relative importance. To achieve this objective, the time-series dataset is divided into two distinct periods: the first period contains 1982 to 2001 and the second period covers 2002 to 2021. The Herfindahl-Hirschman Index (HHI) is used in the study to determine the geographic concentration of importance. The HHI is 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 districts. By computing the HHI value for both time periods in relation to cotton production, it became possible to assess the impact of technology and policy factors on the concentration of cotton production. The results from these calculations shed light on the degree to which technology and policy influences have contributed to the geographic concentration of importance in the context of cotton production.

2 Review of Literature

A careful review of the academic literature to date has produced no evidence of any studies in Pakistan that have analyzed the relative importance of district-level cotton production with reference to their aggregate cotton production in Punjab. A literature review of some previous studies is given below. Najib et al. (2022) empirically examined the impact of cotton production on economic growth in Benin using time series data covering the period from 1965 to 2021. Vector error correction modeling was applied to obtain empirical results. The study's findings indicated a strong positive nexus between cotton exports and economic growth.

Shabbir and Yaqoob (2019) analyzed the impact of technology on the total factor productivity of cotton crops in Pakistan and India using the time series data from 1954 to 2017. Their empirical results for Pakistan indicated that fertilizer consumption, the area under cotton crop, canal irrigation, seed, expenditure on education, and the number of tractors have a positive and significant association with total factor productivity of cotton crop while labor force in agriculture, tube wells irrigation, and electricity consumption have an insignificant negative association with

total factor productivity of cotton crop (Arshad et al., [2022](#)). Empirical findings for India depicted that electricity consumption, expenditure on education, the area under cotton crop, and fertilizer consumption have a negative and significant association with the total factor productivity of cotton crop (Sinha, [2023](#)). At the same time, labor force, number of tractors, tube wells irrigation, and canal irrigation have shown a positive but insignificant association with total factor productivity of the cotton crop. Chaudhry et al. ([2009](#)) empirically examined the comparative advantage of cotton, rice, and sugarcane crops for the Bahawalpur and Multan regions using primary data from 100 farmers from each region. Their empirical results showed that cotton crops were more efficiently produced in the Bahawalpur and Multan regions, while these regions were less efficient in producing rice and sugarcane crops. Moreover, they indicated that cotton crops can compete in domestic markets as well as in foreign markets.

Rehman et al. ([2019](#)) empirically examined the association between agricultural gross domestic product, cotton production, fertilizer consumption, and area under cotton crops using the annual time series data from 1970-2015 in Pakistan. They found that there exists a long-run association between agricultural gross domestic product, cotton production, fertilizer consumption, and area of the cotton crop. Empirical findings revealed that the production of cotton crops and consumption of fertilizer is positively associated with the agricultural gross domestic product of Pakistan, while the area of the cotton crops is negatively associated with the agricultural gross domestic product of Pakistan.

Ahmad and Afzal ([2018](#)) investigated the cost and profit analysis of cotton crops using the primary data from 240 farmers of cotton crops for the years 2012-13 in the Bahawalpur district of Punjab. Their findings revealed that the price and quantity of the crop produced have a positive and significant impact on cotton profit, while the cost of production has a significantly negative impact on cotton profit (Arshad et al., [2022](#)). Moreover, they found that seed, cropped area, fertilizer, land preparation, labor, irrigation, and pesticides are directly associated with cotton production.

Bullock ([2021](#)) has investigated the state-level impact of soybean and corn production on aggregate production in the U.S. The study analyzed the time series data from 1970 to 2017 using a correlated component regression approach. The data was divided into two time periods: a period of Pre-Genetic Modification covering the years 1975 to 1995 and a period of Pro-Genetic Modification indicating the years 1996 to 2017. Absolute values of standardized regression coefficients were applied to determine each state's rank and relative importance. The degree of geographic concentration was calculated using the Herfindahl-Hirschman Index. Empirical results of the study indicated that U.S. soybean production was found to be geographically more concentrated, while U.S. corn production was geographically less concentrated with respect to state-level importance.

Naveed and Hina ([2023](#)) have empirically investigated the division-level influence of wheat production with reference to their aggregate production in Punjab using time series data from 1987 to 2020. A correlated component regression approach was applied to find out the results. The rank and relative importance of each division was determined by the absolute value of the standardized regression coefficients. The degree of geographic concentration of the division-level impact was examined by the Herfindahl-Hirschman Index (HHI). The empirical analysis was analyzed for two-time windows: the first-time frame includes crop years from 1987 to 2003, and the second span covers years from 2004 to 2020. Empirical findings during both the study periods indicated that all divisions had a significant positive impact on aggregate wheat production in Punjab, with the exception of divisions Faisalabad and Sahiwal, which had a positive but insignificant impact on Punjab-aggregate-wheat-production from 2004 to 2020. Furthermore, the estimated results of the correlated component regression showed a rise in the degree of geographic concentration during 2004-2020, as indicated by a 176-point increase in the value of standardized coefficient HHI. This indicates that geographical importance was more dispersed between the period 1987

and 2003 than between 2004 and 2020.

Siddiqui et al. (2012) empirically examined climate change's impact on cotton, rice, wheat, and sugarcane crops using the annual time series data from 1980-2008 in Punjab. Their estimated results indicated that change in climate has a positive impact on the productivity of wheat, while the impact of change in climate was negative for cotton, rice, and sugarcane crops. Rehman et al. (2015) empirically investigated the association of agricultural gross domestic product and important crops, namely cotton, rice, wheat, maize, and sugarcane in Pakistan, by taking the annual time series data from 1950-2015. Empirical findings of the study indicated that the output of cotton, rice, wheat, and maize has a positive association with the agricultural gross domestic product, while the output of sugarcane crops has shown a negative relation with the agricultural gross domestic product of Pakistan.

Abbas (2020) investigated the association between cotton production, area of cotton cultivation, fertilizer consumption, and change in climate using time series data from 1980-2018 in Pakistan. The study applied the autoregressive distributed lag approach to assessing the long-term association between cotton production and independent variables. Estimated results of the study indicated that areas under cotton production and fertilizer consumption have a positive and significant impact on production in the long run and short run as well, while the change in climate variable has a positive but insignificant impact on cotton production both in the long run and short run.

Magsi (2012) estimated the average growth rates of productivity, production, and cotton crop area. The study also analyzed the impact of cotton support price on its production in Pakistan using annual time series data from 1979-80 to 2008-09. The study's findings indicated that productivity, production, and area of cotton crops have grown at a rate of 3.2 percent, 5.2 percent and 1.9 percent, respectively, over the study period. The study also showed that the support price of cotton positively impacts cotton production.

Rehman et al. (2011) estimated the growth pattern of production, area, and yield of important crops in Pakistan using secondary data covering the years from 1972 to 2009. The data was divided into two time periods: the first span includes the years from 1972 to 1988, and the second time window covers 1989 to 2009. Estimated results indicated that the growth rates of production and area were better in terms of cotton, wheat, and sugarcane in the first time period, while rice and sugarcane performance was better in the second time frame. Moreover, the empirical results of the decomposition model revealed that the yield effect was the paramount source of growth in terms of wheat and cotton in both the study periods. However, the area effect regarding sugarcane was the significant source of production growth in both study periods.

Yang et al. (2022) empirically evaluated the cotton production competitiveness in Xinjiang province using secondary time series data covering the period from 2005 to 2018 in China. The comparative advantage of cotton production was evaluated using the aggregate, efficiency, and scale advantage indexes. Correlation matrix and ridge regression were used to determine the factors affecting cotton productivity. Estimated results of the aggregate, efficiency, and scale advantage indexes showed that Xinjiang province had a vast comparative advantage in cotton production. Empirical findings of ridge regression indicated that agricultural machinery power, fiscal expenditure on agricultural support, fertilizer use, and total cotton production had positive and significant impacts on agricultural output. Meanwhile, the proportion of affected areas by insects and diseases, as well as the average cotton yield and area, had a negative impact on agricultural output.

Raza and Ahmed (2015) investigated the impact of climate change on cotton productivity in Pakistan using temperature, precipitation, fertilizer, area, and cotton yield. The time series data

from 1981 to 2010 was used in the study. Production function was applied to investigate the relationship between cotton productivity and climate change. Empirical results of Fixed Effect Model indicated that precipitation and temperature significantly impacted cotton productivity. Moreover, fertilizer, fertilizer nutrients off take per acre of cotton, area under cotton, and technology all positively and significantly impacted cotton yield.

Rashid et al. (2020) empirically evaluated the impact of climate change on cotton production in Pakistan. They used an annual time series dataset from 1981 to 2015 for their investigation. The ARDL technique's empirical results indicated that maximum temperature and rainfall both had a significant positive impact on cotton productivity, while minimum temperature had a significant negative influence. Moreover, areas under cotton, fertilizer, and technology were shown to have positive effects on cotton production.

A number of researchers used various approaches to analyze the impact of cotton production on economic growth or agricultural gross domestic product (Rehman et al., 2015; Rehman et al., 2019; Najib et al., 2022). Some other studies (Abbas 2020; Rashid et al. 2020; Raza & Ahmed, 2015; Siddiqui et al., 2012) investigated the relationship between climate change and cotton productivity. However, less attention has been given to analyzing the relative significance of cotton-producing districts with reference to their aggregate cotton production in Punjab.

3 Data and Methodology

In this study, district-wise, as well as aggregate data of cotton production in Punjab is used for 20 districts, including Vehari, Bahawalpur, Khanewal, Muzaffargarh, R. Y. Khan, Lodhran, Bahawalnagar, Multan, Sahiwal, Layyah, Okara, Faisalabad, Rajanpur, Jhang, D. G Khan, Kasur, Pakpattan, Mianwali, Sargodha and Bhakkar. The dataset on cotton production came from the Directorate of Agriculture Crop Reporting Service, Punjab Lahore (crs.agripunjab.gov.pk), and the data report compiled by Khan et al. (2010) with title crops, area, and production (by districts) volume I issued by the federal bureau of statistics. The cotton production is measured in tones. The dataset in this study has been divided into two time periods: the first-time window contains 1982 to 2001 years, and the second time window covers 2002 to 2021 years. For each time period, the data has been analyzed by using the correlated component regression (CCR) approach. Herfindahl-Hirschman Index (HHI) measures the degree of geographic concentration. A comprehensive overview of all these measures is given below

3.1 Correlated Component Regression (CCR)

In this study, we have applied the correlated component regression (CCR) approach, which Magidson (2011), used to analyze the district-level influence of cotton production on aggregate cotton production in Punjab. A particular problem that has to be considered when using traditional regression analytic approaches is the existence of suppression effects or multicollinearity (Lynn 2003). Generally, multicollinearity or suppression effects occur when there exists a moderate or high degree of correlation between two or more two explanatory variables, which have no direct effect on the dependent variable. These multicollinearity issues or data sparsity lead to large variances and standard errors of estimated regression coefficients, rendering them unstable and statistically insignificant (Pandey & Elliot, 2010).

Therefore, treatment of multicollinearity or suppression effects is very important. Magidson (2013) indicated that for the multicollinear or suppression data sets, the CCR approach gives more reliable and stable estimates of the regression coefficients. In agriculture economics, this methodology was first used by Bullock (2021) for analyzing the state-level impact of soybean and corn production on aggregate production in the U.S. This technique was also applied by Naveed and Hina (2023) for empirically investigating the division-level influence of wheat production with reference to their aggregate production in Punjab, Pakistan. This study also uses CCR regression to assess the contribution of district-level cotton production to aggregate cotton production in Punjab, Pakistan.

The general procedure of the CCR approach is described as follows:

In the first step, we estimate the following equation for each predictor from ordinary least squares (OLS). This proceeds as follows:

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

Where Y is the outcome variable, and X_g shows predictor variables with $g = 1, 2, 3, \dots, P$; $\hat{\gamma}_g^{(1)}$ and $\hat{\lambda}_g^{(1)}$ are respectively the intercept and slope coefficients for a particular predictor variable g .

The component variable S_1 is defined as the weighted average of all 1-predictor effects, while the weights are slope coefficients estimated from equation (1), that is

$$S_1 = \frac{1}{P} \sum_{g=1}^P \hat{\lambda}_g^{(1)} X_g \quad (2)$$

Predictions for Y in the 1-component CCR model are obtained through OLS regression of Y on S_1

$$\hat{Y} = \hat{\alpha}^{(1)} + \hat{\beta}_1^{(1)} S_1 \quad (3)$$

As the component variable, S_1 measures the direct effect of each independent variable upon the dependent variable without considering the suppressor effects. Therefore, it is known as the direct effect component variable. In the same way, the second component variable S_2 is calculated by first predicting the following regression equation for each explanatory variable from simple OLS:

$$\hat{Y} = \hat{\gamma}_g^{(2)} + \hat{\lambda}_{1,g}^{(2)} S_1 + \hat{\lambda}_g^{(2)} X_g \quad (4)$$

The second correlated component variable S_2 is defined as the weighted average of all 2-predictor effects:

$$S_2 = \frac{1}{P} \sum_{g=1}^P \hat{\lambda}_g^{(2)} X_g \quad (5)$$

Predictions for Y in the 2-component CCR model are obtained from simple OLS regression of Y on S_1 and S_2 , that is

$$\hat{Y} = \hat{\alpha}^{(2)} + \hat{\beta}_1^{(2)} S_1 + \hat{\beta}_2^{(2)} S_2 \quad (6)$$

Component variable S_2 and subsequently derived component variables are correlated with S_1 , capturing the effects of suppressor variables. These suppressor variables improve prediction in the component variable model by removing irrelevant variation from one or more explanatory variables directly affecting the outcome variable. This process for the derivations of component variables can continue until the optimal number of components is achieved.

In general, for any K (where K is less than P)-component variables, we first estimate the regression equation (7) for each explanatory variable using OLS, that is

$$\hat{Y} = \hat{\gamma}_g^{(K)} + \hat{\lambda}_{1,g}^{(K)} S_1 + \hat{\lambda}_{2,g}^{(K)} S_2 + \dots + \hat{\lambda}_{K-1,g}^{(K)} S_{K-1} + \hat{\lambda}_g^{(K)} X_g \quad (7)$$

After estimating equation (7), the final correlated component variable S_k is defined as follows:

$$S_k = \frac{1}{P} \sum_{g=1}^P \hat{\lambda}_g^{(k)} X_g \quad (8)$$

Predictions for Y in the k -component CCR model are obtained using simple OLS regression of Y on $S_1, S_2, \dots, S_k$.

$$\hat{Y} = \hat{\alpha}^{(K)} + \hat{\beta}_1^{(K)} S_1 + \hat{\beta}_2^{(K)} S_2 + \dots + \hat{\beta}_k^{(K)} S_k \quad (9)$$

Substituting equations (2), (5), and (8) in equation (9), we get equation (10):

$$\hat{Y} = \hat{\alpha}^{(K)} + \hat{\beta}_1^{(K)} \left(\frac{1}{P} \sum_{g=1}^P \hat{\lambda}_g^{(1)} X_g \right) + \hat{\beta}_2^{(K)} \left(\frac{1}{P} \sum_{g=1}^P \hat{\lambda}_g^{(2)} X_g \right) + \dots + \hat{\beta}_k^{(K)} \left(\frac{1}{P} \sum_{g=1}^P \hat{\lambda}_g^{(k)} X_g \right) \quad (10)$$

From equation (10) we can get equation (11):

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

$$\hat{Y} = \hat{\alpha}^{(K)} + \sum_{g=1}^P \hat{\beta}_g X_g \quad (12)$$

Where

$$\hat{\beta}_g = \frac{1}{P} \sum_{k=1}^K \hat{\beta}_k^{(K)} \hat{\lambda}_g^{(k)} \quad (13)$$

The estimated coefficients $\hat{\beta}_g$ for each explanatory variable X_g in equation (12) is simply the weighted average of all the loadings taking regression coefficients $\hat{\beta}_1^K, \hat{\beta}_2^K, \dots, \hat{\beta}_k^K$ as weights. Equation (13) can be used to calculate the values of non-standardized regression coefficients while the standard errors are calculated from the following formula:

$$SE(\hat{\beta}_g) = \frac{1}{P} \sqrt{\sum_{k=1}^K (SE(\hat{\beta}_k^K))^2 (\hat{\lambda}_g^k)^2} \quad (14)$$

Where $\hat{\lambda}_g^k$ and $SE(\hat{\beta}_k^K)$ are, respectively, the loadings and the standard errors of the estimated coefficients of the final correlated component model. Districts are ranked based on the absolute values of standardized regression coefficients, which are obtained as follows:

$$\hat{\beta}_g^* = \left(\frac{\hat{\sigma}_g}{\hat{\sigma}_y} \right) \times \hat{\beta}_g \quad (15)$$

Where

$\hat{\beta}_g^*$ = standardized regression coefficient for each explanatory variable with $g = 1, 2, 3, \dots, P$

$\hat{\beta}_g$ = regression coefficient for each explanatory variable, with $g = 1, 2, 3, \dots, P$. These are obtained from equation (13)

$\hat{\sigma}_g$ = standard deviation of each of the explanatory variables, with $g = 1, 2, 3, \dots, P$ and

$\hat{\sigma}_y$ = standard deviation of the dependent variable y

Standardized regression coefficients reveal which independent variables impact the dependent variable more. Generally, the standardized regression coefficient measures each explanatory variable's marginal impact (in standard deviations) on the dependent variable. Absolute values of the standardized regression coefficients of each district and the percentage share of the sum of absolute standardized coefficients are used to determine each district's rank and relative importance. Districts with higher rankings indicate that one standard deviation change in their production has a larger impact (in standard deviations) on aggregate production in Punjab.

We have divided the whole data set (1982-2021) into two periods: the first window includes crop years from 1982 to 2001, and the second window covers crop years from 2002 to 2021. Since each time period contains as many observations as the number of explanatory variables, therefore the traditional approach, such as OLS, is not applicable to these data sets. Moreover, the OLS methodology could not be used to calculate the regression coefficients' reliable estimates due to the suppression effects or multicollinearity problem. Therefore, a newly developed regression technique, that provides more reliable and stable estimates of the regression coefficients in the case of a multicollinear or suppression data set, called correlated component regression (CCR), has been followed in this study.

3.2 Herfindahl-Hirschman Index (HHI)

In measuring the degree of geographic concentration of the district-level impacts, a Herfindahl-Hirschman Index (HHI), which was also used by Bullock (2021) and Naveed and Hina (2023) in agriculture economics, has been calculated from the following formula:

$$HHI = \sum_{i=1}^n (s_i \cdot 100)^2 \quad (16)$$

Where n shows the number of the districts included in the study, and s_i represents the share for the district as a percent of the absolute sum of standardized regression coefficients. The HHI is a measurement commonly used to understand the level of market concentration in applied industrial organization research and antitrust policy. However, this index can also be used as a general measure of concentration in many other applications. Its value lies between 0 and 10,000. HHI has a maximum value of 10,000 when a single market participant holds 100% of the market share, indicating a monopoly situation. If there are many market participants, with each holding a market share of almost 0%, the HHI has a value close to zero, indicating its minimum, and this form of market is often classified as perfect competition.

This study examines the degree of change in geographic concentration across two time periods through a change in HHI value. The second objective of this study is to determine whether the first or second time period exhibits greater geographic variation in the relative importance of each district's contribution to aggregate cotton production in Punjab. This objective is examined through the analysis of a change in HHI value. A lower HHI value in a period indicates shares are less concentrated across the districts in that period, implying greater geographical variation in that period compared to other period and vice versa.

4 Results and Discussion

The CCR-Linear regression has been applied to cotton production data. When the data set is divided into two periods, 1982-2001 and 2002-2021, the number of explanatory variables approaches the sample size, implying high-dimensional data. CCR performs very well for high-dimension data with components 2, 3, or 4 (Magidson [2010](#)). Following Magidson ([2010](#)), this study uses 4 correlated component variables in the regression model and retains all the explanatory variables in both time periods.

Table 2: Correlated component non-standardized and standardized regression coefficients of cotton production for a time window (1982-2001)

Component	Coefficient	Standard Error	T-Statistic	Standardized Coefficient	Share (%)
S1	0.0834*	0.0017	49.1256	0.9974	70.92
S2	0.1245*	0.0055	22.6394	0.1477	10.50
S3	0.2698*	0.0228	11.8372	0.1681	11.95
S4	0.5191*	0.0761	6.8183	0.0933	6.63

*, **, and *** indicate that the coefficients significantly differ from zero at the 1%, 5%, and 10% significance levels, respectively.

Estimated values of the non-standardized and standardized coefficients, along with their statistical significance and shares as a percentage sum of standardized coefficients for the time window 1982-2001, are reported in Table 2. All four coefficients of component variables are significantly different from zero at a 1% significance level. The coefficient of the first correlated component variable, S1, measures the direct effect and has a 70.92 percent share of the sum of standardized coefficients. The remaining 29.08 percent share measures the indirect effect, distributed among second, third, and fourth component variables, with S2 and S3 having 10.50 percent and 11.95 percent shares respectively, while S4 has only 6.63 percent share of the total standardized coefficients.

Table 3: Correlated component non-standardized and standardized regression coefficients of cotton production for a time window (2002-2021)

Component	Coefficient	Standard Error	T-Statistic	Standardized Coefficient	Share (%)
S1	0.1127*	0.0016	69.3003	1.1111	70.12
S2	0.1802*	0.0081	22.2505	0.2452	15.48
S3	0.2016*	0.0181	11.1263	0.1568	9.90
S4	0.3231*	0.0587	5.5093	0.0713	4.50

*, **, and *** indicate that the coefficients significantly differ from zero at the 1%, 5%, and 10% significance levels, respectively.

Estimated values of the non-standardized and standardized coefficients, along with their statistical significance and shares as a percentage sum of standardized coefficients for the time window 2002-2021, are reported in Table 3. All four coefficients of component variables are significantly different from zero at a significance level of 1%. The coefficient of the first correlated component, variable S1, measures the direct effect and has a 70.12 percent share of the sum of standardized coefficients. The remaining 29.88 percent share measures the indirect effect, distributed among second, third, and fourth component variables, with S2 and S3 having 15.48 percent and 9.90

percent shares respectively, while S4 has only 4.50 percent share of the total standardized coefficients.

Estimated values of the district-level non-standardized and standardized coefficients, along with their statistical significance and shares as a percentage of the absolute sum of standardized coefficients from 1982 to 2001, are reported in Table 4. The coefficients of all the districts have positive and significant impact on aggregate cotton production in Punjab with the exception of district Sargodha and Mianwali, which have negative impact on Punjab-aggregate-cotton-production during 1982-2001. The production shares, based on the absolute sum of standardized coefficients, indicate that District Vehari has the highest contribution (13.07 percent), followed by Bahawalpur (9.85 percent), Khanewal (9.22 percent), Muzaffargarh (8.47 percent), R. Y. Khan (7.66 percent), Lodhran (6.71 percent) and Bahawalnagar (6.29 percent). There exists not much difference among the shares of districts Multan, Sahiwal, Layyah, Okara, and Faisalabad, which range from 4.44 percent to 4.87 percent. Sargodha has the lowest standardized share, contributing only 0.01 percent. The Herfindahl-Hirschman Index (HHI), calculated from the shares of the absolute sum of standardized regression coefficients, is 726, showing a low level of concentration, as it is less than the 1500 threshold set by the U. S. Department of Justice and the Federal Trade Commission.

Table 4: District-wise non-standardized regression coefficients, standardized regression coefficients, and shares as a percentage of total absolute standardized coefficients for a time window (1982-2001)

Rank	District	Coefficient	Standard Error	T-Statistic	Standardized Coefficient	Share (%)
1	Vehari	1.3757*	0.0455	30.2279	0.1621	13.07
2	Bahawalpur	0.8829*	0.0380	23.2353	0.1222	9.85
3	Khanewal	0.9661*	0.0201	48.1483	0.1144	9.22
4	Muzaffargarh	1.4413*	0.0638	22.5822	0.1050	8.47
5	R. Y. Khan	0.7925*	0.0363	21.8583	0.0950	7.66
6	Lodhran	1.0345*	0.0559	18.4960	0.0832	6.71
7	Bahawalnagar	0.8289*	0.0453	18.2787	0.0780	6.29
8	Multan	0.7754*	0.0581	13.3453	0.0604	4.87
9	Sahiwal	0.5969*	0.0429	13.8992	0.0600	4.84
10	Layyah	4.9954*	0.1436	34.7797	0.0576	4.64
11	Okara	1.2606*	0.0560	22.5237	0.0575	4.63
12	Faisalabad	2.3773*	0.0872	27.2781	0.0551	4.44
13	Rajapur	0.6058*	0.0325	18.6251	0.0492	3.96
14	Jhang	1.2547*	0.0632	19.8667	0.0451	3.64
15	D.G.khan	0.7673*	0.0506	15.1716	0.0445	3.59
16	Kasur	3.3919*	0.2027	16.7297	0.0240	1.93
17	Pakpattan	0.3348*	0.0359	9.3372	0.0178	1.44
18	Mianwali	-1.6390*	0.4562	-3.5931	-0.0057	0.46
19	Bhakkar	0.8718***	0.4238	2.0574	0.0035	0.28
20	Sargodha	-0.0106	0.3250	-0.0325	-0.0001	0.01
Constant		95628.32**	44264.58	2.1604		
Herfindahl-Hirschman Index						726

*, **, and *** indicate that the coefficients significantly differ from zero at the 1%, 5%, and 10% significance levels, respectively.

Estimated values of the district-level non-standardized and standardized coefficients, along with their statistical significance and shares as a percentage of the absolute sum of standardized coefficients from 2002 to 2021, are reported in Table 5. All the districts have positively and significantly impacted aggregate cotton production in Punjab. Standardized regression coefficient shares indicate that District Bahawalpur has the highest contribution (9.75 percent), followed by R. Y. Khan (9.64 percent), Vehari (9.38 percent), Lodhran (9.20 percent), Khanewal (8.12 percent), Muzaffargarh (8.08 percent) and D. G. Khan (7.14 percent). There exists not much

difference among the shares of districts Layyah, Pakpattan, Jhang, Bhakkar, and Okara, as these district's shares range from 2.00 percent to 2.91 percent. During this time period, the lowest production share has been contributed by Mianwali. The Herfindahl-Hirschman Index (HHI), calculated using shares of the absolute sum of standardized regression coefficients, has a 690 value below the 1500 threshold that the U. S. Department of Justice and the Federal Trade Commission would consider a low concentration level.

Table 5: District-wise non-standardized regression coefficients, standardized regression coefficients, and shares as a percentage of total absolute standardized coefficients for the time window (2002-2021)

Rank	District	Coefficient	Standard Error	T-Statistic	Standardized Coefficient	Share (%)
1	Bahawalpur	1.4876*	0.0382	38.9485	0.1329	9.75
2	R. Y. Khan	0.8321*	0.0342	24.3434	0.1313	9.64
3	Vehari	0.8488*	0.0127	66.9246	0.1279	9.38
4	Lodhran	1.1458*	0.0316	36.2243	0.1254	9.20
5	Khanewal	0.9686*	0.0493	19.6621	0.1107	8.12
6	Muzaffargarh	1.1005*	0.0442	24.9157	0.1102	8.08
7	D.G.khan	1.5547*	0.0346	44.8823	0.0973	7.14
8	Bahawalnagar	0.7947*	0.0284	27.9419	0.0841	6.17
9	Rajapur	0.8690*	0.0267	32.4958	0.0826	6.06
10	Faisalabad	1.9962*	0.2620	7.6197	0.0474	3.48
11	Sahiwal	0.9168*	0.1015	9.0285	0.0471	3.46
12	Multan	0.4530*	0.1038	4.3635	0.0434	3.18
13	Layyah	2.0453*	0.0719	28.4377	0.0396	2.91
14	Pakpattan	0.9751*	0.0635	15.3561	0.0395	2.90
15	Jhang	1.1324*	0.1752	6.4622	0.0367	2.69
16	Bhakkar	1.4416*	0.0696	20.7097	0.0342	2.51
17	Okara	1.1330*	0.1361	8.3247	0.0273	2.00
18	Sargodha	4.1175*	0.5360	7.6822	0.0181	1.33
19	Kasur	2.5940*	0.5207	4.9814	0.0173	1.27
20	Mianwali	0.3138**	0.1070	2.9327	0.0099	0.73
Constant		-140485.7	99579.66	-1.4108		
Herfindahl-Hirschman Index						690

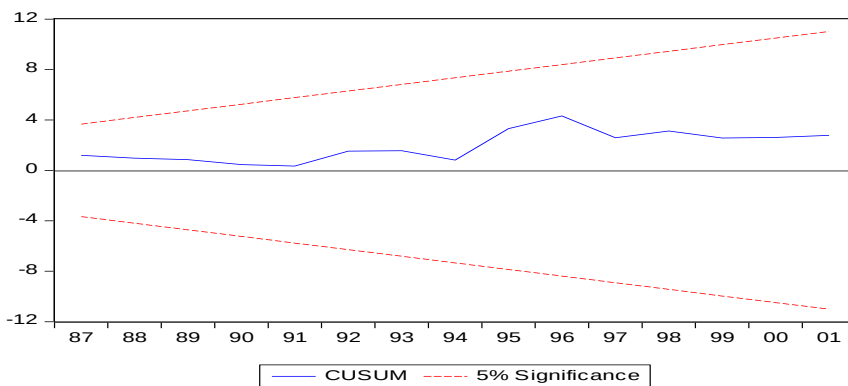
*, **, and *** indicate that the coefficients are significantly different from zero at the 1%, 5%, and 10% levels of significance, respectively.

District-wise changes in rankings, changes in standardized shares, and changes in the value of HHI during two time periods, (1982-2001) and (2002-2021), are reported in Table 6. Our findings show that the concentration level decreased in the period (2002-2021), as indicated by the 36-point decrease in HHI value. Between 1982 and 2001, the top four districts (Vehari, Bahawalpur, Khanewal, and Muzaffargarh) collectively held a 40.61 percent share, with Vehari holding 13.07 percent, Bahawalpur holding 9.85 percent, Khanewal holding 9.22 percent and Muzaffargarh holding 8.47 percent. Between 2002 and 2021, only a 35.33 percent share was contributed by these four districts, with a 9.38 percent share contributed by Vehari, a 9.75 percent share contributed by Bahawalpur, an 8.12 percent share contributed by Khanewal, and an 8.08 percent share contributed by Muzaffargarh. The Districts that indicated the greatest increase in ranking between the two-time windows are D. G. Khan (8 places), Rajapur (+4 places), and R. Y. Khan, Pakpattan, and Bhakkar (+3 places each). The Districts that lose their rankings between two time periods are Okara (-6 places), Multan (-4 places), and Layyah and Kasur (-3 places each). The Districts that made major improvements in standardized coefficient shares are D. G. Khan (+3.55 percent), Lodhran (+2.49 percent), and Bhakkar (+2.23 percent). The largest reduction in the standardized coefficient share occurred in district Vehari (-3.69 percent).

Table 6: District-wise change in shares and rankings during two time periods

District	Rank			Share (%)		
	1982-2001	2002-2021	Change	1982-2001	2002-2021	Change
Vehari	1	3	-2	13.07	9.38	-3.69
Bahawalpur	2	1	+1	9.85	9.75	-0.10
Khanewal	3	5	-2	9.22	8.12	-1.10
Muzaffargarh	4	6	-2	8.47	8.08	-0.39
R. Y. Khan	5	2	+3	7.66	9.64	+1.98
Lodhran	6	4	+2	6.71	9.20	+2.49
Bahawalnagar	7	8	-1	6.29	6.17	-0.12
Multan	8	12	-4	4.87	3.18	-1.69
Sahiwal	9	11	-2	4.84	3.46	-1.38
Layyah	10	13	-3	4.64	2.91	-1.73
Okara	11	17	-6	4.63	2.00	-2.63
Faisalabad	12	10	+2	4.44	3.48	-0.96
Rajanpur	13	9	+4	3.96	6.06	+2.10
Jhang	14	15	-1	3.64	2.69	-0.95
D. G. Khan	15	7	+8	3.59	7.14	+3.55
Kasur	16	19	-3	1.93	1.27	-0.66
Pakpattan	17	14	+3	1.44	2.90	+1.46
Mianwali	18	20	-2	0.46	0.73	+0.27
Bhakkar	19	16	+3	0.28	2.51	+2.23
Sargodha	20	18	+2	0.01	1.33	+1.32
HHI	726	690	-36			

The stability of the estimated parameters is checked for both the time periods by plotting the cumulative sum of residuals (CUSUM) and cumulative sum of square residuals (CUSUMSQ). Plots of CUSUM and CUSUMSQ for the time window 1982-2001 are shown in Figure 1 and Figure 2 respectively. Figure 1 and Figure 2, show that the plots of CUSUM and CUSUMSQ lie within the straight lines at 5% significance level, implying that the parameters of the first component model are structurally stable.

**Figure 1:** Cumulative Sum of Residuals of the component model during 1982-2001

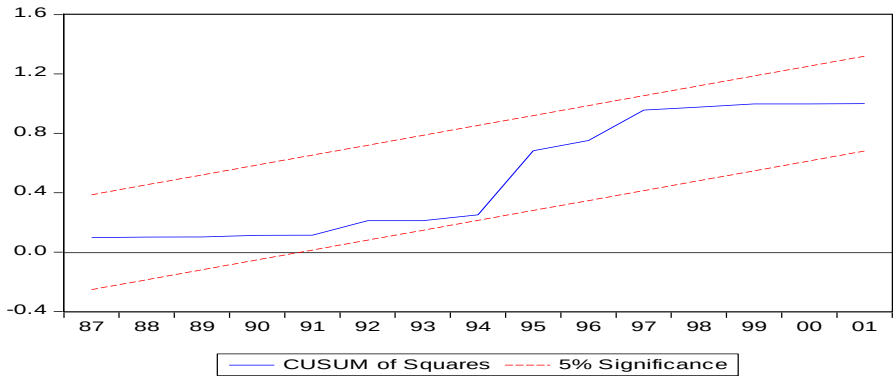


Figure 2: Cumulative Sum of Square Residuals of the component model during 1982-2001

Plots of CUSUM and CUSUMSQ for the time window 2002-2021 are shown in Figure 3 and 4 respectively. Figure 3 and 4 show that the plots of CUSUM and CUSUMSQ lie within the straight lines at 5% significance level, indicating that the parameters of the second component model are also structurally stable.

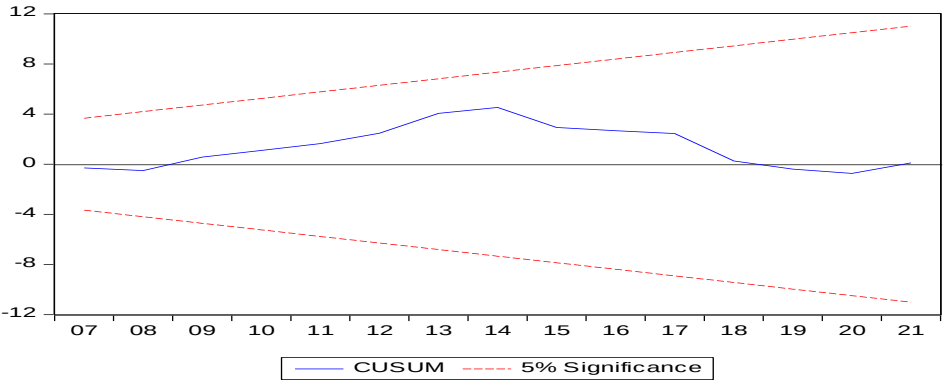


Figure 3: Cumulative Sum of Residuals of the component model during 2002-2021

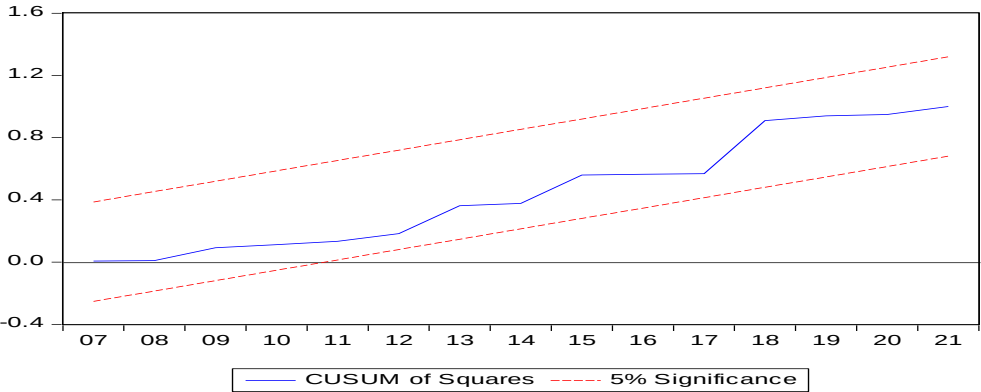


Figure 4: Cumulative Sum of Square Residuals of the component model during 2002-2021

5 Summary and Conclusion

The paramount objective of this study is to analyze the district-level geographic importance of

aggregate cotton production in Punjab using the correlated component regression technique. This methodology is well-suited for multicollinear and sparse datasets like those examined in the present study. The whole dataset (1982-2021) is divided into two periods: the first window includes crop years from 1982 to 2001, and the second window covers crop years from 2002 to 2021. In both time periods, the data has been analyzed through correlated component regression using aggregate cotton production in Punjab as a dependent variable and district-wise cotton production as independent variables. The percent shares of the absolute values of the standardized regression coefficients have examined each district's rank and relative importance. In each time period, the geographic concentration was measured using the Herfindahl-Hirschman Index. This index was calculated for both time periods using the percentage shares of the absolute values of the standardized regression coefficients, and the results are compared across the two periods. The stability of the estimated parameters is checked for both time periods by plotting the cumulative sum (CUSUM) and the cumulative sum of squares (CUSUMSQ) residuals.

Our findings show that the concentration level decreased, as indicated by a 36-point decrease in the value of HHI. A smaller HHI value in the time window 2002-2021 indicates that the shares are less concentrated geographically, implying greater geographic variation (across districts). During the time window 1982 to 2001, the top four districts (Vehari, Bahawalpur, Khanewal, and Muzaffargarh) collectively held a 40.61 percent share, with Vehari holding 13.07 percent, Bahawalpur holding 9.85 percent, Khanewal holding 9.22 percent and Muzaffargarh holding 8.47 percent. During the time window from 2002 to 2021, only a 35.33 percent share was contributed by these four districts, with a 9.38 percent share contributed by Vehari, a 9.75 percent share was contributed by Bahawalpur, an 8.12 percent share was contributed by Khanewal and an 8.08 percent share was contributed by Muzaffargarh.

The Districts that indicated the greatest increase in ranking between the two-time windows are D. G. Khan (+8 places), Rajanpur (+4 places), and R. Y. Khan, Pakpattan, and Bhakkar (+3 places each). The Districts that lose their rankings between two time periods are Okara (-6 places), Multan (-4 places), and Layyah and Kasur (-3 places each). The Districts that made major improvements in standardized coefficient shares are D. G. Khan (+3.55 percent), Lodhran (+2.49 percent), and Bhakkar (+2.23 percent). The largest reduction in the standardized coefficient share occurred in district Vehari (-3.69 percent). The CUSUM and CUSUMSQ plots for both time periods indicate structural stability in the estimated parameters of both correlated component models. From the aforementioned results, the following conclusion can be drawn:

- First, it is very important to note that the rankings determined by the correlated component regressions do not match with the average relative production shares of each district. This difference between the rankings determined by the CCR and the average relative production shares may differ because the CCR regression captures the direct (prime variable) as well as indirect (suppressor variable) effects of each district's production.
- Second, the estimated results of the correlated component regression for 2002-2021 show a decline in concentration, marked by a 36-point decrease in the HHI value. This indicates technological advancements and policy factors have contributed to greater geographic variation in the relative importance of each district during 2002-2021 compared to 1982-2001.

6 References

- Abbas, S. (2020). Climate change and cotton production: an empirical investigation of Pakistan. *Environmental Science and Pollution Research*, 27, 29580–29588. <https://doi.org/10.1007/s11356-020-09222-0>
- Ahmad, D., & Afzal, M. (2018). Estimating the economic perspective of cotton crop in Southern Punjab of Pakistan. *International Journal of Advanced and Applied Sciences*, 5(6), 50–55. <https://doi.org/10.21833/ijaas.2018.06.007>
- Ali, H., Hameed, R. A., Ahmad, S., Shahzad, A. N., & Sarwar, N. (2014). Efficacy of different techniques of nitrogen application on American cotton under semi-arid conditions. *Journal*

- of *Food Agricultural and Environment*, 12(1), 157–160.
- Arshad, M. U., Zhao, Y., Hanif, O., & Fatima, F. (2022). Evolution of overall cotton production and its determinants: Implications for developing countries using Pakistan case. *Sustainability*, 14(2), Article e840. <https://doi.org/10.3390/su14020840>
- Asian Development Bank. (2005, March 31). *Punjab economic report: Poverty reduction and economic management unit South Asia region*. <https://fid4sa-repository.ub.uni-heidelberg.de/3717/1/Pakistan%20Punjab%20Economic%20Report.pdf>
- Baig, I. A., Ullah, S., & Nasir, S. (2023). Policy impacts on comparative advantage and production protection to cotton and its competing crops in Pakistan. *The Pakistan Development Review*, 62(4), 539–552.
- 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. <https://doi.org/10.1017/aae.2020.36>
- Chaudhry, I. S., Khan, M. B., & Akhtar, M. H. (2009). Economic analysis of competing crops with special reference to cotton production in Pakistan: The case of Multan and Bahawalpur Regions. *Pakistan Journal of Social Sciences*, 29(1), 51–63
- Food and Agriculture Organization of the United Nations. (2021). *Recent trends and prospects in the world cotton market and policy developments*. <https://doi.org/10.4060/cb3269en>
- Government of Pakistan Finance Division. (2021). *Pakistan economic survey 2020-21*. https://www.finance.gov.pk/survey_2021.html
- in multiple regression models. *Journal of the Society for Social Work and Research*, 1(1), 28–40.
- Khan, S. B., Ahmad, F., Sadaf, S., & Kashif, R. H. (2010). *Crops area and production (by districts) (1981-82 to 2008-09): Volume I food and cash crops*. Pakistan Bureau of Statistics. https://www.pbs.gov.pk/sites/default/files/agriculture/publications/area_and_production_by_districts_for_28_years.pdf
- Khurshid, W., Qayyum, A., & Shera, M. J. (2021). Pakistan during 2015-16 TO 2020-21. *Journal of ISOSS*, 7(4), 229–240.
- Lynn, H. S. (2003). Suppression and confounding in action. *The American Statistician*, 57(1), 58–61. <https://doi.org/10.1198/0003130031090>
- Magidson, J. (2010, July 31–August 5). *Correlated component regression: A prediction/classification methodology for possibly many* [Paper presentation]. 2010 American Statistical Association Proceedings. Vancouver, British Columbia
- Magidson, J. (2011). *Rethinking regression, prediction and variable selection in the presence of high dimensional data: Correlated component regression* [Paper presentation]. Proceedings of Modern Modeling Methods (M³) Conference. University of Connecticut, Gentry, United States
- Magidson, J. (2013). Correlated component regression: Re-thinking regression in the presence of near collinearity. In H. Abdi, W. W. Chin, V. E. Vinzi, G. Russolillo, L. Trinchera (Eds.), *New perspectives in partial least squares and related methods* (pp.65–78). Springer.
- Magsi, H. (2012). Support Price: a growth rate model of cotton production in Pakistan. *Agricultural Journal*, 7(1), 21–25.
- Najib, D. C. S., Fei, C., Dilanchiev, A., & Romaric, S. (2022). Modeling the impact of cotton production on economic development in Benin: A technological innovation perspective. *Frontiers in Environmental Science*, 10, Article e926350. <https://doi.org/10.3389/fenvs.2022.926350>
- 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.
- Pandey, S., & Elliott, W. (2010). Suppressor variables in social work research: Ways to identify Planning & Development Board. (2023). *Punjab Growth Strategy 2023*. https://pnd.punjab.gov.pk/system/files/PGS_2023%2019-21-145.pdf
- Qazi, T. F., Niazi, A. A. K., & Basit, A. (2021). Assessment of agricultural performance of districts of Punjab. *Global Social Sciences Review*, 6(1), 158–172. [http://doi.org/10.31703/gssr.2021\(VI-I\).16](http://doi.org/10.31703/gssr.2021(VI-I).16)
- Rana, A. W., Ejaz, A., & Shikoh, S. H. (2020). *Cotton crop: A situational analysis of Pakistan*. [PACE Policy Working Paper]. International Food Policy Research Institute (IFPRI). <https://doi.org/10.2499/p15738coll2.133702>
- Rasheed, A., & Nosheen, M. (2024). Role of Women in Economic Development of Pakistan. *International Journal of Business and Management Sciences*, 5(2), 192–212.
- Rashid, M., Husnain, Z., & Shakoor, U. (2020). Impact of climate change on cotton production in Pakistan: an ARDL bound testing approach. *Sarhad Journal of Agriculture*, 36(1), 333–341 <http://doi.org/10.17582/journal.sja/2020/36.1.333.341>
- Raza, A., & Ahmad, M. (2015). *Analysing the impact of climate change on cotton productivity in Punjab and Sindh, Pakistan*. [Working Paper no. 9]. https://file.pide.org.pk/pdf/Climate_Change_9.pdf
- Rehman, A., Jingdong, L., Chandio, A. A., Hussain, I., Wagan, S. A., & Memon, Q. U. A. (2019). Economic perspectives of cotton crop in Pakistan: A time series analysis (1970–2015) (Part 1). *Journal of the Saudi Society of Agricultural Sciences*, 18(1), 49–54. <https://doi.org/10.1016/j.jssas.2016.12.005>
- 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), 145–58. <https://doi.org/10.1016/j.psrb.2016.09.002>
- Rehman, F. U., Saeed, I., & Salam, A. (2011). Estimating growth rates and decomposition analysis of agriculture production in Pakistan: pre and post sap analysis. *Sarhad Journal of Agriculture*, 27(1), 125–131.
- Sarfraz, S., Ali, F., Hameed, A., Ahmad, Z., & Riaz, K. (2023). Sustainable agriculture through technological innovations. In C. S. Prakash, S. Fiaz, M. A. Nadeem, F. S. Baloch, & A. Qayyum (Eds.), *Sustainable agriculture in the era of the OMICs revolution* (pp. 223–239). Cham: Springer International Publishing.
- Shabbir, M. S., & Yaqoob, N. (2019). The impact of technological advancement on total factor productivity of cotton: a comparative analysis between Pakistan and India. *Journal of Economic Structures*, 8(1), 1–16. <https://doi.org/10.1186/s40008-019-0160-4>
- Shuli, F., Jarwar, A. H., Wang, X., Wang, L., & Ma, Q. (2018). Overview of the cotton in Pakistan and its future prospects. *Pakistan Journal of Agricultural Research*, 31(4), Article e396. <http://doi.org/10.17582/journal.pjar/2018/31.4.396.407>
- 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*, 51(4), 261–274.
- Sinha, J. K. (2023). Economies of the factors affecting crop production in India: Analysis based on Cobb-Douglas Production Function. *Sumerianz Journal of Economics and Finance*, 6(1), 10–18.
- Tiffin, R., & Irz, X. (2006). Is agriculture the engine of growth? *Agricultural Economics*, 35(1), 79–89. <https://doi.org/10.1111/j.1574-0862.2006.00141.x>
- Yang, Z., Tang, J., Yu, M., Zhang, Y., Abbas, A., Wang, S., & Bagadeem, S. (2022). Sustainable cotton production through increased competitiveness: Analysis of comparative advantage and influencing factors of cotton production in Xinjiang, China. *Agronomy*, 12(10), Article e2239. <https://doi.org/10.3390/agronomy12102239>