Journal of Agricultural and Applied Economics (2021), **53**, 55–74
doi:10.1017/aae.2020.36

CAMBRIDGE UNIVERSITY PRESS

RESEARCH ARTICLE

The Influence of State-Level Production Outcomes upon U.S. National Corn and Soybean Production: A Novel Application of Correlated Component Regression

David W. Bullock

Agribusiness and Applied Economics, North Dakota State University, Fargo, ND 58102, USA
Corresponding author. Email: david.w.bullock@ndsu.edu

Abstract

The relative importance of key state-level outcomes upon U.S. national corn and soybean production was examined using *correlated component regression*, a recently developed regression technique for application to multicollinear and sparse data sets. Standardized coefficients were used to rank the states' relative importance. A Herfindahl-Hirschman Index was used to measure the degree of concentration among the top ranked states. The empirical analysis looked at two time periods: a pre-Genetic Modification (1975–1995) and a post-Genetic Modification (1996–2017) period. The results indicate that U.S. corn production is becoming less geographically concentrated in terms of state-level importance while the opposite holds true for soybean production.

Keywords: Correlated component regression; crop production; relative importance in prediction; geographic concentration; climate change; technological change; farm policy changes

JEL Classifications: C19; Q15; Q54; Q55

1. Introduction

During the growing season for major U.S. crops, most of the crop production information is provided at the state level. This holds true even through the harvest period. For example, the widely followed USDA *Crop Progress* reports provide information on crop progress (planting, development, and harvest percentages) and condition (categorical rating percentages) for the major production states for each major U.S. crop. These are reported on a weekly basis during the growing season. Additionally, soil moisture (sub- and topsoil strata) ratings are reported for the lower 48 states. Projected planted and harvested areas are also reported on a state level for the major states at multiple times during the growing season.

However, for most economists and industry analysts, the relevant and important production numbers for forecasting purposes are generally at the national level which feeds into the projected supply and demand balance tables such as those regularly reported in the USDA World Agricultural Supply and Demand Estimates monthly reports. This raises the question of whether certain states' production outcomes can provide a representative signal for the national production outcome. This question has particular importance to those engaged in a wide range of crop production and marketing activities ranging from logistics to storage to research and development on crop technology and cropping alternatives. The question is also important from a risk management perspective since history has shown that major crop production events, both negative and positive, tend to be limited spatially. Therefore, the more concentrated the geographic importance of a particular crop's production, from a national perspective, the more significant the

impact of any production event (drought, hurricane, etc.) if the event area contains a substantial portion of the total U.S. crop production.

Changes in crop production technology, farm policy, and global climate are having profound effects upon not only the level but also the geographic distribution of crop production in the U.S. and around the world. For example, since the first half of the 1990s, the acreage planted to soybeans in North Dakota has increased by over 800%—an increase of over 5.2 million acres which moves North Dakota from #18 to #9 in production ranking over the same period. In a recent article in the *Wall Street Journal* (Bunge, 2018), the effect of climate change and shorter-maturing crop varieties is evidenced by the changes in crop production occurring in Upper Alberta, Canada. Since 1950, average temperatures around La Crete, Alberta have increased by 3.6° Fahrenheit which has increased the growing season by nearly 2 weeks. While wheat and canola still dominate crop production in Canada, the area planted to corn has increased by 20% and soybeans has roughly doubled over the past decade alone. The article quotes Cargill CEO David MacLennan from a 2016 interview:

> "Today, the U.S. corn belt is in Iowa, Illinois, Indiana. In 50 years, it may be in Hudson Bay, Canada."

The main purpose of this study is to devise a methodology for measuring the relative importance of state-level production outcomes in predicting the national U.S. outcome for two crops: corn and soybeans. Relative importance is defined as the marginal impact of an individual state's over- or under-production, relative to recent history, upon the over- or under-production of U.S. aggregate crop production.

To measure relative importance, a production performance metric is constructed that measures the annual production level versus the normally observed level over the previous 5 years. These metrics were calculated for the 1975 through 2017 crop years using historical data (1970 to 2017) from the USDA-NASS. To measure the relative importance of each individual state upon the national crop production outcome, regressions were set up with the U.S. production performance metric as the dependent variable and the individual states' production performance metrics as the explanatory variables. The regression standardized coefficient values from each regression were used to rank each state in terms of its importance in impacting the national metric. Because of data sparsity and multicollinearity issues, the standard ordinary least squares regression model could not be used to derive accurate coefficient estimates. Therefore, a relatively new regression procedure designed to directly handle the issues of data sparsity and/or multicollinearity, called *correlated coefficient regression* (CCR), was applied to the data set.

A secondary purpose of this study was to measure the impact of recent technological, climatic, and policy factors upon the state ranking and also the geographic concentration of this importance. To accomplish this purpose, the data set containing the performance metrics was divided into two time periods: (1) the 1975 to 1995 crop years representing the period just prior to the commercialization of the first GM crop varieties, and (2) the 1996 to 2017 crop years which covers the period following the commercialization of GM varieties. The latter period can also be characterized by the increasing importance of bio-fuel production, major changes in farm policy moving away from supply control to a market-based income support emphasis, and the highly publicized increases in global temperatures and weather volatility which had been occurring even prior to 1996.

To measure the geographic concentration of importance, a *Herfindahl-Hirschman Index* (HHI) was calculated on the absolute percentage shares of the standardized coefficients. By comparing the derived HHI from each time period for each crop, the impacts of the technological, climatic, and policy factors upon the concentration in production risk can be observed.

The results of this study show an increase in the HHI concentration measure for corn between the two periods while declining for soybeans. Changes in the state rankings for corn indicate that expansion into states with significant irrigation (Nebraska and Texas) has contributed to the

https://doi.org/10.1017/aae.2020.36 Published online by Cambridge University Press

geographic dispersion. For soybeans, the increase in concentration appears to be driven by increased competition of Roundup Ready and shorter-maturing varieties of soybeans that has supplanted acres previously planted to wheat, particularly in the northern Corn Belt. These results hold a number of implications for those involved in assessing and forecasting U.S. national corn and soybean production.

This study is organized as follows. In the next section, a review of the previous literature is conducted with a particular focus on the impacts of climate change, technology, and farm policy upon crop yields and the geographic distribution of acreage and production. This is followed by a section that describes the data and the research methodology utilized in this study. The final two sections present the regression results, and a discussion of the major observations and conclusions from this research.

2. Previous Studies

An exhaustive review of the literature produced no evidence of previous studies that directly examine the main question posed in this study. Instead, the historical focus has been focused primarily on the examination of recent observed climatic, technological, and farm policy changes and their impact upon either crop yields and/or the geographic distribution of planted area and production.

Among the earliest studies examining the impact of climate change and crop technology upon regional corn and soybean yields were those by Thompson (1969, 1970, 1986, and 1988). Among the major findings was that there existed overlaps in the optimal seasonal growing conditions for both corn and soybeans (1970), that the observed trend of increasing CO_2 in the earth's atmosphere was a major contributor to increases in corn yields (1986), and that cyclical patterns of approximately 18 months in corn yields might be tied to lunar and El Nino cyclical phenomena (1988).

Menz and Pardey (1983) also attempted to separate the impacts of weather from crop technology with regard to the question of whether corn yields were reaching a plateau given the increasing rates of nitrogen fertilization which was an explanatory variable in their model. Their results indicated a much reduced but still positive marginal physical product from nitrogen application and a near constant contribution (1 bushel per year) that could be attributed to other technologies.

A more recent and comprehensive analysis of the impacts of climate and technology upon corn and soybean yields is contained in two concurrent studies by Tannura, Irwin and Good (2008a, 2008b) which employed the unknown breakpoint tests of Quant (1960) and Andrews (1993) to test for structural change in relationship between weather and technology with regard to corn and soybean yields. They used a modified version of Thompson's (1988) model. Their results indicated an asymmetry between yields that were negatively impacted by unfavorable weather versus those positively impacted by favorable weather. Additionally, they found no conclusive evidence of any structural changes in the relationship between the weather/technology variables and crop yields. Finally, they compared the forecasting accuracy of the modified Thompson model with the USDA yield trendline-based forecasting models using a variety of forecast metrics. Their results indicated that the modified Thompson model outperformed the USDA trendline model later in the growing season (August 1 or later for corn, September 1 or later for soybeans). Also, an application of encompassing tests indicated that combining the modified Thompson model with the USDA model could significantly improve forecast accuracy by an average of 10% for corn and 6% for soybeans.

Among the other studies that examined the impacts of technology and climate change upon crop yields include Garcia et al. (1987) who assessed the impact of weather and technology upon U.S. corn yields across two time periods: 1931–1960 and 1961–1982. They found that weather was more important in explaining the yield variability between the two time periods. Kaufmann and Snell (1997) estimated a hybrid model that accounted for both climatic and social impacts upon

https://doi.org/10.1017/aae.2020.36 Published online by Cambridge University Press

corn yields in the U.S. They found that social impacts had an important impact upon yields—particularly in assessing the relative costs and benefits for adoption of specific cropping practices. Lobell, Schlenker, and Costa-Roberts (2011) developed a database of yield response models to evaluate the impacts of climate trends on global yields by country over the 1980–2008 period. For corn and wheat, the results indicated a global net loss of 3.8% and 5.5%, respectively, while regional variations essentially balanced out the impacts on rice and soybeans.

Cai et al (2013) developed a principal components-based climate index to estimate the linkage between climate and regional U.S. crop yields. Their results indicated that future hot and dry weather conditions were more likely to have a significant impact upon crop yields in the southern U.S. when compared to other states. Schlenker and Roberts (2009) examined nonlinear and asymmetric relationships between temperature and crop yields with regard to long-term temperature projections. They projected that area-weighted average yield will decline by 30%–46% under the slowest warming scenario and by 63%–82% under the most rapid warming scenario (Hadley III model). Lobell et al. (2014) examined yield sensitivity to drought conditions for Central U.S. corn and soybeans using data from 1995 to 2012. They found that a key factor impacting sensitivity to drought was the planting density although new varieties were more robust to the crowding effect.

Tolhurst and Ker (2015) modeled the impact of technical change upon crop yields using the mixture of normals (EM algorithm) model with embedded trend functions to account for technological change for corn, soybeans, and wheat between 1955 and 2011. They found that the rate of technical change altered the shape of the crop yield distributions particularly with relation to the higher moments. Du et al. (2015) examined the impact of geographic and climatic factors upon the skewness of crop yield distributions for corn and soybeans in 13 Midwestern states from 1990 to 2009. Their results indicated that better natural resource endowments (climate and soils) decreased the observed skewness in yields. Chen, Chen, and Xu (2016) examined the impact of climate change upon corn and soybean yields in China. Their results indicated a loss of about US$820 million to both sectors over the previous decade due to climate change with corn and soybeans yields projected to decline by 3%–12% and 7%–19%, respectively, by 2100.

Kukal and Irmak (2018) examined the long-term variability in climate and yields for corn, soybeans, and sorghum using 46 years of county-level data from 9 Great Plains states. They found that irrigated yields were more robust to the impacts of climate change with considerable geographic variations in the results. Huffman, Jin, and Xu (2018) examined a half-century of panel data on U.S. Midwest non-irrigated corn yields using a component model that attributed yield effects to nitrogen fertilization levels, public investment in corn research, biotechnology, weather, and pests. They found that nitrogen use, public investment in corn research, and biotechnology increased corn yield potential while excessive heat significantly reduces nitrogen productivity. Another important finding was that biotechnology primarily reduced yield damage due to soil moisture stress but had little effect upon damage due to excessive heat.

A number of studies have examined the impacts of technology and climate change upon other crop production variables such as planted area, resource endowments, land values, and profits. Among them are Rosenzweig and Parry (1994) who combined data from individual crop yield studies to obtain a global picture of the simulated change in crop production associated with different climate scenarios. They observed that climate change was creating major disparities in agricultural production vulnerability between developed and developing countries. Mendelsohn, Nordhaus, and Shaw (1994) measured the economic impact of climate change upon land prices using a Ricardian modeling approach using cross-sectional data from over 3,000 counties in the U.S.. They found that the most negative impacts of climate change would be realized upon cropland values in the southern U.S. when compared to other regions. Schlenker, Hanemann, and Fisher (2006) also utilized a Ricardian regression approach to examine the link between climatic variables (degree days and precipitation) and U.S. farmland values east of the 100th meridian (i.e., farmland not dependent upon irrigation). Their results also showed varying positive and negative

https://doi.org/10.1017/aae.2020.36 Published online by Cambridge University Press

impacts upon farmland values on a county-by-county basis with the number of counties exhibiting losses exceeding those exhibiting gains.

Deschenes and Greenstone (2007) examined the impact of climate change (i.e., year-to-year variations in temperature and precipitation) upon agricultural profits using U.S. county-level panel data. Their projections indicated that long-run climate predictions (i.e., Hadley 2 Model) would result in a $1.3 billion (2002 $) or 4.0% increase in aggregate U.S. agricultural profits by December 2099 with a high level of variation among the states with South Dakota having the largest increase (+$720 million) and California having the largest decrease (−$750 million). Marshall et al. (2015) examined the impact of climate change upon regional water balances in the U.S. and the resultant impact upon the distribution between irrigated and non-irrigated cropping area. They found that the impact of irrigation water supply was small relative to the direct biophysical impacts upon crop yields. Miao, Khanna, and Huang (2016) investigated the effect of price and climate variables upon non-irrigated U.S. corn and soybean production using a county-level panel data set covering the 1977–2007 period. They found that climate change was directly responsible for declines in corn production ranging from 7% to 41% and declines in soybean production from 8% to 45% when controlling for the price effects.

Burke and Emerick (2016) examined the impact of climatic variations upon long-run climate change adaptation strategies used by U.S. crop producers. Their results indicated that a lack of adaptation by producers was more the result of inadequate available options or implementation costs rather than a lack of knowledge regarding climate change. Haile et al. (2017) examined the global determinants of crop production for corn, wheat, rice, and soybeans over the 1961–2013 period using country-level data. Using climate prediction models, they projected that climate change will reduce global food production by 9% in the 2030s and 23% in the 2050s with a large variability across countries and crops. Fei, McCarl, and Thayer (2017) examined the effect of historical patterns in precipitation, temperature, and atmospheric gas composition along with the frequency of extreme weather events upon cereal grain acreage in the U.S. Their results indicated that climate change would induce regional shifts in planted wheat area with northward shifts in the southern Great Plains, westward shifts in the northern Great Plains, and eastward shifts in the Pacific Northwest.

A recent study focusing upon policy implications related to geographic crop area is Li, Miao and Khanna (2019) who examined the expansion of ethanol production in the U.S. and its impact on land-use when controlling for the effect of changes in relative crop prices. They found that an increase in ethanol capacity has led to a modest 3% increase in corn acreage and less than a 1% increase in total crop acreage between 2008 and 2012. The price effect was found to be twice as large as ethanol capacity, but the effect was essentially reversed by the sharp downturn in prices after 2012.

3. Data and Methodology

To calculate the relative over- or under-performance of annual corn and soybean production, a *production performance index* (PPI) was calculated as a proxy for the state and national production level relative to recent history. The formula for the PPI is as follows:

$$PPI_t = P_t - O(P_{t-1}, \cdots, P_{t-5}), \tag{1}$$

where P_t is the production level in time period t, and $O(\cdot)$ is the Olympic average function (drop minimum and maximum values and average the remaining three values). Therefore, the PPI measures the degree by which the current year's production either exceeded (over-performed) or fell short of (under-performed) the normal production level from the preceding 5 years. Because the PPI uses lagged differencing, the resulting series is generally stationary and the effects of autocorrelation in crop yields are reduced.

https://doi.org/10.1017/aae.2020.36 Published online by Cambridge University Press

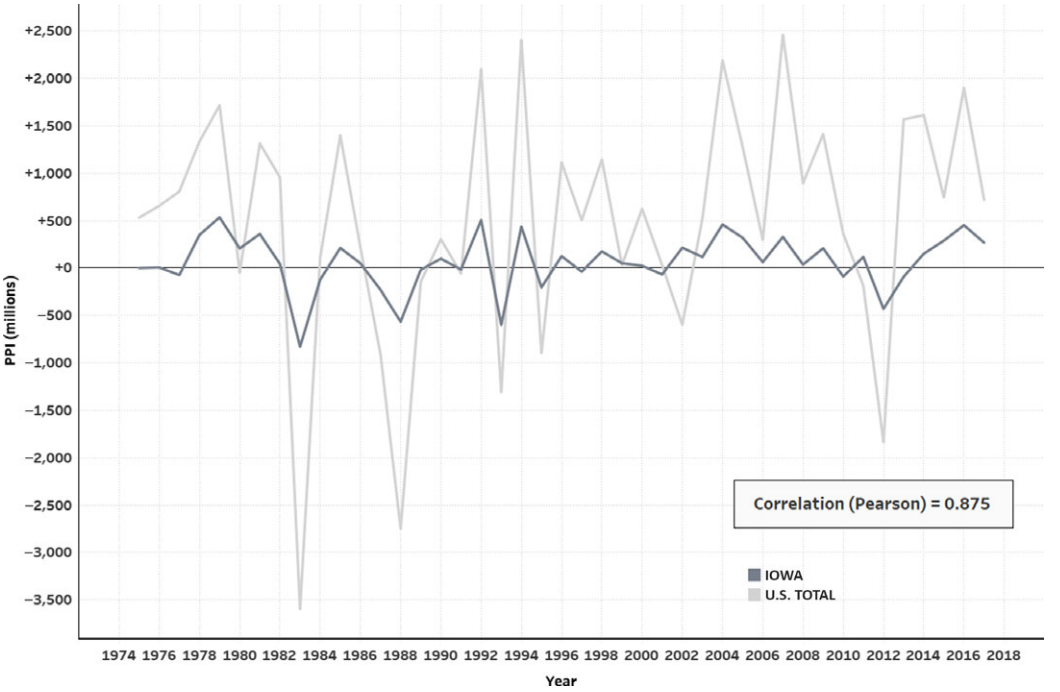

Figure 1. Production performance index (PPI) values for U.S. and Iowa Corn, 1975–2017.

The Olympic average was used to set a benchmark of what could be considered a "normal" production level for a particular region (state or national) at the time of comparison. This measure was chosen over other measures (such as trendline fits and moving averages) because it has a lower data requirement and minimizes the effect of recent extreme values. Also, the Olympic average is used in many USDA crop insurance programs in calculating an actual production history yield, particularly when there is insufficient data to estimate a trendline yield. The 5-year lag was used to make sure that the average represented a more recent measure that was representative of what would be considered a "normal" level of crop production for a particular state.

The PPI is illustrated for the U.S. and Iowa in Figure 1. As would be expected, there is a high degree of correlation (87.5%) between the two measures since Iowa is a major source of U.S. corn production. However, the Iowa measure is less volatile than the U.S. measure due to its smaller production area, deep soils of high quality, and ideal weather for corn production.

The data set used in this study is comprised of the state-level corn and soybean production from 1970 to 2017 for the 18 major corn producing (Colorado, Illinois, Indiana, Iowa, Kansas, Kentucky, Michigan, Minnesota, Nebraska, North Carolina, North Dakota, Ohio, Pennsylvania, South Dakota, Tennessee, Texas, and Wisconsin) and soybean producing (Arkansas, Illinois, Indiana, Iowa, Kansas, Kentucky, Louisiana, Michigan, Minnesota, Mississippi, Missouri, Nebraska, North Carolina, North Dakota, Ohio, South Dakota, Tennessee, and Wisconsin) states as determined by the USDA. The data set also contains the national-level production for both crops, and an "Other States" aggregate is derived as a residual when subtracting the sum of the 18-state production from the national aggregate. The data come from the USDA-NASS Quick Stats online database (https://quickstats.nass.usda.gov).

Using the national and state-level production data, the PPI was calculated for the national and state totals from 1975 to 2017 (allowing for 5-year lag in PPI formula). For estimation purposes, the PPI data were split into two time periods: (1) the period preceding the introduction of

https://doi.org/10.1017/aae.2020.36 Published online by Cambridge University Press

Roundup Ready soybeans and Bt corn in 1996 referred to as the *pre-GM* period, and (2) the period coinciding with and following the introduction of the GM events referred to as the *post-GM* period. The former time period covers the 1975–1995 crop years while the latter covers the 1996–2017 crop years. While it is noted that there are many factors that have influenced crop production over the 1975–2017 period, none has likely produced such a major point of demarcation in U.S. corn and soybean production as the introduction of GM varieties in 1996.

For corn, the 18 major states comprised an average 94.3% share of total U.S. production over the 48 years in the database (1970–2017). The minimum share was 91.5% in 1976 while the maximum share was 95.6% in 2006. For the pre-GM time window (1975–1995), Iowa had the largest average production share at 19.2% followed by Illinois (17.3%), Nebraska (10.9%), Indiana (9.0%), and Minnesota (8.4%). For the post-GM time window (1996–2017), Iowa also had the largest average production share at 18.2% followed by Illinois (16.0%), Nebraska (11.7%), Minnesota (9.9%), and Indiana (7.3%).

For soybeans, the 18 major states comprised an average 94.3% share of total U.S. production over the 48 years in the database. The minimum share was 89.1% in 1982 and the maximum share was 97.4% in 2002. For the pre-GM time window (1975–1995), Illinois had the largest average production share at 17.6% followed by Iowa (16.1%), Indiana (8.4%), Minnesota (8.2%), and Missouri (8.2%). For the post-GM time window (1996–2017), Iowa had the largest average production share at 15.3% followed by Illinois (14.7%), Minnesota (9.4%), Indiana (8.1%), and Nebraska (7.2%).

The change in average corn production by leading state from the last 5 years of the pre-GM period (1991–1995) to the last 5 years of the post-GM period (2013–2017) is geographically illustrated in Figure 2. Iowa had the largest absolute change (+961.3M bushels) in average production while North Dakota (+919.1%) had the largest percentage increase. In terms of yield, North Dakota had the largest average gain (+72.8 bushels per acre) followed by Nebraska (+67.9), Minnesota (+64.7), Tennessee (+62.2), and Iowa (+62.0). States with the largest gain in planted area were Kansas (+2.6M acres), North Dakota (+2.4M), South Dakota (+2.2M), Minnesota (+1.5M), and Nebraska (+1.4M).

Figure 3 shows the corresponding production change information for the top soybean producing states. Nebraska had the largest absolute change (+187M bushels) in average soybean production between the two periods while North Dakota had the largest percentage change (+1,124.7%). States with the largest yield increases were Louisiana (+22.4 bushels per acre), Mississippi (+22.3), Arkansas (+18.6), Nebraska (+18.5), and South Dakota (+15.8). The largest increases in planted area were North Dakota (+5.2M acres), South Dakota (+2.9M), Nebraska (+2.6M), and Minnesota (+1.9M).

To analyze and rank the relative contribution of each state to the overall U.S. PPI index, one standard approach is to estimate a linear regression model with the U.S. PPI as the dependent variable and the individual states' PPI as the independent variables. The standardized coefficient values from the linear regression can be used to rank the individual states and test for significance. However, there are a couple of issues that arise with the application of linear regression to this particular data set: (1) with the data set split between 1975 to 1995 and 1996 to 2017 there are nearly as many independent variables as there are observations (referred to as the *sparsity* problem), and (2) the independent variables have a high level of correlation (*multicollinearity*) which can overstate the variance of the individual estimators and result in inaccurate individual coefficient values and *t*-statistics (Kennedy, 1998).

Traditional regression approaches to the problems of sparse data and multicollinearity typically involve one of four approaches (Kuhn and Johnson, 2013): (1) *acknowledge and ignore* the problem, (2) *leave out or combine* the most problematic variables, (3) utilize a *penalized regression method* such as *ridge* (Hoerl and Kennard, 1970) or *lasso* (Tibshirani, 1996) regression which trade off increased bias in exchange for reduced variance in the parameter estimates, or (4) utilize a dimension reduction method such as *principal components regression* (PCR; Massy, 1965), *supervised PCR* (Bair et al., 2006), or *partial least squares* (PLS; Wold, 1966) which reduces

https://doi.org/10.1017/aae.2020.36 Published online by Cambridge University Press

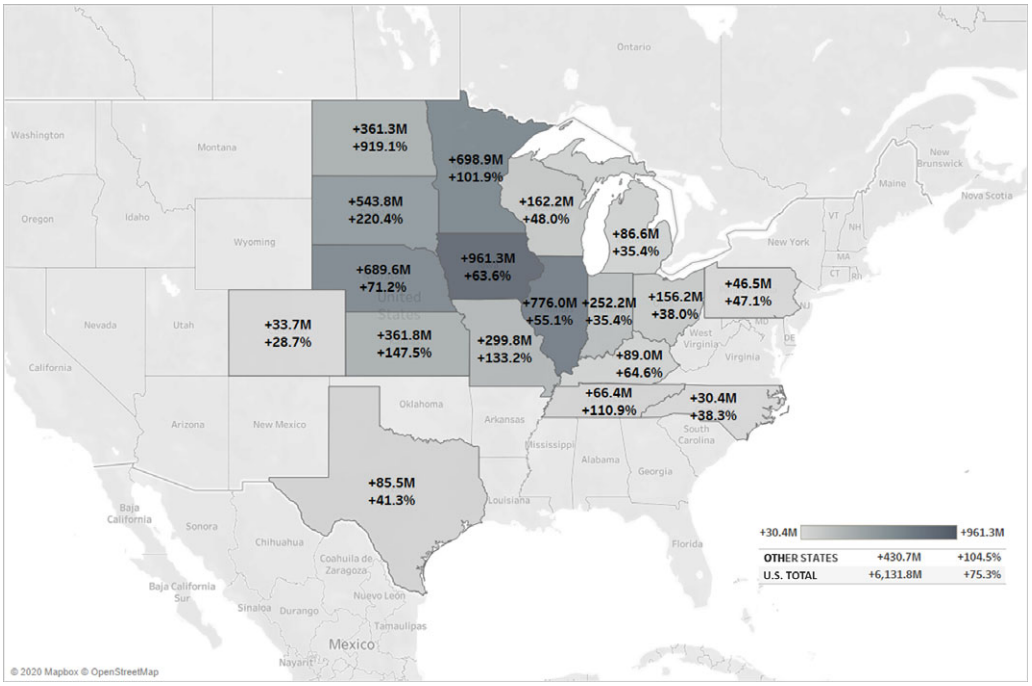

Figure 2. Change in average corn production by state, 2013–2017 versus 1991–1995.
(Source: USDA-NASS, Stats Online Database)

multicollinearity by removing redundant features in the data set through the condensation of a large number of explanatory variables into a smaller set of component or latent variables that preserve a high percentage of the information in the original data set.

Another, more recent development in this area has been the introduction (Magidson, 2010) of *correlated component regression* (CCR) which is a dimension reduction method. While the other dimension reduction methods rely upon extracting orthogonal component variables to reduce multicollinearity, CCR relies exclusively upon dimension reduction and the inclusion of *suppressor variables* which directly reduce the confounding effects of multicollinearity, thus obtaining more reliable parameter estimates. CCR also has the added benefit in that its component (latent) variables are much easier to interpret when compared to the other dimension reduction methods.

The CCR model is based upon two tuning parameters: the number of components (latent variables) to be derived (k) and the number of independent variables to retain in the model (p). Tuning of these parameters is done using a cross-validation (CV) procedure such as *m-fold validation*. A CV metric such as R^2, mean squared error (MSE), or area under the receiver operating curve (AUROC) is chosen for optimization under the CV procedure. The CCR procedure is extremely flexible and versions have been developed for modeling ordinary least squares (CCR-Linear), logistic (CCR-Logistic), linear discriminant (CCR-LDA), survival (CCR-Cox), and latent variable (CCR-Latent) models.

The CCR-Linear algorithm begins with all P of the independent variables and estimates the following P single-variable regression equations:

$$Y_i = \delta_g^{(1)} + \lambda_g^{(1)} \cdot X_{g,i}, \tag{2}$$

where Y_i is observation i of the dependent variable with $i = 1, \ldots, N$; $X_{g,i}$ is observation i of independent variable $g = 1, \ldots, P$; $\delta_g^{(1)}$ and $\lambda_g^{(1)}$ are the regression intercept and slope parameters for independent variable g.

https://doi.org/10.1017/aae.2020.36 Published online by Cambridge University Press

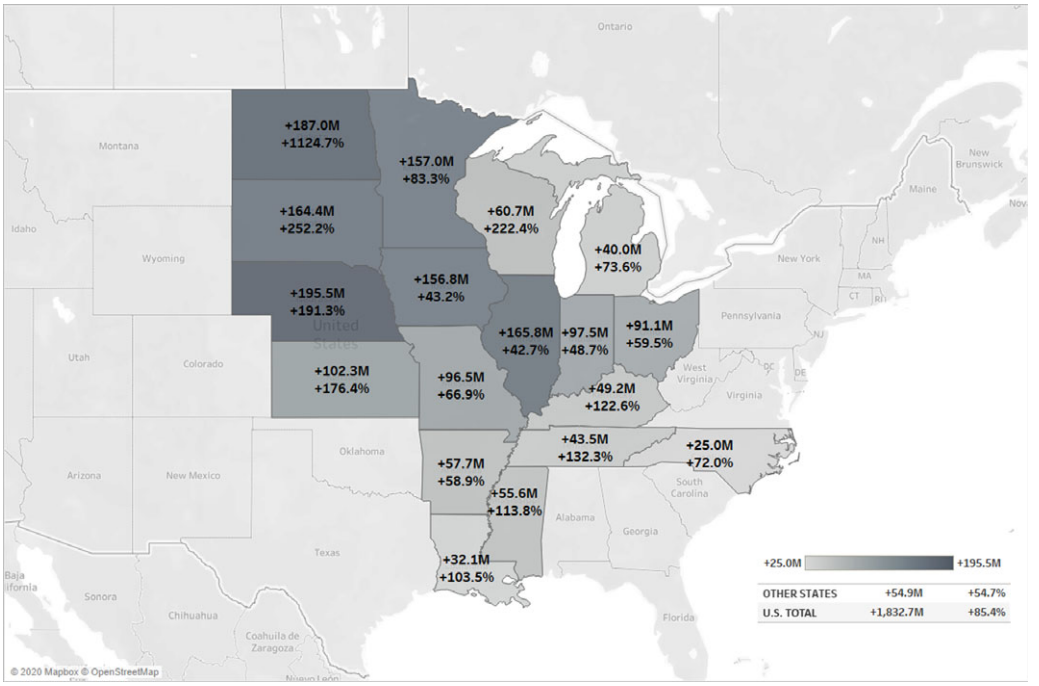

Figure 3. Change in average soybean production by state, 2013–2017 versus 1991–1995.
(Source: USDA-NASS, Stats Online Database)

The first correlated component variable, CC_1, is then constructed as the linear, weighted average of each predictor using the single equation slope coefficients as the weights:

$$CC_{1,i} = \frac{1}{P} \cdot \sum_{g=1}^{P} \hat{\lambda}_g^{(1)} \cdot X_{g,i}, \tag{3}$$

From (3), the 1-component model is then estimated as:

$$Y_i = \alpha^{(1)} + \beta_1^{(1)} \cdot CC_{1,i}, \tag{4}$$

with the relevant CV metric (CV-R^2 or CV-MSE) stored from this regression model for later determination of the optimal number of retained components in the model. The CC_1 component is called the *direct effects* component since it measures the direct impact of each independent variable upon the dependent variable without any latent suppressor effects.

The second component variable (CC_2) is constructed by first estimating the following regression equation for each of the independent variables:

$$Y_i = \delta_g^{(2)} + \gamma_{1,g}^{(2)} \cdot CC_{1,i} + \lambda_g^{(2)} \cdot X_{g,i}, \tag{5}$$

then, using the results from (5) to derive the correlated component:

$$CC_{2,i} = \frac{1}{P} \cdot \sum_{g=1}^{P} \hat{\lambda}_g^{(2)} \cdot X_{g,i}, \tag{6}$$

https://doi.org/10.1017/aae.2020.36 Published online by Cambridge University Press

The second component CV metric is then derived from the following regression equation:

$$Y_i = \alpha^{(2)} + \beta_1^{(2)} \cdot CC_{1,i} + \beta_2^{(2)} \cdot CC_{2,i}, \tag{7}$$

Note that CC_2 and subsequent derived component variables represent *latent suppressor effect* variables in the correlated component model.

The progression of component derivations can continue up to the point where the number of components equals the number of independent variables if the model is fully identified and non-sparse, or equals the maximum number specified by the user. In the case where the number of components equals the number of independent variables, the initial regression model would be:

$$Y_i = \delta_g^{(P)} + \gamma_{1,g}^{(P)} \cdot CC_{1,i} + \cdots + \gamma_{P-1,g}^{(P)} \cdot CC_{P-1,i} + \lambda_g^{(P)} \cdot X_{g,i}, \tag{8}$$

With the final component equal to:

$$CC_{P,i} = \frac{1}{P} \cdot \sum_{g=1}^{P} \hat{\lambda}_g^{(P)} \cdot X_{g,i}, \tag{9}$$

and

$$Y_i = \alpha^{(P)} + \beta_1^{(P)} \cdot CC_{1,i} + \cdots + \beta_P^{(P)} \cdot CC_{P,i}, \tag{10}$$

to determine the final CV metric.

Once the optimal number of components (K^*) is determined using the maximum value of the CV metric, the number of independent variables can either be specified by the user (up to P) or can be optimally determined using a *step-down procedure*. The first step in the step-down procedure is to estimate the model with all of the predictors and calculate the CV metric. In the next step, the independent variable with the smallest absolute value of its standardized coefficient is removed and the model is re-estimated with the reduced set of independent variables. This is repeated until there is only one independent variable left in the data set. The optimal set of independent variables (P^*) is the set that produces the maximum value of the CV metric.

Initial results presented by Magidson (2010) indicated that CCR can perform as well or better than PCA-based and penalty function approaches when forecasting out-of-sample values from sparse and multicollinear data sets. In another paper (Magidson and Wassmann, 2010), the potential value of latent suppressor variables in the detection of prostate cancer was demonstrated using an application of CCR-LDA to patient data.

To date, most published applications of CCR have been in the medical and sociological fields (Alkerwi et al., 2015). Given its recent development, applications of CCR in the economics literature have been relatively sparse with the exception of a paper presented at a math and engineering conference (Trivedi and Birau, 2013) that examined correlation between several international stock indices and a recent logistics paper (Garver and Williams, 2018).

In this study, the Magidson CCR model was applied to the derived PPI data sets to rank the relative importance of each state in predicting the PPI on a national scale using the absolute value of the standardized coefficients to produce the ranking. This was done for both the pre-GM and post-GM time windows for both corn and soybeans.

The results are presented in a standardized coefficient format which is retained by converting all of the variables (dependent and independent) into equivalent z-scores and running the regressions on the transformed variables. Standardized coefficients are often used to answer the question of which of the independent variables have a greater effect upon the dependent variable. Essentially, the standardized coefficient shows the marginal effect (in standard deviations) on the dependent variable of a one-standard deviation increase in the independent variable. This provides a better benchmark of comparison as it removes any bias due to different unit measurements and/or variable size. Note that the t-statistics are the same whether the coefficient is standardized or non-standardized. The states are

https://doi.org/10.1017/aae.2020.36 Published online by Cambridge University Press

ranked based upon the absolute value of their standardized coefficient and the percentage share of the total sum of absolute coefficients. Note that the ranking includes states that can have both positive and negative impacts upon national production. States that have a higher ranking indicate that a one standard deviation increase in their PPI metric has a larger impact (in standard deviations) upon the national measure. The percentage shares are reported and used in the calculation of the measure of geographic concentration.

To measure the geographic concentration of the state-level impacts, a HHI was calculated using the following formula:

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

where n is the number of states in the regression and s_i is the standardized coefficient share for the i-th state (in decimal format). The HHI is a commonly used measurement of market concentration in applied industrial organization research and antitrust policy; however, the index can be used as a general measure of concentration in many applications. For a market monopoly with one firm holding a 100% share of the market, the HHI has a value of 10,000 representing its maximum.

Changes in the HHI can be used to examine the degree of change in geographic concentration across the two time periods using the standardized coefficient shares. A primary hypothesis examined was the introduction of new crop technologies, combined with changes in climate and farm policy, would result in a greater geographic dispersion of the relative importance of each state upon the national aggregate. This would imply a decline in the HHI when comparing the latter 1996–2017 period to the former 1975–1995 period for each crop.

The CCR-Linear regression was applied to the PPI data sets using the COREXpress™ software package which is sold by Statistical Innovations (www.statisticalinnovations.com). The maximum number of correlated components (p) was set to eight for each estimation and the step-down procedure of variable selection was not utilized (all of the states were retained in the estimated model). Each model was estimated using an out-of-sample four-fold CV option with CV-R^2 as the CV metric in the CCR-Linear procedure. Application of the other available CV metric (CV-MSE) produced identical results to those from CV-R^2. Note that the CV-AUROC metric is only available for qualitative dependent models (CCR-Logistic and CCR-LDA).

4. Results

The organization of this section will first present the CCR results for corn followed by soybeans. For each crop, the pre-GM (1975–1995) period results are presented first followed by the post-GM (1996–2017) results. A discussion contrasting the changes between the two time periods for both crops is included in the section that follows.

4.1. Corn: Pre-GM (1975–1995) Period

The CV procedure for corn in the pre-GM time frame (1975–1995) resulted in a maximum CV-R^2 of 98.7% with four components retained. All of the estimated correlated component coefficient values ($\hat{\beta}_i$) had positive values and were highly significant at the 99% confidence level.

The individual states' non-standardized and standardized coefficient values along with the percent standardized share (absolute value) are shown in Table 1. The states are ranked in order of their share of the absolute standardized coefficient values. Note that each state's (g) non-standardized coefficient value ($\hat{\phi}_g$) is calculated as:

$$\hat{\phi}_g = \sum_{i=1}^{K} \hat{\lambda}_g^{(i)} \cdot \hat{\beta}_i, \tag{11}$$

https://doi.org/10.1017/aae.2020.36 Published online by Cambridge University Press

Table 1. Individual states' coefficients and shares, corn, 1975–1995 time period

Rank[a]	State	Coefficient	Standard Error	*T*-Statistic	Pr > |*t*|	Signif[b]	Std Coefficient	Share (%)[a]
1	IA	1.167	0.061	19.225	<0.0001	***	0.282	23.1
2	IL	0.993	0.053	18.659	<0.0001	***	0.201	16.5
3	MN	0.810	0.074	10.990	<0.0001	***	0.092	7.6
4	IN	0.899	0.071	12.702	<0.0001	***	0.088	7.2
5	SD	1.577	0.122	12.896	<0.0001	***	0.068	5.6
6	MI	2.091	0.126	16.641	<0.0001	***	0.068	5.6
7	KS	2.373	0.092	25.890	<0.0001	***	0.062	5.1
8	WI	1.050	0.074	14.236	<0.0001	***	0.060	4.9
9	OTHER	0.952	0.052	18.193	<0.0001	***	0.058	4.8
10	MO	0.995	0.053	18.823	<0.0001	***	0.044	3.6
11	NC	2.197	0.115	19.028	<0.0001	***	0.039	3.2
12	OH	0.604	0.072	8.356	<0.0001	***	0.038	3.1
13	KY	1.341	0.227	5.911	<0.0001	***	0.031	2.5
14	ND	3.313	0.727	4.556	0.0003	***	0.025	2.0
15	PA	1.243	0.355	3.496	0.0030	***	0.024	2.0
16	NE	0.126	0.082	1.532	0.1450		0.012	1.0
17	CO	0.835	0.367	2.277	0.0369	**	0.010	0.9
18	TN	0.866	0.875	0.990	0.3369		0.010	0.8
19	TX	−0.184	0.107	−1.714	0.1059		−0.005	0.4
[Constant]		23.144	19.423	1.192	0.2508			

[a]Rank and share based upon absolute value of standardized coefficient.
[b]***Significantly different from zero at 99% confidence level, **95%, *90%.

where K is the number of retained correlated components. One of the weaknesses of the COR-Express software package is that it does not provide coefficient standard errors and t-statistics for the regression coefficients; however, these can be derived from the regression standard errors of the component coefficients as follows:

$$se(\hat{\phi}_g) = \sqrt{\sum_{i=1}^{K} \left(\hat{\lambda}_g^{(i)}\right)^2 \cdot \left(se(\hat{\beta}_i)\right)^2}, \tag{12}$$

where $se(\cdot)$ is the standard error for the regression coefficient. The constant coefficient represents the intercept ($\hat{\alpha}$) from the optimal CV-R^2 regression.

The results indicate that during the 1975–1995 period, production outcomes in Iowa and Illinois had a dominant share (39.6%) of the impact upon the overall national corn production outcome. The next tier of states (Minnesota and Indiana) had an approximate 15% share. Note all of the coefficients are statistically significant at the 95% confidence level with the exception of Nebraska, Tennessee, and Texas. The HHI on the standardized coefficient shares was equal to 1,103 which is below the 1,500 threshold that the U.S. Department of Justice would consider a moderate level of concentration in antitrust applications.

https://doi.org/10.1017/aae.2020.36 Published online by Cambridge University Press

Table 2. Individual states' coefficients and shares, corn, 1996–2017 time period

Rank[a]	State	Coefficient	Standard Error	*T*-Statistic	Pr > |*t*|	Signif[b]	Std Coefficient	Share (%)[a]
1	IL	0.874	0.022	39.089	<0.0001	***	0.228	13.5
2	NE	1.540	0.048	32.145	<0.0001	***	0.215	12.8
3	IA	0.927	0.043	21.724	<0.0001	***	0.194	11.5
4	IN	1.526	0.107	14.267	<0.0001	***	0.181	10.7
5	KS	1.393	0.180	7.738	<0.0001	***	0.133	7.9
6	SD	1.066	0.031	34.041	<0.0001	***	0.097	5.8
7	MO	0.877	0.112	7.837	<0.0001	***	0.079	4.7
8	OTHER	0.668	0.040	16.525	<0.0001	***	0.076	4.5
9	PA	2.694	0.181	14.853	<0.0001	***	0.070	4.2
10	TX	1.202	0.121	9.944	<0.0001	***	0.061	3.6
11	NC	3.085	0.154	19.994	<0.0001	***	0.059	3.5
12	CO	−2.176	0.522	−4.167	0.0007	***	−0.053	3.1
13	MN	0.523	0.027	19.321	<0.0001	***	0.051	3.1
14	TN	−2.887	0.587	−4.920	0.0002	***	−0.049	2.9
15	ND	0.706	0.033	21.116	<0.0001	***	0.043	2.5
16	WI	0.784	0.099	7.912	<0.0001	***	0.037	2.2
17	MI	1.123	0.119	9.461	<0.0001	***	0.029	1.7
18	KY	0.488	0.231	2.111	0.0509	*	0.015	0.9
19	OH	0.178	0.149	1.193	0.2503		0.014	0.8
[Constant]		84.279	15.860	5.314	0.0001	***		

[a]Rank and share based upon absolute value of standardized coefficient.
[b]***Significantly different from zero at 99% confidence level, **95%, *90%.

4.2. Corn: Post-GM (1996–2017) Period

The four-fold CV procedure for corn in the post-GM period (1996–2017) resulted in an optimal CV-R^2 value of 97.3% with five components retained. The component regression coefficients are all positive and significant at the 99% confidence level.

Table 2 shows the individual states' coefficients (non-standardized and standardized) ranked by their absolute percent share of the total standardized values. Almost half (48.5%) of the total value is in the top four states (Illinois, Nebraska, Iowa, and Indiana). All of the coefficients are statistically significant at the 90% level or higher except for Ohio which is also the lowest ranked among the states. The HHI on the standardized coefficient shares is equal to 818 which is lower than the pre-GM period.

4.3. Soybeans: Pre-GM (1975–1995) Period

The four-fold CV procedure resulted in an optimal CV-R^2 of 98.3% with three correlated components retained. All three coefficients were positive and statistically significant at the 99% confidence level.

Table 3 shows the state non-standardized and standardized coefficient values along with the statistical significance and standardized share indicators. All of the coefficients were statistically significant at the 99% level with the exception of Michigan which was significant at the 95% level. In terms of standardized share, there is not much distance between the top four states (Missouri,

https://doi.org/10.1017/aae.2020.36 Published online by Cambridge University Press

Table 3. Individual states' coefficients and shares, soybeans, 1975–1995 time period

Rank[a]	State	Coefficient	Standard Error	*T*-Statistic	Pr > |*t*|	Signif[b]	Std Coefficient	Share (%)[a]
1	MO	1.156	0.048	24.106	<0.0001	***	0.140	9.8
2	IA	0.818	0.037	21.934	<0.0001	***	0.137	9.6
3	IL	0.743	0.024	31.443	<0.0001	***	0.127	8.9
4	MN	1.169	0.055	21.324	<0.0001	***	0.127	8.9
5	OTHER	0.705	0.067	10.458	<0.0001	***	0.113	7.9
6	OH	1.294	0.043	30.132	<0.0001	***	0.106	7.4
7	TN	2.358	0.248	9.523	<0.0001	***	0.102	7.1
8	MS	1.252	0.125	10.009	<0.0001	***	0.083	5.8
9	NE	1.583	0.087	18.150	<0.0001	***	0.081	5.6
10	LA	1.215	0.111	10.908	<0.0001	***	0.069	4.9
11	IN	0.806	0.039	20.479	<0.0001	***	0.066	4.6
12	SD	1.441	0.109	13.232	<0.0001	***	0.051	3.6
13	KS	1.077	0.089	12.128	<0.0001	***	0.048	3.3
14	KY	1.297	0.214	6.072	<0.0001	***	0.046	3.2
15	NC	1.511	0.146	10.375	<0.0001	***	0.042	2.9
16	AR	0.414	0.064	6.509	<0.0001	***	0.033	2.3
17	WI	1.652	0.141	11.707	<0.0001	***	0.030	2.1
18	ND	1.361	0.150	9.084	<0.0001	***	0.018	1.3
19	MI	0.651	0.272	2.392	0.0286	**	0.012	0.8
[Constant]		−5.211	6.382	−0.816	0.4255			

[a]Rank and share based upon absolute value of standardized coefficient.
[b]***Significantly different from zero at 99% confidence level, **95%, *90%.

Iowa, Illinois, and Other States) ranging from 9.8% down to 8.9%. There is also not much distance separating the second tier of states (Ohio, Tennessee, and Mississippi) ranging from 7.9% to 7.1%. The HHI value of 680 indicated a very low level of concentration among the standardized coefficient shares and was much lower than the value for corn (1,103) in the same time period.

4.4. Soybeans: Post-GM (1996–2017) Period

The four-fold CV procedure produced an optimal CV-R^2 of 98.9% with eight correlated components retained. All of the coefficient values were positive and the first six coefficients (CC_1 through CC_6) were statistically significant at the 99% level with CC_7 significant at the 95% level. The final component (CC_8) had a *P* value (0.107) just outside the 90% confidence range, but it was still retained due to the improvement in the out-of-sample CV-R^2 validation metric.

Table 4 shows the individual states' non-standardized and standardized coefficients along with indicators for the level of statistical significance and standardized share. All of the states' regression coefficients were significant at the 99% level with the exception of the Other States (90% significant), and Louisiana and Michigan (not significant). Wisconsin is notable in having a highly significant but negative regression coefficient. In terms of the share of total standardized coefficient values, Iowa stands out in the top spot with 15.9% followed by Minnesota (12.5%) and Illinois (12.3%). There was a big gap before reaching the next tier of states (Indiana, Kansas,

https://doi.org/10.1017/aae.2020.36 Published online by Cambridge University Press

Table 4. Individual states' coefficients and shares, soybeans, 1996–2017 time period

Rank[a]	State	Coefficient	Standard Error	*T*-Statistic	Pr > |*t*|	Signif[b]	Std Coefficient	Share (%)[a]
1	IA	1.361	0.075	18.110	<0.0001	***	0.259	15.9
2	MN	1.572	0.099	15.809	<0.0001	***	0.204	12.5
3	IL	1.090	0.039	28.071	<0.0001	***	0.200	12.3
4	IN	1.092	0.080	13.688	<0.0001	***	0.102	6.3
5	KS	1.038	0.061	17.014	<0.0001	***	0.094	5.8
6	OH	1.264	0.072	17.480	<0.0001	***	0.092	5.7
7	TN	1.903	0.205	9.290	<0.0001	***	0.079	4.9
8	MO	0.736	0.077	9.512	<0.0001	***	0.078	4.8
9	NE	0.967	0.066	14.691	<0.0001	***	0.077	4.7
10	SD	0.811	0.032	25.691	<0.0001	***	0.074	4.6
11	AR	1.490	0.223	6.691	<0.0001	***	0.072	4.4
12	ND	1.009	0.036	27.797	<0.0001	***	0.072	4.4
13	MS	1.685	0.162	10.401	<0.0001	***	0.069	4.3
14	WI	−1.560	0.214	−7.279	<0.0001	***	−0.059	3.6
15	KY	0.916	0.126	7.291	<0.0001	***	0.039	2.4
16	NC	0.915	0.151	6.072	<0.0001	***	0.026	1.6
17	OTHER	0.230	0.112	2.055	0.0605	*	0.014	0.8
18	LA	−0.259	0.202	−1.282	0.2224		−0.009	0.5
19	MI	−0.182	0.150	−1.219	0.2446		−0.006	0.4
[Constant]		10.460	3.401	3.076	0.0088	***		

[a]Rank and share based upon absolute value of standardized coefficient.
[b]***Significantly different from zero at 99% confidence level, **95%, *90%.

and Ohio) beginning at 6.3%. The HHI value on the standardized coefficients was equal to 838 and represented a slightly higher value when compared to corn (818) in the same time period.

5. Summary and Conclusions

Over the past 20 years, U.S. agriculture has witnessed profound changes with respect to technology, climate, farm policy, and other factors (ethanol production, Chinese demand, etc.) that have major repercussions with regard to the geographic distribution of crop production. There have been many recent studies that have examined both the direct and indirect impacts of these production factors upon crop yields, acreage, and production from both a temporal and spatial perspective. However, little to no attention has been paid to the impact of these factors upon the relative importance of each individual state's crop production outcome as it relates to the national outcome.

The purpose of this study was to address this question of state-level geographic importance for U.S. corn and soybeans by utilizing a novel new regression technique called CCR that is suitable for application to sparse and/or multicollinear data sets such as those examined in this study. A metric, called a PPI, was constructed that measured each state's and the national production performance by comparing the current year's value to the recent average levels (using an Olympic average). This metric was applied to USDA-NASS corn and soybean production history over two time periods:

https://doi.org/10.1017/aae.2020.36 Published online by Cambridge University Press

a pre-GM (1975–1995) period representing the years leading up to the commercialization of the first GM varieties of corn and soybeans in 1996, and a post-GM (1996–2017) period representing the years following commercialization.

A CCR regression model for each crop and time period was estimated with the national aggregate PPI as the dependent variable and the state-level PPI as the independent variables. The percent shares of the absolute value of the standardized coefficients from the regressions were used to rank each state and measure its relative importance. To measure concentration in each time period, a HHI was calculated on the standardized coefficient shares for each time period and compared across the time periods. The primary hypothesis introduced was that the introduction of new crop technologies (such as GM), along with observed climatic and policy changes, has resulted in a lower level of geographic concentration of production importance as measured by the HHI from the pre- to the post-GM time periods for each crop.

A summary of the standardized coefficient rankings between the two time periods (pre-GM and post-GM) is presented for both corn and soybeans in Table 5. Also, the HHI metric measuring the degree of concentration in the upper coefficient shares is also listed at the bottom of the table.

For corn, the level of concentration in the standardized coefficient shares declined moderately as evidenced by the 285-point decline in the HHI. In the pre-GM period, the top two states (Iowa and Illinois) held almost a 40% share with Iowa holding almost a 25% share by itself. In the post-GM period, nearly the same 40% share was held by the top three states (Illinois, Nebraska, and Iowa) with a near even distribution across the top four states (adding Indiana). States that made major jumps in the ranking between the two time periods included Nebraska (+14 places), Texas (+9), Pennsylvania (+6), and Colorado (+5). States with the largest declines in rankings were Michigan (−11), Minnesota (−10), Wisconsin (−8), and Ohio (−7). Nebraska had the largest increase in coefficient share (+11.8%) while Iowa had the largest decrease (−11.6%).

For soybeans, the level of concentration in the standardized coefficient shares increased slightly with a 158-point increase in the HHI. Note, however, that the post-GM HHI for soybeans is only slightly higher than corn as its pre-GM value was significantly less. In the pre-GM period, the coefficient shares are remarkably evenly distributed across a 3% range for the top seven states (Missouri, Iowa, Illinois, Minnesota, Other States, Ohio, and Tennessee). In the post-GM time frame, slightly over 40% of the shares are concentrated in just the top three states (Iowa, Minnesota, and Illinois). States that made major upward jumps in the rankings between the two time periods included Kansas (+8), Indiana (+7), North Dakota (+6), and Arkansas (+3). States with the largest declines were Other States (−12), Louisiana (−8), and Missouri (−7). The largest gain in standardized coefficient share was Iowa (+6.3%). The largest drop in coefficient share was Other States (−7.1%).

From these results, the following observations can be made. First, it is important to note that the rankings produced by the correlated correlation regressions do not mimic the average relative production shares of each state. A comparison of the rankings (regression versus average production shares) across the two crops and two periods produced correlation estimates (Kendall's tau) that ranged from 0.404 to 0.661 in value. This difference is because the CCR regressions measure both the direct and indirect (suppressor variable) effects of each state's PPI metric.

Second, the corn regression results exhibited a moderate reduction in concentration as measured by the standardized coefficient HHI. This indicates that geographic importance is more dispersed in the post-GMO period when compared to the pre-GMO. Notable is the strong increase in rankings for two states with significant areas under irrigation (Nebraska and Texas). This may be indicative of the importance of irrigation as a suppressive factor in times of extreme drought as was observed in Kukal and Irmak (2018). Also, it is also likely to be a side-effect of biotech (Bt) corn varieties expanding into areas where hard red winter wheat was once dominant.

Second, the soybean regression results exhibited a moderate increase in the standardized coefficient HHI which counters the initial hypothesis that technological, climate, and policy changes

https://doi.org/10.1017/aae.2020.36 Published online by Cambridge University Press

Table 5. Summary of state rankings by standardized coefficient share for both crops by time period

	Corn				Soybeans			
	1975–1995		1996–2017		1975–1995		1996–2017	
Rank	State	Share (%)[a]	State	Share (%)[a]	State	Share (%)[a]	State	Share (%)[a]
1	IA	23.1	IL	13.5	MO	9.8	IA	15.9
2	IL	16.5	NE	12.8	IA	9.6	MN	12.5
3	MN	7.6	IA	11.5	IL	8.9	IL	12.3
4	IN	7.2	IN	10.7	MN	8.9	IN	6.3
5	SD	5.6	KS	7.9	OTHER	7.9	KS	5.8
6	MI	5.6	SD	5.8	OH	7.4	OH	5.7
7	KS	5.1	MO	4.7	TN	7.1	TN	4.9
8	WI	4.9	OTHER	4.5	MS	5.8	MO	4.8
9	OTHER	4.8	PA	4.2	NE	5.6	NE	4.7
10	MO	3.6	TX	3.6	LA	4.9	SD	4.6
11	NC	3.2	NC	3.5	IN	4.6	AR	4.4
12	OH	3.1	CO	3.1	SD	3.6	ND	4.4
13	KY	2.5	MN	3.1	KS	3.3	MS	4.3
14	ND	2.0	TN	2.9	KY	3.2	WI	3.6
15	PA	2.0	ND	2.5	NC	2.9	KY	2.4
16	NE	1.0	WI	2.2	AR	2.3	NC	1.6
17	CO	0.9	MI	1.7	WI	2.1	OTHER	0.8
18	TN	0.8	KY	0.9	ND	1.3	LA	0.5
19	TX	0.4	OH	0.8	MI	0.8	MI	0.4
HHI[b]		1,103		818		680		838

[a]Percentage share of sum of standardized coefficient values (absolute value).
[b]Herfindahl-Hirschman Index applied to standardized coefficient shares.

have led to less concentration in the latter period. This is a somewhat puzzling result; however, it should be noted that the soybean post-GM HHI is nearly the same value as for corn. The increase for soybeans came from a much lower base in the pre-GMO period. Notable is that two of the top three states gaining in coefficient share rankings (North Dakota and Kansas) happen to be the largest producers of hard red spring and hard red winter wheat in the U.S. Minnesota (ranked #2 in post-GM rankings) is also a major producer of hard red spring wheat. Iowa, Illinois, and Indiana also were more significant soft red winter wheat producers in the pre-GMO period. The introduction of glyphosate resistant (Roundup Ready) soybeans has had a profound competitive impact upon wheat—probably more than any other major competing crop.

Another notable change in the soybean regression results was the increase in the number of indirect effects components (two versus seven) from the pre-GMO to post-GMO periods. The standardized share of the direct effects coefficient (CC_1) declined from 75.0% in the pre-GMO period to 59.7% in the post-GMO period. An examination of the component values in both periods indicates that the indirect effect coefficients are more correlated with fluctuations in the El Niño—Southern Oscillation cycle. For corn, a cursory examination of the component values indicates a greater sensitivity to temperature extremes—an indication that soybeans may be more susceptible to extreme moisture events while corn is more susceptible to extreme temperature events.

https://doi.org/10.1017/aae.2020.36 Published online by Cambridge University Press

Finally, a related observation is that the states that are ranked lower (for both corn and soybeans) tend to have more diverse cropping mixes when compared to the higher ranked states. For example, Michigan is ranked last for both periods for soybeans and has one of the most diverse crop production economies in the U.S. (outside of California). Wisconsin is also ranked low for both crops and has a very diverse crop production mix with a significant portion of corn production geared towards silage for its dairies. Another factor that may be driving these rankings relative to cropping mix is the dominance of the corn–soybean rotation in some states such as Iowa and Illinois. Soybeans provide advantages in a rotation as they are nitrogen fixing and are legumes which offer greater weed control management options.

These results provide significant information and a tool that can be used by USDA and industry economists in weighing geographic information during the growing season when making national crop production projections. For example, in predicting national corn production, reports of severe heat damage in Nebraska are more likely to have a severe impact upon the national aggregate when compared to flooded acres in Missouri. For soybeans, a localized drought in Iowa would have greater impact than if it occurred in North Dakota.

For those involved in risk management, the results indicate that technology can work both ways in terms of either spreading out or concentrating geographic production risk. The decline in geographic concentration for corn indicates that it would take a very widespread drought (such as in 2012) to have the same impact as more narrowly defined geographic drought in the past (such as in 1988). For soybeans, what was once a broad geographic distribution of risk has now narrowed with an event impacting three geographically close states (Iowa, Minnesota, and Illinois) having a significant impact upon the national soybean production.

In addition to the problem presented in this study, CCR can provide direct benefits over other estimation methods when applied to problems involving qualitative dependent variables or hazard rate modeling. This is due to the wider range of models available such as CCR-Logistic and CCR-Cox. Other variable reduction techniques, such as PCR and PLS, are basically limited to linear discriminant models. Therefore, there exists a wide range of potential research areas in agricultural economics, such as credit and insurance modeling, where CCR can provide major impacts. This is particularly true since data sparsity and multicollinearity can often arise when working with these types of data sets.

Additionally, a direct extension of this study would be to look at the correlated component latent variables themselves and determine what are the primary explanatory variables behind each component. These variables could include meteorological data (both national and regional), technological adoption data, farm policy indicators, costs of production measures, and many other categories of factors.

Acknowledgements. The author acknowledges Bryon Parman, William W. Wilson, and four anonymous reviewers for their helpful suggestions and comments in the preparation of this manuscript.

Funding Statement. This research received no specific grant from any funding agency, commercial, or not-for-profit sectors.

Conflict of Interest. The author declares none.

Data Availability Statement. All of the data used in this study comes from the USDA-NASS Quick Stats online database found at https://quickstats.nass.usda.gov.

References

Alkerwi, A., C. Vernier, N. Sauvageot, G.E. Crichton, and M.F. Elias. "Demographic and Socioeconomic Disparity in Nutrition: Application of a Novel Correlated Component Regression Approach." *BMJ Open* **5**,5(May 2015):e006814.

Andrews, D.W. "Tests for Parameter Instability and Structural Change with Unknown Change Point." *Econometrica* **61**,4(July 1993):821–56.

Bair, E., T. Hastie, D. Paul, and R. Tibshirani. "Prediction by Supervised Principal Components." *Journal of the American Statistical Association* **101**,473(March 2006):119–37.

https://doi.org/10.1017/aae.2020.36 Published online by Cambridge University Press

Bunge, J. "A Warming Climate Brings New Crops to Frigid Zones." *The Wall Street Journal*, November 25, 2018.

Burke, M., and K. Emerick. "Adaptation to Climate Change: Evidence from US Agriculture." *American Economic Journal: Economic Policy* **8**,3(August 2016):106–40.

Cai, R., J.D. Mullen, J.C. Bergstrom, W.D. Shurley, and M.E. Wetzstein. "Using a Climate Index to Measure Crop Yield Response." *Journal of Agricultural and Applied Economics* **45**,4(November 2013):719–37.

Chen, S., X. Chen, and J. Xu. "Impacts of Climate Change on Agriculture: Evidence from China." *Journal of Environmental Economics and Management* **76**(March 2016):105–24.

Deschenes, O., and M. Greenstone. "The Economic Impacts of Climate Change: Evidence from Agricultural Output and Random Fluctuations in Weather." *The American Economic Review* **97**,1(2007):354–85.

Du, X., C.L. Yu, D.A. Hennessy, and R. Miao. "Geography of Crop Yield Skewness." *Agricultural Economics* **46**,4(July 2015):463–73.

Fei, C.J., B.A. McCarl, and A.W. Thayer. "Estimating the Impacts of Climate Change and Potential Adaptation Strategies on Cereal Grains in the United States." *Frontiers in Ecology and Evolution* **5**(June 2017):1–12.

Garcia, P., S.E. Offutt, M. Pinar, and S.A. Changnon. "Corn Yield Behavior: Effects of Technological Advance and Weather Conditions." *Journal of Climate and Applied Meteorology* **26**(1987):1092–102.

Garver, M.S., and Z. Williams. "Improving the Validity of Theory Testing in Logistics Research Using Correlated Components Regression." *International Journal of Logistics Research and Applications* **21**,4(July 2018):363–77.

Haile, M.G., T. Wossen, K. Tesfaye, and J. von Braun. "Impact of Climate Change, Weather Extremes, and Price Risk on Global Food Supply." *Economics of Disasters and Climate Change* **1**,1(June 2017):55–75.

Hoerl, A.E., and R.W. Kennard. "Ridge Regression: Biased Estimation for Nonorthogonal Problems." *Technometrics* **12**,1(1970):55–67.

Huffman, W.E., Y. Jin, and Z. Xu. "The Economic Impacts of Technology and Climate Change: New Evidence from U.S. Corn Yields." *Agricultural Economics* **49**,4(July 2018):463–79.

Kaufmann, R.K., and S.E. Snell. "A Biophysical Model of Corn Yield: Integrating Climatic and Social Determinants." *American Journal of Agricultural Economics* **79**,1(February 1997):178–90.

Kennedy, P. *A Guide to Econometrics*. 4th ed. Cambridge, MA: The MIT Press, 1998.

Kuhn, M., and K. Johnson. *Applied Predictive Modeling*. New York, NY: Springer, 2013.

Kukal, M.S., and S. Irmak. "Climate-Driven Crop Yield and Yield Variability and Climate Change Impacts on the U.S. Great Plains Agricultural Production." *Scientific Reports* **8**,1(December 2018):1–18.

Li, Y., R. Miao, and M. Khanna. "Effects of Ethanol Plant Proximity and Crop Prices on Land-Use Change in the United States." *American Journal of Agricultural Economics* **101**,2(March 2019):467–91.

Lobell, D.B., M.J. Roberts, W. Schlenker, N. Braun, B.B. Little, R.M. Rejesus, and G.L. Hammer. "Greater Sensitivity to Drought Accompanies Maize Yield Increase in the U.S. Midwest." *Science* **344**,6183(May 2014):516–9.

Lobell, D.B., W. Schlenker, and J. Costa-Roberts. "Climate Trends and Global Crop Production Since 1980." *Science* **333**,6042(July 2011):616–20.

Magidson, J. "Correlated Component Regression: A Prediction/Classification Methodology for Possibly Many Features." *2010 JSM Proceedings of the American Statistical Association, Biometrics Section*, Vancouver, British Columbia, 2010.

Magidson, J., and K. Wassmann. "The Role of Proxy Genes in Predictive Models: An Application to Early Detection of Prostate Cancer." 2010 *JSM Proceedings of the American Statistical Association, Biometrics Section*, Vancouver, British Columbia, 2010.

Marshall, E., M. Aillery, S. Malcolm, and R. Williams. "Agricultural Production under Climate Change: The Potential Impacts of Shifting Regional Water Balances in the United States." *American Journal of Agricultural Economics* **97**,2(March 2015):568–88.

Massy, W. "Principal Components Regression in Exploratory Statistical Research." *Journal of the American Statistical Association* **60**(1965):234–46.

Mendelsohn, R., W.D. Nordhaus, and D. Shaw. "The Impact of Global Warming on Agriculture: A Ricardian Analysis." *The American Economic Review* **84**,4(1994):753–71.

Menz, K.M., and P. Pardey. "Technology and U.S. Corn Yields: Plateaus and Price Responsiveness." *American Journal of Agricultural Economics* **65**,3(August 1983):558–62.

Miao, R., M. Khanna, and H. Huang. "Responsiveness of Crop Yield and Acreage to Prices and Climate." *American Journal of Agricultural Economics* **98**,1(January 2016):191–211.

Quant, R. "Tests of the Hypothesis that a Linear Regression Obeys Two Different Regimes." *Journal of the American Statistical Association* **55**(1960):324–30.

Rosenzweig, C., and M.L. Parry. "Potential Impact of Climate Change on World Food Supply." *Nature* **367**,6459(January 1994):133–8.

Schlenker, W., and M.J. Roberts. "Nonlinear Temperature Effects Indicate Severe Damages to U.S. Crop Yields Under Climate Change." *Proceedings of the National Academy of Sciences* **106**,37(September 2009):15594–8.

Schlenker, W., W.M. Hanemann, and A.C. Fisher. "The Impact of Global Warming on U.S. Agriculture: An Econometric Analysis of Optimal Growing Conditions." *The Review of Economics and Statistics* **88**,1(February 2006):113–25.

https://doi.org/10.1017/aae.2020.36 Published online by Cambridge University Press

Tannura, M.A., S.H. Irwin, and D.L. Good. *Are Corn Trend Yields Increasing at a Faster Rate?* Marketing and Outlook Briefs No. MOBR 08-02. Champaign-Urbana, IL: Department of Agricultural and Consumer Economics, University of Illinois at Urbana-Champaign, 2008a.

Tannura, M.A., S.H. Irwin, and D.L. Good. *Weather, Technology, and Corn and Soybean Yields in the U.S. Corn Belt*. Marketing and Outlook Research Report No. 1. Champaign-Urbana, IL: Department of Agricultural and Consumer Economics, University of Illinois at Urbana-Champaign, 2008b. Internet site: http://www.ssrn.com/abstract=1147803 (Accessed October 5, 2018).

Thompson, L.M. "Weather and Technology in the Production of Corn in the U. S. Corn Belt." *Agronomy Journal* **61**,3(1969):453–6.

Thompson, L.M. "Weather and Technology in the Production of Soybeans in the Central United States." *Agronomy Journal* **62**,2(1970):232–6.

Thompson, L.M. "Climatic Change, Weather Variability, and Corn Production." *Agronomy Journal* **78**,4(1986):649–53.

Thompson, L.M. "Effects of Changes in Climate and Weather Variability on the Yields of Corn and Soybeans." *Journal of Production Agriculture* **1**,1(1988):20–27.

Tibshirani, R. "Regression Shrinkage and Selection Via the Lasso." *Journal of the Royal Statistical Society: Series B (Methodological)* **58**,1(January 1996):267–88.

Tolhurst, T.N., and A.P. Ker. "On Technological Change in Crop Yields." *American Journal of Agricultural Economics* **97**,1(January 2015):137–58.

Trivedi, J., and R. Birau. "Co-Movements Between Emerging and Developed Stock Markets in Terms of Global Financial Crisis." *1st WSEAS International Conference on Mathematics, Statistics, and Computer Engineering*. Dubrovnik, Croatia: WSEAS, 2013, pp. 146–51.

Wold, H. "Estimation of Principal Components and Related Models by Iterative Least Squares." In *Multivariate Analysis*. P.R. Krishnaiaah, ed. New York: Academic Press, 1966, pp. 391–420.

Cite this article: Bullock DW (2021). The Influence of State-Level Production Outcomes upon U.S. National Corn and Soybean Production: A Novel Application of Correlated Component Regression. *Journal of Agricultural and Applied Economics* **53**, 55–74. https://doi.org/10.1017/aae.2020.36