A Bio-Economic Assessment of the Spatial Dynamics of U.S. Corn Production and Yields

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Jason Michael Beddow

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Advisor: Philip G. Pardey, Co-Advisor: Terrance Hurley

February, 2012

Acknowledgements

The process of creating a dissertation is not unlike most other research projects—the ideas and work of colleagues are combined to create a report. What is unique about a dissertation is that only one person may claim authorship. Philip Pardey and I have engaged in countless conversations regarding the content of this dissertation and it would be difficult to determine which ideas were his and which were my own. I must give Professor Pardey equal credit for anything the reader might find useful in this work. At the same time, I am wholly responsible for any deficiencies since Prof. Pardey neither selected which topics would be addressed nor how I would go about the analysis. Terrance Hurley also contributed much to this work, and I especially appreciate his analytical insight, clarity of thought and ability to, as Einstein is said to have said, make things "… as simple as possible, but no simpler." The other members of my committee, Robert King and Stanley Wood, have also provided a great deal of advice, and their steadfast support of this work is deeply appreciated. Bridging the gap between economics and biology proved a difficult undertaking, and I benefited greatly from the sage advice of many, including Bob Sutherst, Darren Kriticos, Jawoo Koo and Robert Hijmans.

The research presented here required, at least by the standards of economists, vast amounts of data, a substantial portion of which was entered by hand from old census volumes. Michelle Hallaway spent many long hours entering and checking data, assisted by Toby Pardey. Louise Letnes provided substantial help in identifying and locating a number of fairly obscure articles, reports and datasets. I also thank Michelle and her family for sharing their knowledge of corn farming.

This work would not have been possible without the generous financial support of the Bill and Melinda Gates Foundation, Dr. and Mrs. William Hueg (via a Hueg-Harrison Fellowship) and the International Science and Technology Practice and Policy Center.

Finally, I thank my wife, Lucy Angelis, who sacrificed much so that I could pursue this work, and whose support and love were unwavering. This dissertation is dedicated to her.

Dedication

For Lucy.

Abstract

This dissertation reports on an investigation into the effects of location on corn production and productivity. The landscape of crop production is dynamic—where crops are produced changes dramatically over time. The answers to important questions about the potential impacts of global climate change and whether agriculture will be able to meet the world's increasing need for food are affected by the moving footprint of production. However, most studies of agricultural productivity and the effects of global warming do not consider that agriculture moves, and that the concomitant changes in natural services have important effects. A full set of county-level census data on corn production and area in the United States have been digitized and assembled for the first time, and new methods have been applied to account for changing geopolitical boundaries. Concepts adapted from economic index number theory are used to show that some 15 to 20 percent of the change in U.S. corn output over the past 130 years has come about due to shifts in where corn is produced. A newly developed, long-run, corn-specific weather dataset is used with the county data to show that, because of changes in the location of production, U.S. corn is now grown in cooler climates than it was a century ago, possibly offsetting some of the potential impacts of climate change. Finally, methods from ecological modeling, spatial econometrics, and crop modeling are combined to create a corn yield model that is then used to develop a location- and season-specific crop suitability indicator that takes into account the intra-seasonal dynamics of weather and the complex relationships between weather and yields. It will be shown that the suitability metric developed in this study gives results that are both consistent and more interpretable than more traditional approaches.

Table of Contents

List of Figures

List of Tables

Chapter 1 Introduction[1]

This dissertation reports on an investigation into the effects of location on agricultural production. Of primary concern is how changes in where production takes place—the geographical footprint of agriculture—affect agricultural output, inputs, and productivity. Newly digitized county-level data on U.S. corn[2] output and area spanning nearly thirteen decades, along with newly developed analytical and mapping techniques, make it possible to answer important questions that have hitherto been unanswerable. The first portion of the research, presented in Chapter 2, develops parsimonious, but powerful, methods for estimating the impacts of the changing spatial footprint of production on output. This procedure enables identification of the relative importance of changes in yield, area and the relative spatial allocation of production in determining a change in national output.

Chapter 3 continues the focus on location by considering an often misunderstood (and mischaracterized) attribute of location: weather. A long-run, county-level weather dataset is developed and utilized to investigate how the changing footprint of production has changed the effective climate in which corn is

---

[1] Portions of this section draw from Beddow et al. (2010a). I acknowledge and appreciate the contributions of the co-authors.

[2] Historically, the known grains were referred to generally as "corn," and the term "maize" was used in specific reference to *Zea mays*. For example, as reported in John Harris' 1748 volume *Voyages and Travels,* "Their Corn is of two sorts, English Wheat… and Maize, or Indian Corn (Fussel 1992, p. 19)." Nineteenth-century publications and the early U.S. Censuses of Agriculture referred to the crop as "Indian Corn," as was apparently common practice among Americans at the time. The terms do not cause confusion among modern speakers, but for the sake of consistency this document will adopt the American English convention, using "corn" in reference to maize except in direct quotations. In all cases, "corn" refers to field corn (which is used primarily for feed, industry and biofuels), as distinct from the horticultural varieties known as "sweet corn." Field and sweet corn are treated separately in U.S. agricultural statistics.

grown in the United States. It will be shown that the movements in corn production that occurred between the late 19th and early 21st centuries have profound implications for the potential effects of climate change on corn production. To get a better handle on the interaction between location and production, the historical use of weather in econometric yield models is investigated, and a practical method for introducing weather, and hence space, into productivity models is developed based on principles from ecological modeling.

A companion to this document is a Corn Atlas, which contains over 200 relevant maps and figures, the majority of which were not selected for inclusion in this thesis. The presentation in this document may be complemented by the Corn Atlas for readers who desire additional information on a particular topic and for those who wish to verify that the data used in this study are reasonable. In addition to static maps, figures and tables, a set of animated maps are also provided so that the reader may develop an intuitive sense of how variables have changed across space and time. The Corn Atlas files are available from the author on request, and shall also be available from the University of Minnesota's International Science and Technology Practice and Policy center (InSTePP), which has graciously agreed to permanently host the files on their website at www.instepp.umn.edu.

The dissertation will address three primary objectives, as discussed in the following sections.

Objective 1: Assess the Effect of Location on Agricultural Production

The notion that the agricultural output and productivity of a particular region tend to change over time is widely recognized, and economic productivity models often take account of those changes to gain insights into the temporal and inter-regional dynamics of productivity. However, many analysts do not recognize the more explicit spatial dynamics of production. Agriculture is an inherently local process, with yields (and hence output) being driven by local factors such as climate, soils, and terrain, as well as technology adoption and its use. Consequently, agricultural production and productivity are especially sensitive to spatial and inter-temporal variations in the natural factors of production. Precisely where crops are produced (the spatial footprint of agriculture) changes over time, with corresponding changes in the spatially heterogeneous factors that affect crop production. Giving more explicit attention to the spatial dimensions of agriculture and how they change over time deepens our understanding of the production and productivity performance of this sector.

Since movements in the footprint of agriculture necessarily imply underlying changes in the natural factors that drive productivity, one must take care when interpreting a difference in yield or land productivity,[3] namely by recognizing that phenomena exhibited by spatial aggregates are driven by local conditions. For example, a region's average corn yield will likely decrease if that region increases corn output by expanding into less productive areas. Although the region's average yield did indeed decrease, the underlying change in the agricultural landscape requires a more nuanced interpretation of the change than would be necessary if the geography

---

[3] "Yield" refers to the physical output of a single commodity per unit of land (per year) used in its production while "land productivity" generally has the output of an aggregation of two or more commodities (usually a summation of their value) as its numerator. Thus, yields are a specific type of partial factor productivity.

of production were constant.[4] Such examples shed light on the difficulty of specifying a counterfactual when assessing productivity differences, primarily because the analyst must ask whether a productivity metric is comparable across time or space when the underlying amounts and composition of natural and other inputs have changed, either due to spatial movements or other factors.

This thesis develops novel methods for assessing the effect of location on agricultural production. The effect of a change in the footprint of agricultural production is difficult to quantify since year-to-year changes in national weather, other (non-natural) inputs, and technological patterns can easily confound the effect of location. With some modification, widely used methods from economic index number theory are used to develop new ways to describe the portion of a national (or regional) output change that is attributable to changes in the spatial footprint of production. The metrics developed in this study provide direct answers to questions such as:

- How much of the increase in national corn output can be attributed to changes in the spatial allocation of production?
- How much would output have changed if the footprint of production had not changed?
- How much did changes in national yield versus area contribute to the observed change in national output?

---

[4] Olmstead and Rhode (2002) broached a similar idea with respect to U.S. wheat production (namely, regarding the results of Parker and Klein 1966), but they weren't focused on climate per se. A global-scale assessment of the spatial dynamics of crop production is provided by Beddow et al. (2010b).

Objective 2: Assess the Effect of Location on the Production Environment

While Objective 1 provides insight into the effects of location on production, the analysis is not explicit about the mechanisms through which location influences output. There are a variety of location-specific factors that influence agricultural yields. Adoption of technologies, availability and use of irrigation water, management regimes and weather differ from one location to another. This study puts particular focus on the influence of weather on yields. Climate change has received much attention (e.g., Mendelsohn, Nordhaus and Shaw 1994; Lobell and Asner 2003; Reilly et al. 2003; Lobell and Field 2007; Deschênes and Greenstone 2007; Schlenker and Roberts 2008; Massetti and Mendelsohn 2011); however most of this literature has neither explicitly nor clearly recognized that the changing footprint of production has altered the average growing climate of various crops. Thus, depending on the location or crop, the changing footprint may mute or amplify the crop productivity consequences of global warming.

Over the nearly thirteen decades for which county-level corn production data are now available for the United States, there has been a marked trend toward production in the cooler climes of the Northern Plains and the Lake States and away from the warm, humid, sub-tropical environment of the Southeastern United States. Newly developed, long-run, county-level weather data, combined with the long-run corn production data are used to show how the moving footprint of production has altered the effective growing climate of corn in the United States. As might be guessed, the average corn plant is now grown in a cooler climate than in the late 1800s.

Objective 3: Develop Methods for Including Weather in Productivity Models

Even if productivity models were to explicitly incorporate important local factors, it is unlikely that they could be modeled effectively using standard economic techniques. The indexing and production function approaches typically used in productivity assessment[5] are well-suited to account for inter-seasonal and cross-sectional variability in natural services and other factors of production; however the models temporally aggregate cropping seasons into discrete (calendar-year) units. The implicit assumption is that the intra-seasonal timing of input usage such as rainfall, fertilizers, and solar radiation is not important so long as the inputs are available during the cropping season. But the timing of events does matter: crops respond differently to temperature and rainfall at different stages of development (Neild 1982), daily light/dark cycles influence flowering (Warrington and Kanemasu 1983), and a hard freeze can result in crop loss, even if otherwise sufficient degree days are accumulated. The extent of temporal aggregation inherent in productivity models limits our ability to consider such dynamic phenomena and the myriad complex interactions among them.

Multifactor productivity (MFP) models sometimes include some correction for space by including one or more indicators of agroecological suitability (for example, see Chan-Kang et al. 1999). As they are typically implemented, these indicators present a number of methodological difficulties. Chief among these is the tendency to treat weather variables in ways that are inconsistent with the underlying biological process being modeled. For example, average temperature or rainfall over

---

[5] See Alston et al. (2010), pp. 101-103, for a short introduction to this literature.

an entire year (or season) is unlikely to provide a good econometric correction for weather since the crop's responses to these variables (and their interactions) are highly non-linear. Moreover, exactly how temperature and rainfall (and other natural inputs) are distributed throughout the crop's lifecycle have important implications for final yield.

The importance of the timing of intra-seasonal events has long been recognized. Wallace (1920) is an early attempt to empirically link weather to crop productivity, wherein the author considers the correlation between monthly temperature, rainfall and yield for a number of states.[6] A multitude of crop-yield studies that followed Wallace's work have tended to include temperature and rainfall in certain months as statistical corrections for weather. However, the months chosen are typically months during which a particularly important phenological phase of corn growth, silking, occurs in the Corn Belt states. Thus, these variables do account for some of the year-to-year variability in corn yields, but only because they happen to coincide with important stages in the biological growth process where a large share of the nation's corn is grown. There is no reason to suspect that their inclusion in a model should provide meaningful information for all corn growing locations, or for all years. Insofar as the timing of phenological stages differs across locations and by season, a rainfall or temperature variable for a given month will have an unstable relationship with final yield.

One solution is to use one of the various crop models (most of which are descended from the CERES Maize model) to assess the relationship between weather

---

[6] The study by Wallace (1920) and a number of other efforts to empirically link yield with weather (and other) inputs are presented in more detail in Chapter 3.

and location. However, this approach is not tractable for application to a long-run (century plus), national set of locations due to the location specificity and data requirements of the models. The present study develops a practicable approach for deriving a location-specific, bio-climatic weather variable across a large panel of data by:

- developing relatively simple, yet accurate, methods for predicting when the various phenological stages will occur in a given location-year, and
- providing a set of methods for aggregating the effects of weather into a time- and location-specific metric of crop suitability.

The final section of Chapter 3 presents an econometric assessment of the ability of the newly developed weather metric to address weather effects. It is neither expected nor required that the weather metric should account for all differences in yield across space and time. Rather, the success of the model shall be assessed based on its ability to provide a statistically significant and readily-interpretable measure of the effects of weather relative to the measures developed in previous work. It will be shown that a simple, dynamic, and location-specific crop suitability metric can yield results that are more interpretable than achieved using other approaches.

Chapter 2 A Spatial Look at Corn in the United States

2.1 *National Area and Output Over Time*

In 2010, 81 million acres of corn were grown in the United States, around 20 percent of the corn area harvested worldwide that year (accounting for about 37 percent of global corn output). The 2010 U.S. corn crop occupied an area larger than the entire land mass of Poland and, if the 2010 corn crop had been grown in a single row, the length of that row would best be measured in astronomical units (the average distance from the earth to the sun)—the single corn row would extend nearly three astronomical units, a distance that would take the Concorde over 23 years to complete at its cruising speed of mach two.[7] But corn is neither grown in a single row nor in a contiguous Poland-sized area: its production is spread throughout the country, and the particular way in which it is distributed has important consequences for the quality and quantity of grain that is produced.

The known history of corn is a history of cultivation. Landrace (naked) varieties of other major cereal grains are well-known but, while some have raised conjectures about the natural ancestors of modern corn, the "…wild strains of maize…no longer exist or have not yet been discovered" (Vavilov 1992, p. 398). Candolle (1884 p. 394) supposed that wild corn might have existed at the higher altitudes in Chile and Mexico. Candolle describes some early assertions that corn originated in the global East, but later shows these assertions to be at best mistaken, and at worst "complete forgeries" (p. 17). Nevertheless, there is no longer

---

[7] Assuming a 30 inch row spacing.

disagreement among scholars of the subject that corn originated in the Americas, that it has been under cultivation during the entire tenure of agriculture in the New World, and was likely under cultivation in Southern Mexico at least as long ago as 7,000 BCE (Schery 1972, p. 411).[8]

There is little statistical information on the area devoted to corn or the production of that area before the mid-19th century. Kirkby (1973) used the relationship between cob length and yield, along with archeological evidence, to estimate early yield changes in the Oaxaca Valley in southern Mexico. He surmised that yields in the valley increased from about four or five bushels per acre in 1300 BCE to about 10-18 bushels per acre by 900 CE. By some estimates, pre-Columbian native yields in the United States were around 20 bushels per acre and there is some evidence that Colonial yields in the Northeast were sometimes higher. For example, Douglas (1760) reported that "…good land in Maryland and Virginia may yield per acre…30 bushels [of] Indian corn (p. 375)." There is, however, no doubt that corn was widely grown in the Americas before the arrival of Europeans, with production extending from southern Mexico to the Northeastern United States, where Europeans first encountered the crop (Schery 1972, pp. 412-413).

---

[8] The debate over the origins of corn is much more nuanced than has been described here. Wilkes (2004) provides a fairly recent history of the debate. Much conjecture has related to the relationship between corn and wild members of the species *Zea mays,* known generically as "tesointe." As summarized by Doebley (2001), two competing theories arose: Mangelsdorf and colleagues proposed that corn had descended from an extinct wild maize while Beadle proposed that corn is a domesticated form of teosinte. Doebley concludes that the evidence from molecular genetics is consistent with Beadle's teosinte hypothesis and offers no support for the Mangelsdorf hypothesis. Diamond (2010) assesses how and why crops spread from their centers of origin (see especially Ch. 10).

The decadenal U.S. census of 1840 included the first nationwide enumeration of corn production, recording 378 million bushels produced among the 29 states and territories that were surveyed. The census of 1850 recorded 592 million bushels of corn production, an increase of over 57 percent from the previous census. Several new states and territories were included in the 1850 census, but inspection of the state data reveals that the added areas explain at most three percent of the increase. Rather, production in several Midwestern states increased markedly, with Illinois and Michigan more than doubling output and Wisconsin and Iowa more than quintupling production.

The following decade saw a similar absolute increase in production, with the 1860 census recording a total production of 839 million bushels. As in the previous census, several states and territories were added to the enumeration in 1860 (notably, Kansas and Nebraska), with the new areas again accounting for about three percent of the increase in national production. Over half of the increased output derived from increases in Iowa, Illinois and Missouri. The decade also witnessed the first recorded decreases in state production, with the New England states collectively decreasing their output by about ten percent and South Carolina dropping over seven percent. As recorded in the 1870 census, 1869 production decreased by almost 78 million bushels relative to the level of production recorded in the 1860 census. However, some states, notably Illinois, Iowa and Kansas, markedly increased production while this was offset by large reported decreases in other states, especially, Indiana, Kentucky, Virginia and Alabama (with the latter two about halving their production, likely due to residual effects from the Civil War).[9]

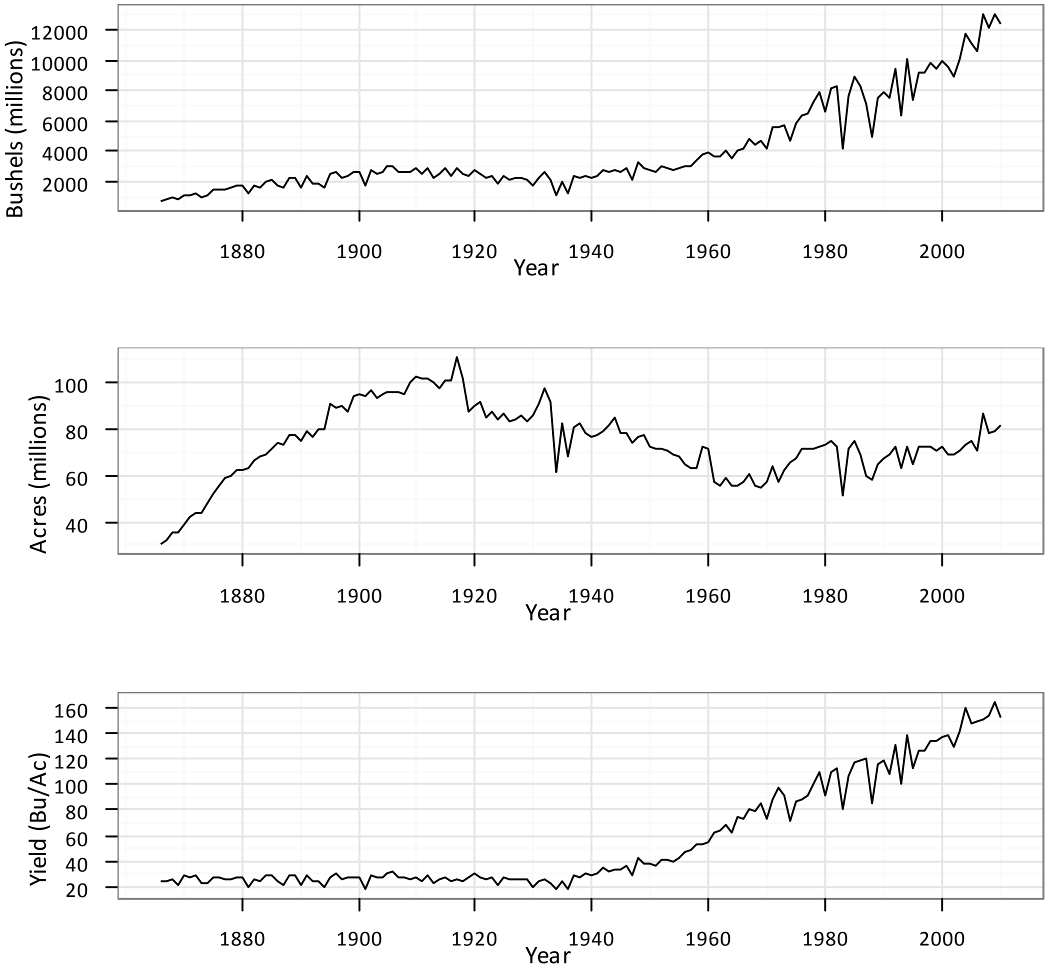
The National Agricultural Statistics Service (NASS) of the USDA provides national estimates of corn area and production stretching back to 1866. Thus, the beginning of this series marks the first year for which reliable, survey-based estimates of national yield can be calculated. Figure 1 shows the production (upper panel), harvested acreage (middle panel) and yield (lower panel) of corn in the United States between 1866 and 2010. Corn output increased dramatically over the 145 year period, from about 730 million bushels in 1866 to over 12.4 billion bushels in 2010, a 17-fold increase. Over the same period, harvested acreage increased less than three-fold, from about 30 million acres to 81.4 million acres, while the average yield increased over six-fold (from 24.3 to 152.8 bushels per acre). This implies that much of the increased output is attributable to increased corn yield, and was not simply the result of expanded acreage. Indeed, over the entire period, harvested acreage grew at an average annual rate of less than 0.7 percent while yield increased by just under 1.3 percent per year.[10]

---

[9] The 1869 reported production of the former slave-holding states is questionable. The authors of the Ninth Census noted: "[t]he plantations of the old slave States are squatted all over by the former slaves, who hold small portions of the soil, often very loosely determined as to extent, under almost all varieties of tenure." Apparently this situation rendered it difficult to obtain a complete enumeration and although efforts were made to establish procedures that would produce an accurate accounting of the production of the South, "…after a weary and unprofitable struggle, the Superintendent was fain to accept whatever could be obtained in regard to the agriculture of that region…[t]o have insisted upon a logical treatment of the subject would have been equivalent to giving up the agricultural statistics of the year (Ninth Census, 1870, Industry and Wealth, p. 72)."

[10] Unless noted otherwise, growth rates for a period are calculated as the log difference of the values of the endpoints of the period divided by the number of years in the period. See section A.1 for a justification of this approach.

Figure 1: U.S. Production, Area and Yield



Source: Created using data from USDA-NASS (2011).

Corn output has not closely followed the acreage trends, especially since the end of the First World War. Indeed, total output and acreage are negatively correlated during the period beginning in 1919 (ρ = -0.15 [-0.34, 0.06]),[11] while yield and output show a near-perfect linear correlation (ρ = 0.98 [0.97, 0.99]). This does not necessarily imply that changes in acreage are unimportant, but rather that it is crucial

---

[11] Throughout this paper, "correlation", denoted "ρ", refers to Pearson's product-moment correlation unless noted otherwise. The square bracketed range (when provided) indicates a 95 percent confidence interval assuming, generally without assessment, that the data vectors represent normally distributed data.

to consider exactly *which acres* were taken into and out of production—for example, national output could be increased by taking lower-yielding acres out of production and adding back (fewer) high-yielding acres. This phenomenon will be explored later in this thesis, but the fact that acreage does not show a significant linear correlation with output provides some evidence that such a scenario may hold true.

The rate of growth in corn acreage varied markedly over time. Corn acreage increased rapidly through the end of World War I, reaching an all-time high of nearly 111 million acres during the 1917 crop year. The harvested acreage decreased sharply after the war, with 14.7 million acres lost between the 1918 and 1919 crop years. This began a general declining trend in corn acreage that continued for nearly half a century. Corn acreage began a marked ten-year increase in 1970 and, after sharp acreage drops in the 1980s, acreage began the current slowly increasing trend.

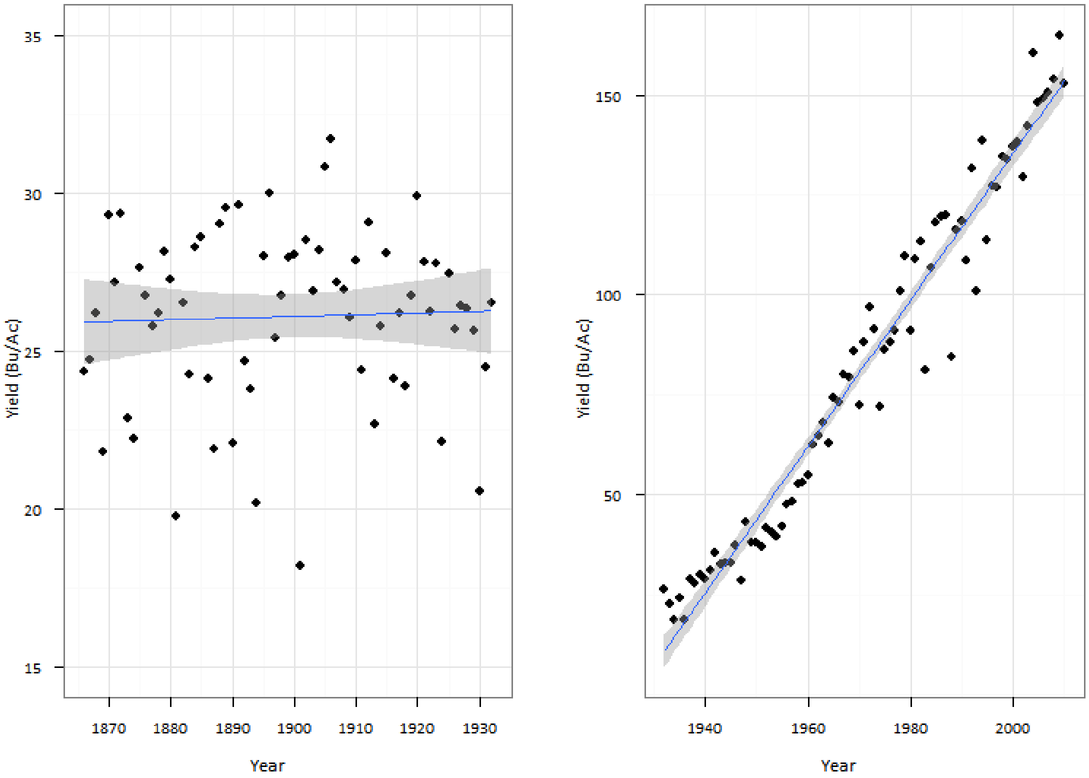
Inspection of the bottom panel of Figure 1 reveals at least two distinct periods of yield increase. Bray and Watkins (1964) determined that corn yields were constant over the period 1870-1925, began decreasing from 1925 through 1937 then rapidly increased from 1937 through 1960. The breakpoints found by Bray and Watkins are based on a five-year moving average of the yield series and could not be replicated using optimal breakpoint algorithms.[12] The best linear fit to the 1870-1960 yield data is obtained with a single breakpoint at 1932; the yield growth rate in the earlier period was near zero while the annual growth rate afterward averaged 3.2 percent (as determined using a semi-log regression).

---

[12] The determination of the number and location of optimal breakpoints uses the implementation described by Zeileis et al. (2003) of the methods developed by Bai and Perron (1998).

The best linear fit to the yield data over the entire period from 1866 to 2010 is obtained with two breakpoints, one at 1944 and another at 1987. The initial period, from 1866 through 1944, saw an average yield growth rate of less than 0.1 percent per year while the growth rate was 3.1 percent per year during 1945-1987 and 2.0 percent per year from 1988 through 2010. Visually, there is one clear breakpoint, optimally located at 1943 as shown in Figure 2, where the line represents a locally-weighted regression (loess) and the shaded area represents a 95 percent confidence interval around that line.

Figure 2: U.S. Corn Yield, 1866-1942 and 1943-2010



Source: Created using data from NASS (2011).

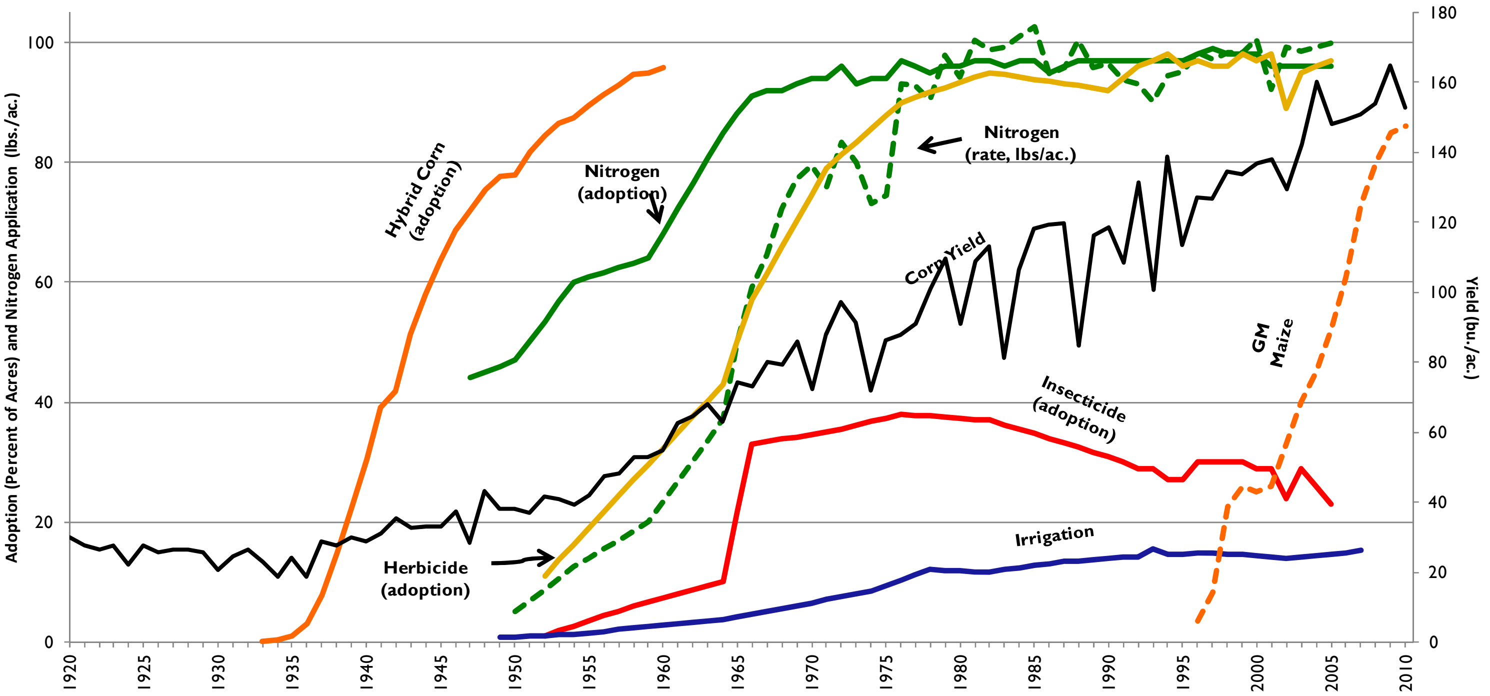
Notes: The blue line is derived from a loess regression and the grey area represents a 95 percent confidence interval about that line.

The increased yield growth rates beginning in the late 1930s are often attributed to the widespread adoption of yield-enhancing technology, especially adoption of both hybrid maize, and a corresponding increase in nitrogen fertilizer use (Brown 1967). In reference to Brown's conclusions, Evans (1993, p. 273) notes that "…the initial rise in yield coincided with the adoption of hybrids, while the ensuing faster rise more or less coincided with the greater use of fertilizer on maize crops." Figure 3 shows U.S. national corn yield along with the adoption of various technologies over time. Hybrid maize adoption does indeed appear to correlate with the increase in national yield that began during the late 1930s.

Nitrogen fertilizer application rates (shown in pounds per acre) increased quickly from the 1950s through 1972 coinciding with larger year-to-year increases in corn yield over the period (shown in bushels per acre on the right axis). The 1973-1974 oil embargo was associated with increased fertilizer prices, and by 1974 nitrogen application had decreased by nearly ten pounds per acre (over 12 percent) relative to the 1972 crop year, with corresponding decreases in national yield. The trend was quickly restored after the oil crisis largely due to an average 18.5 pound per acre increase in application rates between 1975 and 1976. The highest ever national average fertilizer application rate, 102.5 pounds per acre, was achieved in 1985, and fertilizer application rates have been largely stagnant since. From both an economic and a biological standpoint, there is little question that increases in fertilizer application rates resulted in increased yields.

It is difficult to assess the extent to which increased national yields can be attributed to adoption of hybrid varieties and increased fertilizer use per se. Certainly, national fertilizer application rates and yield are highly correlated ($\rho = 0.88$ [0.81, 0.93]), but fertilizer application may be associated with more intensive, highly-

Figure 3: Corn Technology Adoption



Source: Data for nitrogen use on corn acres for 1947-1959 are from Ibach and Adams (1967), except for 1950 which is from Mehring et al. (1957) and 1959 which is from Ibach et al. (1964). Nitrogen use data for 1964-2005 are from USDA-ERS (2011b). Herbicide data for 1952 and 1958 are from Eichers et al. (1978) and values for 1993-2005 are from USDA-NASS (2011a). All other herbicide use data are from Lin et al. (1995). Irrigation data for 1969-1994 are from USDA-ERS (1994) and data for 1997-2007 are from USDA-NASS (2011b). Insecticide use data for 1952 and 1958 are from Eichers et al. (1978) and from Lin et al. (1995) for other years. See Alston et al. (2010a, p. 253) for information on hybrid maize and GM maize adoption.

Notes: Yield is plotted on the right axis; all other variables are plotted on the left. Nitrogen use is in pounds per harvested acre of corn and the adoption variables represent the percentage of harvested corn acres on which the technology was adopted in each year.

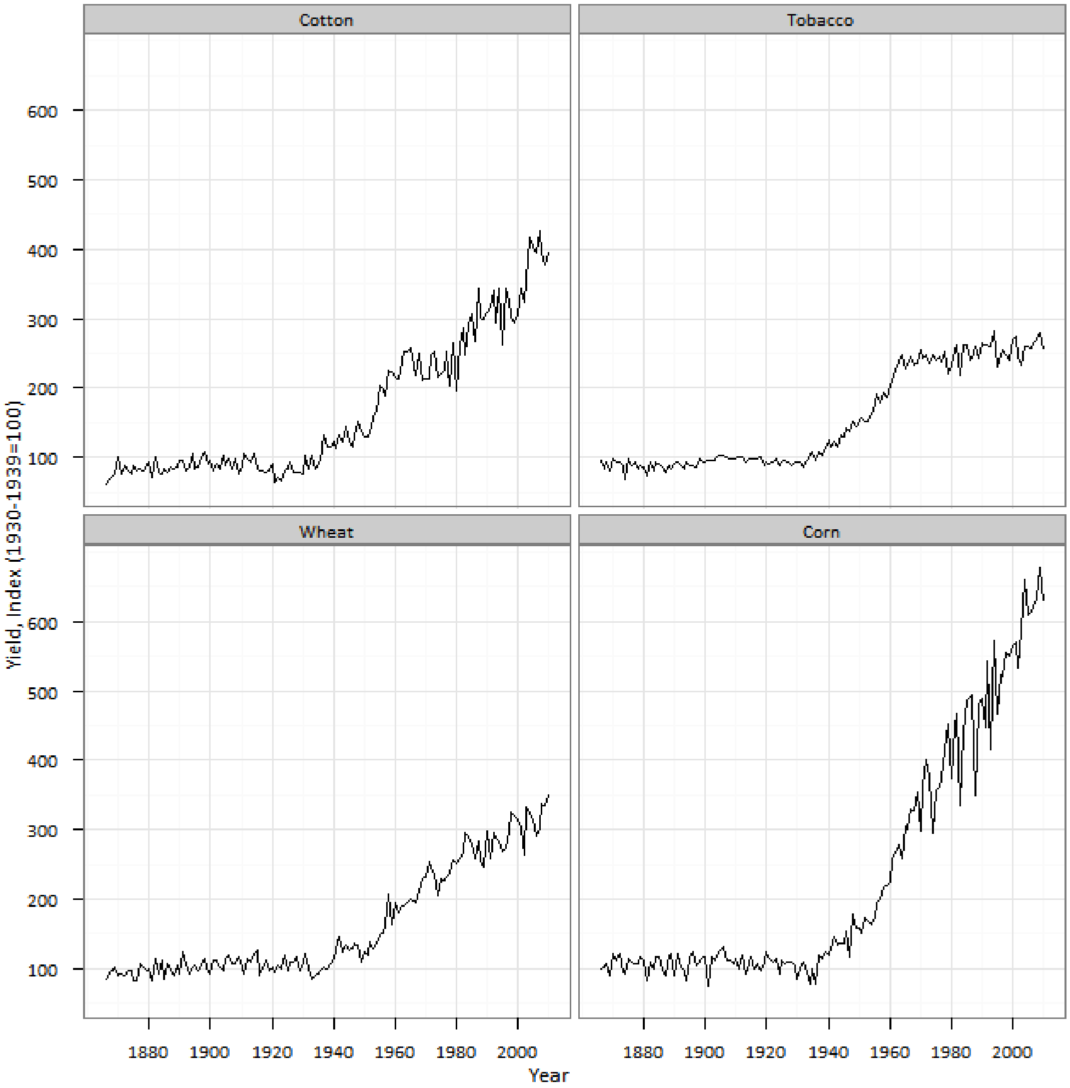
managed production in general. For example, Figure 3 (above) shows herbicide adoption increasing in step with nitrogen application rates, along with large increases in insecticide adoption in the mid-1950s.

Figure 4 (p. 19) shows the yield of corn, tobacco, cotton and wheat in the United States over the period 1866-2010. So that the series are comparable, yield values for each crop are indexed relative to their mean value over a ten year period beginning in 1930. The crops all appear to have yield take-off points in the late 1930s,[13] which implies that at least some of the corn yield increases may have been due to general improvements in farming methods since the 1930s (Evans 1993, p. 273) rather than changes in corn-specific technology. Others point to a general trend towards increased use of commercial fertilizers to explain the similarity of the yield trends of different crops (e.g., Sutch 2008, p. 6). As outlined by Warren (1998), wheat yields benefitted from the introduction of insect- and disease-resistant varieties and more effective herbicides (especially 2,4-D) beginning in the 1940s, while cotton yields increased markedly due to the introduction of more effective insecticides after World War II and the introduction of cultivars that directly increased yields. The national yield is also influenced by policy-induced spatial reallocation, for example Agricultural Adjustment Act (AAA) and related policies that tended to idle less-productive cotton (Luttrell and Gilbert 1976) and tobacco acres (USDA-BAE 1948, after Sutch 2008).

---

[13] For U.S. wheat, Calderini and Slafer (1998) found a single breakpoint in 1936 to be optimal.

Figure 4: U.S. National Yield of Cotton, Tobacco, Corn and Wheat, 1866-2010



Source: Created using data from USDA-NASS (2011b).

Notes: The yield of each crop is indexed so that its average national yield over the period 1930-1939 equals 100.

2.2 *Sub-National Area, Output and Yield*

Robinson (1950) famously pointed out that spatial aggregation matters when statistical methods are applied, noting that correlations between summary statistics taken across large aggregated regions will be higher than correlations taken across smaller regions or individuals. This was not, however, a new insight, for example Gehlke and Biehl (1934) reported a similar result in a study comparing median rental costs to juvenile delinquency: using various levels of spatial aggregation, the authors found that simple correlations between areas were stronger as the area of aggregation became larger. However, they also noted that grouping areas randomly had little effect on the calculated correlations. Openshaw and Taylor (1979) coined the term "modifiable areal unit problem" (MAUP) to refer to spatial aggregation issues such as Robinson's "ecological fallacy." The term is used more generally to refer to the notion that, insofar as geopolitical boundaries are arbitrary with respect to a variable of interest, any arbitrary aggregation across such boundaries will be equally (in)valid (a zoning effect). The problem arises because any aggregation will necessarily result in a loss of information (a scale effect) (Reynolds 1998). For example, Jelinski and Wu (1996) investigated both scale and zoning effects by considering various aggregations of Landsat raster data, finding a notable scale effect on spatial autocorrelation.

The term "modifiable" is intended to contrast with the atomic (individual) level at which data generating processes operate. In the present case, the "individual" is properly a corn plant. Any aggregation, even to the farm- or field-level, will induce MAU problems by definition insofar as the zonal boundaries are arbitrary with respect to the variable(s) of interest. However, data and tractability concerns dictate that large-scale plant-level productivity assessments are not feasible, and one must determine the most appropriate spatial units to employ in a given analysis. As implied by notions of the MAU problem and the ecological fallacy, lower units of aggregation are generally preferable to larger units if one is concerned with the processes underlying aggregate phenomena. The present goal is to investigate the long-run history of U.S. corn production, so it is prudent to explore the availability and quality of county-level data since counties are the smallest spatial unit upon which U.S. agricultural data are consistently reported.

2.2.1 A Consistent Area and Production Dataset for Corn

This section briefly describes the production (output and area) dataset used in this assessment. A more detailed accounting of the data sources along with the methods and procedures used to generate the dataset are included in Section A.2. That section also includes more information on the assumptions embedded in the resulting data and various caveats relevant to the interpretation of the data.

There are two primary sources of long-run, county-level corn production and area data for the United States: those assembled by the USDA's National Agricultural Statistics Service (NASS), and the data presented in the various agricultural censuses. The census data have the advantage of presenting a near-complete enumeration of U.S. corn production and area for each census year, leaving out only values that might reveal information about individual enterprises or individuals. The NASS data are provided on an annual basis, but the sampling approach used by NASS means that the geographical coverage of the data varies from year to year. Choosing a dataset to employ in the analysis of corn production required making a tradeoff between the spatial completeness of the census data and the temporal richness of the NASS data.

While annual data can enable rich time-series analyses, the present concern is with spatial change, the analysis of which is made substantially more difficult if spatial coverage is incomplete. Thus, the agricultural censuses were selected as the primary source of area and production data used in this study.

The various agricultural censuses together provide 23 years of data, one for each census enumeration, and stretch back to the 1879 crop year, the first year in which both production and area data were compiled. Despite the richness and unquestionable utility of these data, no previous effort had been made to develop a unified electronic compilation of the census data. To enable the present analysis, many of these data were entered by hand from census books, and other data were extracted from somewhat disparate electronic files in various formats.[14] Further, a great deal of effort was required to explicitly georeference the county data to shapefiles representing the county boundaries in each year. Because of these efforts, this thesis is the first known study to employ all available county-level census data for any crop, and development of the dataset itself is a noteworthy contribution of this work.

Long-run assessments using county-level data are complicated by changes in county boundaries over time. Individual counties often split to form multiple counties, but there are also numerous cases of multiple counties combining, and of boundaries shifting in seemingly arbitrary ways. For example, consider the county maps of the Dakotas presented in Figure 5. The left panel of the figure shows the

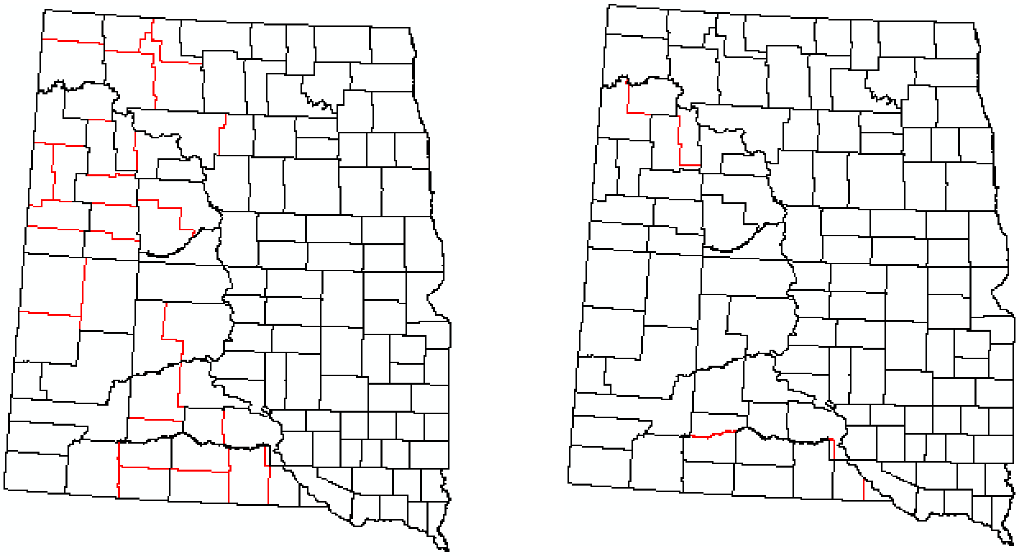
---

[14] While mentioned in the acknowledgements, the contributions of others to the development of the corn dataset cannot be overstated. Michelle Hallaway, Toby Pardey and others spent many tedious hours entering (and checking) data. Cornell's Mann Library and the University of Minnesota's National Historical GIS (NHGIS) program generously provided access to various electronic files.

counties with their boundaries at the beginning of the 20th Century in black, and the right panel shows the boundaries as they were 100 years later. In the left panel, boundaries added since 1900 are indicated in red, while red boundaries in the right panel indicate boundaries that were removed. Inspection of the figure reveals that several new counties were formed from existing counties over the century, while several other counties combined. Other states exhibited similar patterns.

Figure 5: The Dakotas, 1900 and 2000

North and South Dakota, 1900 North and South Dakota, 2000



Source: Created using shapefiles as described in Section A.2.

Notes: In the map for 1900 (left), boundaries added by 2000 are in red. In the map for 2000 (right), boundaries deleted since 1900 are in red.

Given the boundary changes, it would be difficult to use the county-level dataset directly. A great deal of empirical overhead would be required to keep track of the size, shape and location of each county over time, and, for example, it would be difficult to determine the meaning of changes in output, area and yield when the basic spatial units of analysis are themselves changing. One way to attack the problem would be to assume that the crop is distributed evenly across the county. The data could then be re-aggregated to a consistent set of boundaries by allocating the production and area of the old area to new areas according to the share of the new areal units that is also in the old units.

While the above approach would provide a workable disaggregation, such a procedure could bias the results by inducing (re-)zoning effects. For example, consider a case in which a year-1900 county splits such that in 2000 its entire area is composed of two counties: county A and county B. If most of the year-1900 production of this area occurred within the extent of county A, production and area estimates would be biased downward for county A and upward for county B. To remedy this problem, additional information on the location of production within each county was used.

Information from the HarvestChoice Spatial Allocation Model (SpAM)[15] was used to determine which parts of each county were most likely to produce corn, and, whenever possible, the county data were spatially disaggregated accordingly. SpAM provides estimates of the spatial distribution of production for the year 2000, and it is assumed that the relative spatial allocation *within* each county-year is the same as that implied by the SpAM model (but the model was *not* used to add information about the spatial allocation across counties). All county data were first mapped to the county boundaries for the closest available year, and the counties were divided into arrays of five arc-minute pixels. Thus, the data for each year were converted from aereal (county) to raster (pixellated) data, allocating the county's production and area to each pixel in proportion to the pixel's share of the county's area or production in

---

[15] You, Wood and Wood-Sichra (2006) describe the methods used to produce the SpAM estimates. This analysis uses version 3, release 2 of the data from You et al. (2011).

the SpAM results. Finally, the raster data were aggregated according to year-2000 county boundaries, resulting in a panel of output and area data, all mapped to year-2000 county boundaries. Using this procedure, the U.S. county-level corn production and area data are now assembled in a consistent panel format, where each spatial unit represents a consistent entity over the entire 128 year period for which data are available.[16]

It is almost certain that previous studies have limited their temporal or geographic coverage because the areal units were not comparable over long time periods. The re-aggregation approach developed here produces, for the first time, a long-run county panel dataset that corrects for the otherwise confounding problems associated with changing county boundaries. As will be seen, the resulting consistent (but unbalanced) panel will enable some analyses that otherwise would not be possible, and will greatly simplify others.

2.2.2 Moving Production—Regional and County Patterns and Trends

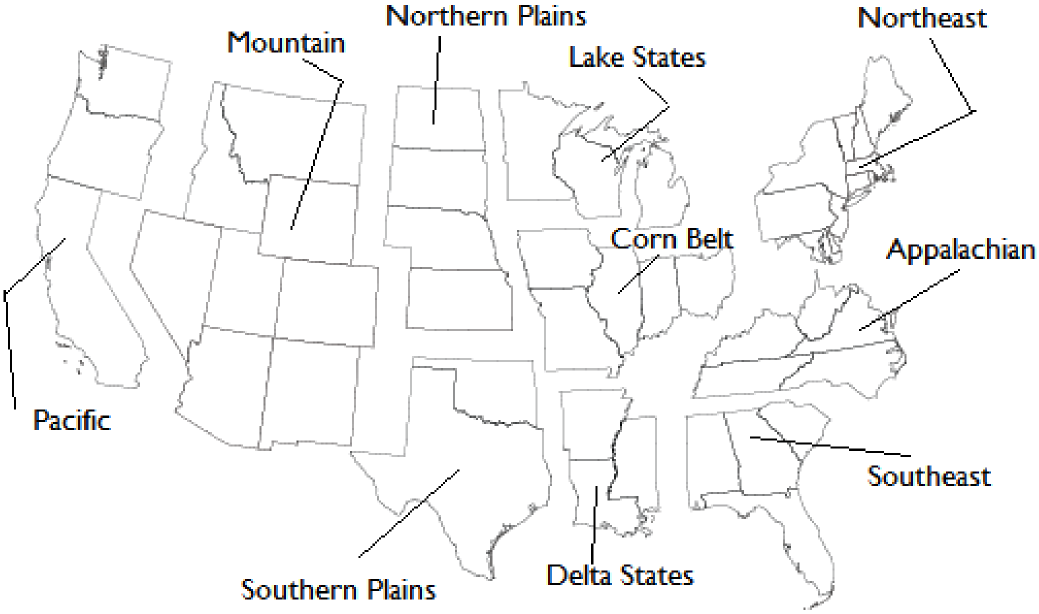
The footprint of corn area changed dramatically during the thirteen decades beginning in 1879. At both the beginning and end of the period, the Corn Belt states of Indiana, Ohio, Illinois, Iowa and Missouri combined accounted for a little less than half of overall U.S. corn acreage (45 percent in 1879 and 47 percent in 2007). While the relative importance of the Corn Belt remained more-or-less stable, there were dramatic shifts away from production in the Southeast and Appalachian regions and toward production in the Northern Plains and Lake States (see Figure 6 for region

---

[16] The reader is again reminded that a more detailed accounting of the methods and procedures used to develop this dataset is provided in the Appendix (Section A.2).

definitions). In 1879, the Appalachian and Southeast regions accounted for about 27 percent of U.S. corn acres, but this share had decreased to just over five percent by 2007; over the same period, the Northern Plains and Lake States increased from a combined share of less than 10 percent of corn acreage to over 38 percent.

Figure 6: U.S. Crop Production Regions



Source: Recreated based on USDA (1998, p. 18).
Notes: A listing of the states in each region is provided in the appendix (Table 12, p. 170).

National acreage increased from 62 million acres in 1879 to 86 million acres in 2007, an increase of 24 million acres. This represents the combined effect of a loss of over 12 million corn acres in the Southeast and Appalachian regions and a 25.5 million acre increase in the Northern Plains and Lake States, netting about half of the additional corn acres. Most of the remaining new acres are accounted for by a 12 million acre increase in the Corn Belt.

The 2007 census reported a dramatic 18 million acre increase in corn harvested area over that reported in the previous census (2002), accounting for threequarters of the net increase in harvested acres since 1879. Between 2002 and 2007, all regions posted increased corn acreage, and there was a small shift away from production in the Corn Belt and Lake States. While the Corn Belt's *share* of production slightly decreased over this five-year period, the region saw a near seven million acre increase in corn production, accounting for almost 60 percent of that region's total increase in output since 1879. This dramatic change occurred in just half a decade, underscoring the ability of farmers to quickly adapt to changing market and environmental conditions, rapidly changing the footprint of production in the process.[17]

Between 1879 and 2007, national corn output increased by nearly 11 billion bushels, from 1.7 billion bushels in 1879 to 12.7 billion bushels in 2007. Production increased in all regions of the country, but the majority (89 percent) of the increased production was in the Corn Belt (50 percent), Northern Plains (23 percent) and Lake States (16 percent). By contrast, the Southeast and Appalachian regions saw their share of national production decrease from over 16 percent to under four percent. The Delta States maintained around a three percent share of national production, and now account for about the same amount of production as the Appalachian states.

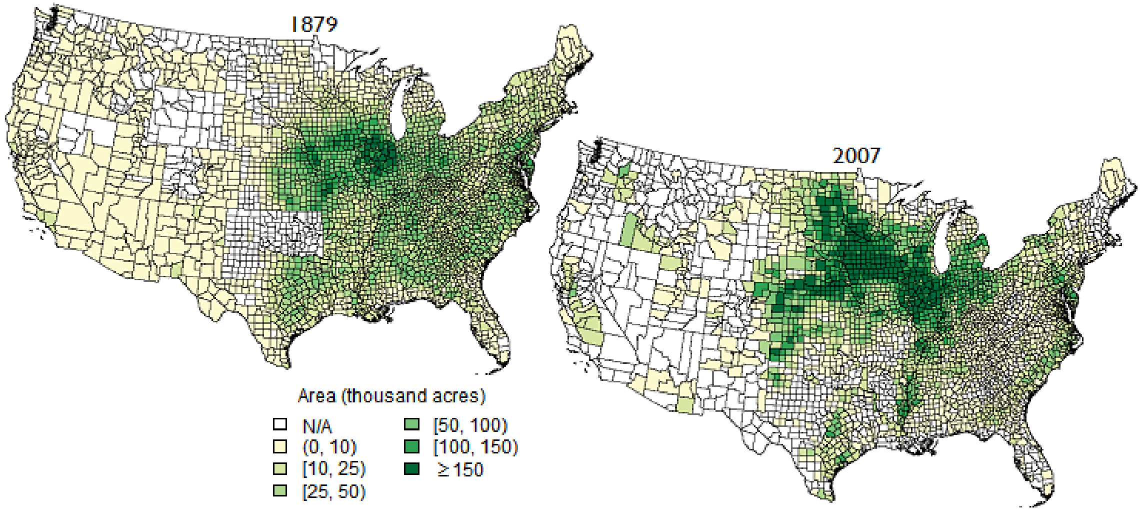
A comparison of county-level area and production data are shown in Figure 7 and Figure 8 (below), respectively. The county maps largely reflect the changes in regional production and area described above, but they also reveal significant intra-

---

[17] Nominal prices of corn reached a long-run high in 2007 and the real price was the highest in approximately a quarter century (see Alston, Beddow and Pardey 2010). The price increase has been attributed to a variety of factors, including increased biofuels demand, poor weather in important grain-producing regions, increases in energy prices, and increased demand for grains (especially due to economic growth in China and India). Reviews are provided by Trostle (2008), Mitchell (2008), and Wright (2011) among many others.

regional changes—the relative county production and area *within* each region did not remain constant. For example, additional area was not added to the Lake States and Northern Plains in proportion to the existing area—counties in the northern portions of these regions saw disproportionate increases relative to the other parts of the regions. Similarly, while the Southeast and Atlantic regions decreased their area and output, the counties near the Atlantic Coast actually increased both their corn acreage and output. These intra-regional trends will be further explored later in this chapter. County-level maps of production, area and yield for each of the 23 census years are available in the Corn Atlas.

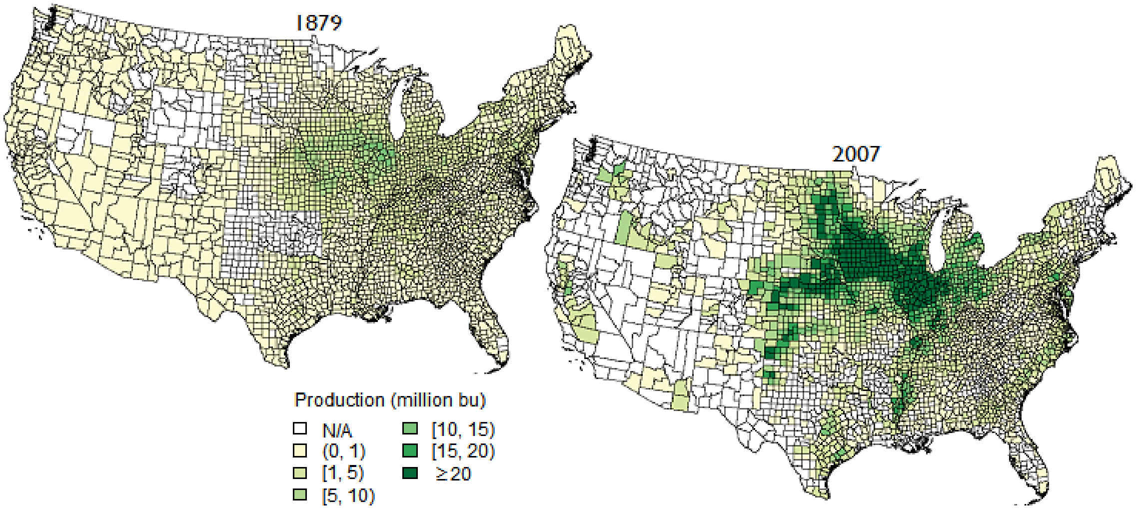
Figure 7: Corn Acreage by County, 1879 and 2007



Source: Created using data described in Section A.2.

Notes: The color of each county represents the number harvested corn acres in that county (in thousands).

Figure 8: Corn Production by County, 1879 and 2007



Source: Created using data described in Section A.2.

Notes: The color of each county represents the number harvested corn bushels in that county (in thousands).

One way to characterize the changing spatial footprint of corn production is to consider changes in the centroid of output and area. The centroid of a spatial variable, also referred to as the mean center, is the point that minimizes the sum of squared distances to all other points. In the present case, each bushel of corn or acre under corn production can be considered a point, so that the goal is to minimize the weighted sum of squared distances.[18] To simplify calculations it is assumed that all corn is grown at the geographic center of the county in which it is produced. So, for example, the centroid of output for a given year is the point on the map that minimizes the sum of squared distances from each county center to the point, with each distance weighted by the county production for the year—in essence, it represents the geographical pivot point.

---

[18] When a weighting is used, the centroid is sometimes referred to as the "mean center," while "centroid" is reserved for unweighted, purely spatial, calculations. No such distinction is made in the present assessment as the meaning shall be clear from the context.

To simplify calculations for this and later analysis, all points were projected onto a flat surface using the Albers Equal Area Conic projection formulae provided by Snyder (1987) and represented in meters east and north of a central point, thus allowing for use of Euclidian geometry to derive distances between points. This projection results in some minor distortions of distance (up to 1.25 percent for the conterminous United States),[19] as Snyder (1987, pp. 97-100) and the references therein discuss in more detail.

Over the 130 years for which complete county-level data are available, the mean center of corn output moved nearly 279 kilometers west and 157 kilometers north of the 1879 mean center, from central Illinois to southeastern Iowa. The mean center of corn area has moved even more, shifting 343 kilometers to the west and 278 kilometers to the north. While the 2007 area and output mean centers are at about the same location in southern Iowa, the 1879 area centroid fell 78 kilometers to the east and 123 kilometers to the south of the production centroid. This implies that in 1879, areas north and west of the centroid were relatively higher-yielding than average, and that the movement of the centroid caused (or was caused by) expansion into relatively high-yielding areas.

The mean centers took a circuitous route from their 1879 locations to their 2007 locations. Table 1 summarizes movements in the mean center of output and area for nine- or ten-year periods ending in the indicated census years. The changes are described as period-to-period centroid movements to the east (easting) and to the

---

[19] The projection was specified with standard parallels of 29.5°N and 45.5°N and a central meridian of 96°W. The location of any point is therefore expressed in meters relative to 96°Wx37.5°N, which falls in the Elk County, Kansas (located in the southeastern portion of the state).

Table 1: Movement in the Corn Mean Centers, 1879-2007

	Easting			Northing		
Year	Output	Acres	Relative	Output	Acres	Relative
			kilometers			
1889	-147	-134	1.1	-23	11	-2.2
1899	24	-27	-0.9	2	-15	-0.1
1909	21	-46	-0.4	12	-4	-3.4
1919	19	43	0.5	-5	-8	0.7
1929	-99	-104	0.9	26	37	0.7
1939	118	123	1.0	33	-37	-0.9
1949	-42	-61	0.7	41	128	0.3
1959	-25	-4	5.7	35	73	0.5
1969	-10	9	-1.1	26	41	0.6
1978	-40	-28	1.4	11	33	0.3
1987	-32	-27	1.2	7	18	0.4
1997	-61	-62	1.0	-9	-2	4.1
2007	-7	-24	0.3	2	4	0.5
Total	-279	-343	0.8	157	278	0.6

Source: Calculated using data described in Appendix A.2.

Notes: Each value for Output and Acres represents the movement of the corresponding mean center to the east (easting) or north (northing), relative to the previous centroid. The values for 1889 represent the movement from the 1879 mean centers.

north (northing), and negative values indicate movements in the opposite direction. Both centroids moved to the north and west in most years, and both trends became more established after 1939. Some of the centroid movement is likely explained by relative increases in higher-yielding areas, but other factors also affected the centroid. For example, settlement and urbanization patterns influenced the landscape of production, although the relationship between the changing population frontier and farming locations is rather complex (see Maizel et al. 1999). Improvements in transportation and communication infrastructure also influenced the footprint of production, namely by allowing for regional specialization in the crops most suited to the local agroecological environment (Kim 1998).[20]

Figure 9 describes the relative contribution of each county to the movement of the area centroid. The red arrow shows the movement in the national centroid, with the origin point representing the 1879 national area centroid and the tip of the arrow representing the 2007 centroid.[21] The area centroid moved northeast from southeastern Illinois to southeastern Iowa. Each of the black arrows originates at the geographic center of a county, and longer arrows indicate that the county had a relatively large impact on the centroid movement; counties without arrows (represented by dots) did not have much effect on the movement of the national centroid.[22] The arrowheads indicate the direction in which the national centroid was pulled by the change in area in the corresponding counties. Similar maps for different beginning and ending years and for the production centroid are available in the Corn Atlas.

The figure highlights a number of important trends. Although the Corn Belt exhibited an increase of 12 million acres over the period, changes in the spatial

---

[20] Kim (1998) asserts that the economic integration enabled by improvements in transportation (among other things) enabled regional specialization, and invokes notions of comparative advantage to explain the changing footprint of production.

[21] These centroids roughly correspond to those calculated by Reilly et al. (2003) using state-level data.

[22] Specifically, the analysis for a given county considered how the centroid would have changed had the county maintained its 1879 area share relative to how the centroid actually moved. Thus, the arrows represent the effect of changing area *shares* on the national area centriod. To accomplish this, the relative torque of each county was calculated to account for the fact that a change in a county located farther away from the centroid has a more pronounced effect than the same change in a county near the centroid. Each arrow is constructed such that its extent in both directions is 100 times the distance the county moved the centroid in that direction. Counties represented by dots moved the national centroid by less than 100 meters in any direction.

location of production within that region did not markedly affect the centroid. This is at least partially because the centroid was already located in the Corn Belt, so the counties in that region had little torque in moving the centroid. Among the Lake States, a number of counties in southern and western Minnesota pulled the centroid markedly to the north, as did the eastern counties of the Dakotas. Changes in Texas largely cancelled each other, with increases in the counties on the Ogallala Aquifer and offsetting decreases in the eastern portion of the state. Finally, reduced acreage shares in the counties of the Southeast and Appalachian states on net had a large impact on the centroid, almost uniformly moving it to the northwest.

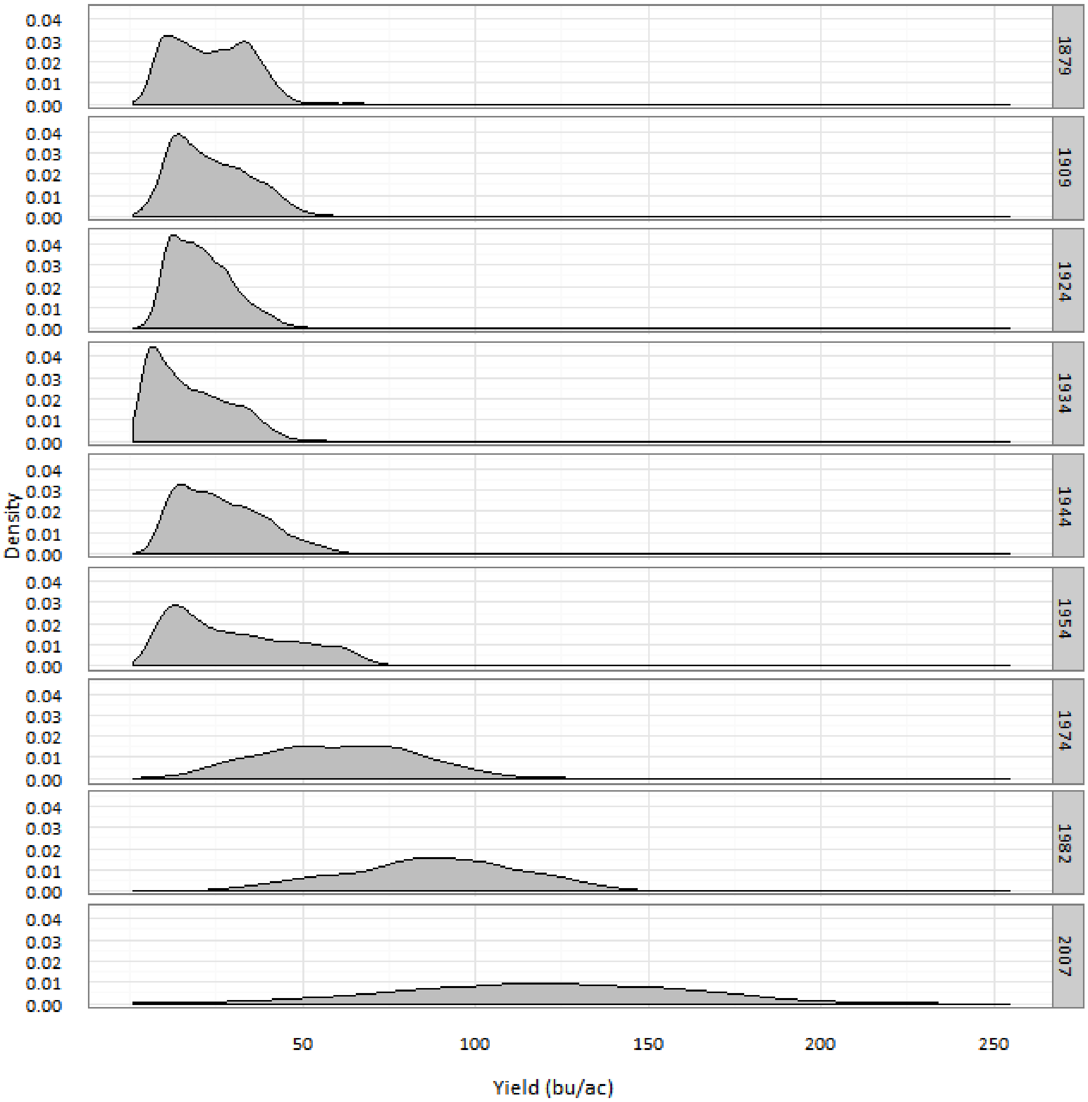
As implied by Figure 9 (and the similar figures in the Corn Atlas), there have been substantial changes in the relative distribution of production and area across counties, however it remains unclear how the distribution of yields has changed over time. Figure 10 (p. 36) shows smoothed county-level yield densities for selected years. As expected, mean yields clearly increase over time. However, neither (non-area weighted) mean nor median yields change appreciably from 1879 to 1924, although the distribution becomes increasingly positively skewed over this period (from μ_3=0.17 in 1879 to μ_3=0.71 in 1924). The 1934 distribution is of particular interest since the represented year occurred during the Dust Bowl; as a result, both the mean and median yield decreased, and the coefficient of variation increased sharply, from about 0.44 in 1924 to 0.64 in 1934. The most pronounced effect was in the movement of the first quartile from 13.5 bushels per acre to 7.7 bushels per acre; the effect was less pronounced in the higher-yielding portion of the distribution such that the inter-quartile range was larger than the median by about 20 percent. The years after the Dust Bowl saw steady increases in median and mean yields, and steadily decreasing skewness—by 1982, the distribution became slightly negatively skewed, although the distributions were more-or-less symmetric starting in the 1960s.

Figure 9: Contributions to the Movement of the Area Centroid, 1879-2007

Source: Derived using data described in Section A.2.

Notes: The movement of the national area mean center of corn production is shown by the bold red arrow. The amount by which each county affected the movement of the mean center is indicated by the length of the black arrow that starts at each county's centroid. The angle of each arrow represents the direction in which the area change in the county moved the centroid. Counties that did not have much effect on the national mean center are represented by dots. See Footnote 22 for more details.

Figure 10: Distribution of County Corn Yields, 1879-2007



Source: Calculated using data described in Appendix A.2.

Notes: The plots represent densities smoothed with a Gaussian kernel estimated using bandwidths determined by "Silverman's rule of thumb."[23]

---

[23] Namely, $1.06 \cdot \min(\sigma, IQR / 1.34) \cdot n^{-1/5}$, where 1.06 is the parameter recommended for Gaussian kernels, IQR is the interquartile range, and σ is the sample standard deviation.

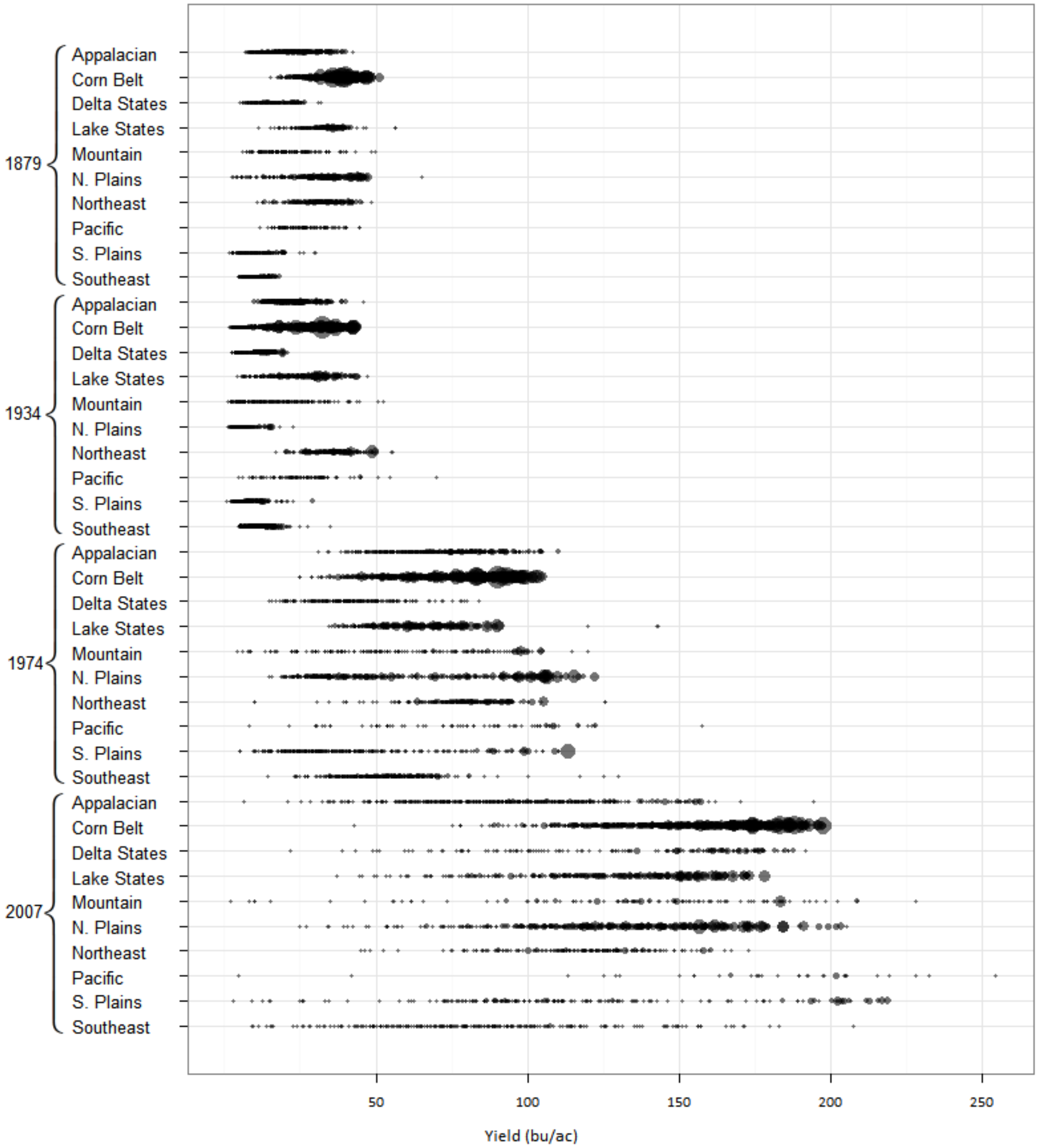
Figure 10 is aspatial, and before proceeding it would be useful to gain some insight into the spatial patterns underlying the raw county yield distributions. Figure 11 plots county-level yields by region for 1879, 1934, 1974 and 2007—each dot in the plot represents an individual county. The dots were made semi-transparent to avoid overplotting the series, and therefore darker points indicate that more than one county had a similar yield. The raw county yield distributions gave no information on the relative importance of the various counties; Figure 11 (p. 39) adds this information by scaling the county dots in proportion to the county's relative share of national output for each year—larger dots represent counties that contributed more to national output than did those represented by smaller dots.

From the figure, it is readily apparent that there have long been a number of counties in the Corn Belt with relatively high production and yields. The two modes apparent in the 1879 distribution (Figure 10, above) appear to have resulted from low yields in the Southeast, Southern Plains and Delta States (lower mode) contrasted with higher yields in the Corn Belt, Northern Plains and Lake States. By 1974 a number of counties in the Northern Plains had surpassed the highest yields in the Corn Belt, although the Northern Plains also contained a number of counties with yields among the lowest nationwide. Over the subsequent three decades, the Corn Belt, Lake States and Northern Plains saw marked increases in the yields of their largest-producing counties although, as in 1974, the Southern Plains presented with several moderately important high-yielding counties.

Moreover, the spatial distribution of yields has widened. In 1879, the average yield of the highest five percent of the yielding counties was 5.6 times greater than the yield of the lowest five percent of counties. By 2007, the difference had increased to

7.3 times. About a fifth of the highest-yielding counties in 2007 entered production after 1879, while over a third of 1879's lowest-yielding counties stopped producing corn. Thus, some of the national yield increase is attributable to counties moving into and out of production. About 40 percent of 2007's highest yielding counties were either already high-yielding in 1879, or entered production after 1879.

Figure 11: Relative County Yield Distribution



Source: Calculated using data described in Appendix A.2.

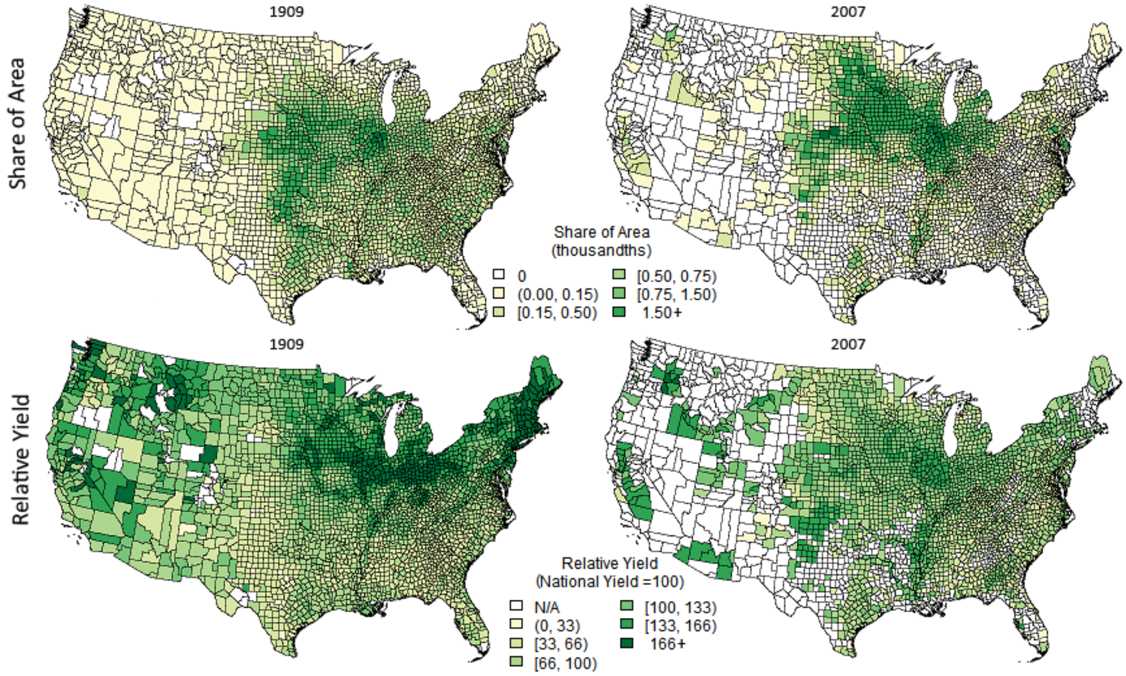
Notes: Each dot represents the yield in a county (shown on the horizontal axis). The size of each dot is proportional to the output of the corresponding county.

2.2.3 Moving Matters—Yield Indexing

The preceding sections of this chapter can be summarized rather succinctly. First, U.S. corn output has increased dramatically over the past 14 decades (about 17-fold) while increases in acreage have been more muted, and there have been long periods over which acreage decreased while output was increasing—for example the area devoted to corn dropped by over one third between 1944 and 1965, while output increased by almost 50 percent. This reflects the general long-run trend (Figure 1, p. 13) of output changes being much more correlated with yield changes than area changes. Notably, these changes in output, area and yield were not uniformly distributed across the United States—the *relative* output and area among the various counties changed as well, resulting in differential yield growth rates. This section will develop and apply a procedure to decompose the change in production into changes in the footprint of production, yield and area under production.

Figure 12 shows the national share of area accounted for by each corn producing county in 1909 and 2007 (top maps). The share of area in the Corn Belt, Lake States and Northern Plains counties has increased, at least partially due to relative decreases in the other parts of the country. The bottom maps show each county's yield relative to the national yield for each of the years (where national yield equals 100). The changes in relative yield have been similar to, but noticeably different from the relative area changes. For example, the relative yields in many of the counties that most markedly increased their acreage share have decreased, so that the Corn Belt now looks much more like the other major corn producing areas in terms of its yields (see Figure 11 and the corresponding discussion).

Figure 12: Relative Yield and Area, 1909 and 2007



Source: Calculated using data described in Appendix A.2.

Notes: The top figures show the share of national area accounted for by each county as a decimal percentage. The bottom figures show the yield of each county, indexed so that the national yield equals 100 in the given year.

The often confounding changes in relative yield, relative area and in the total area devoted to corn make it difficult to understand exactly what has caused national output to change. This section develops a new, spatially-sensitive method for decomposing a change in national output into changes in national area, yield and the relative county-level areas under production.[24] As will be seen, relatively simple concepts from economic index theory can be used, with slight modification, to decompose the national yield change. The analysis will take the yield changes as

---

[24] Hazell (1984) presented a method to decompose changes in output and output variability in terms of changes in yields, areas, their covariances and interactions. Among other things, the approach allows for attribution of changes in average yields of crop categories (e.g., cereals) to changes in the constituent crops. The method does not, however, allow for estimation of the direct effect of spatial changes. Several examples of this method can be found in Anderson and Hazell (1984).

given; the effects of spatial changes (and other factors) on yields will be explored in-depth in Chapter 3. The analysis would not be possible without having production and area data mapped to consistent county boundaries.

Let a_t and y_t represent vectors of county-level corn acreages and yields for some year, *t*. The length of each vector is equal to the number of counties, and they are arranged such that the *i*th element of each vector corresponds to the same county (thus, $a_{it} \cdot y_{it}$ is the total output of county *i* in year *t*). Then, total production in year *t* is the sum $y_t{}'a_t$, and the ratio of national output between year-*t* and some base year, *b*, is a simple index of output:

$$I_O^S = \frac{y_t{}'a_t}{y_b{}'a_b}.$$

Multiplying the numerator and denominator by $y_t{}'a_b$ and rearranging terms reveals that:

$$I_O^S = \frac{y_t{}'a_t}{y_t{}'a_b} \cdot \frac{y_t{}'a_b}{y_b{}'a_b},$$

where the first term is a Paasche area index (hereafter denoted I_A^P), and the second term is a Laspeyres yield index (I_Y^L).[25] A more convenient notation can therefore be employed:

$$I_O^S = I_A^P \cdot I_Y^L$$

---

[25] These yield indexes should not be confused with the "crop yield indexes" described by Hirsch (1942, 1943), Working (1940) and others. The latter indexing approach endeavors to aggregate the yields of different crops, usually during the same year, with the goal of comparing total yield across multi-crop enterprises or regions. The goal here is to decompose the total output of a single crop. The earlier notion of crop yield indexing did not generate much attention after the first half of the century.

By symmetry, the output index can also be decomposed into a Laspeyres area index and a Paasche yield index,[26] namely $I_O^S = I_A^P I_Y^L = I_A^L I_Y^P$, so that the output ratio can be defined by the product of a set of hypothetical quantities:

- the output ratio that would result if the yield *in each county* were held constant at the year-*t* (year-*b*) value, allowing the area in each county to change, as represented by I_A^P (I_A^L), and
- the output ratio that would result if the area in each county were held constant at the year-*b* (year-*t*) value, allowing the county yield levels to change, as represented by I_Y^L (I_Y^P).

Both specifications are equally valid, and which is appropriate for a given assessment depends on the questions to be answered.

The area indexes are of particular interest as they are well-suited to answer questions about the effects of the spatial allocation. For example:

- How much did the change in county areas contribute to current output (I_A^P)?
- What would the change in production have been if the change in county areas had occurred in the base year (I_A^L)?

The area indexes confound the effects of two factors: changes in national area and changes in relative production. If, for example, all counties increased their harvested area by 25 percent between a base period and year-*t*, both area indexes would equal 1.25 for time *t* (irrespective of any changes in yield). Alternatively, if harvested area is reallocated to counties with higher yield, but the *total* area under production and the

---

[26] The index number terminology used by economists can cause some confusion. The present indexes are named by reference to standard terminology for Laspeyres and Paasche price and quantity indexes. In a price (quantity) index, the quantity (price) vector is held constant at either the base-year value (Laspeyres) or the current (year-*t*) value (Paasche). Thus, in the present assessment, area indexes hold yield constant and yield indexes hold area constant.

yield vector are held constant, the resulting increase in production will also be reflected exclusively in the area indexes (even though national average yield will have increased). It would be useful to separate the scaling effect of a change in overall (national) area from the effect of spatially reallocating production.

Consider the Laspeyres area index, defined as:

$$I_A^L = \frac{\mathbf{y}_b{'}\mathbf{a}_t}{\mathbf{y}_b{'}\mathbf{a}_b} = \frac{\sum_{i=1}^{n} y_{bi}a_{ti}}{\sum_{i=1}^{n} y_{bi}a_{bi}}.$$

Suppose that each county changes its area by the same percentage, say p percent, so that $a_t = a_b(1+p)$. This equi-proportional increase in area will generate an index value of $\frac{\mathbf{y}_b{'}\mathbf{a}_b(1+p)}{\mathbf{y}_b{'}\mathbf{a}_b} = (1+p)$, where p exactly equals the percentage change in national area (namely, $1+p = 1'\mathbf{a}_t / 1'\mathbf{a}_b \equiv A_t / A_b$). However, the same *total* area change can generate a different index value if the new (or deleted) areas are not distributed across counties in proportion to their base-period areas. For example, suppose that the total area under production again increases by p percentage points, but one county, say *N*, changes its area by a factor of α percentage points. Relative to the equi-proportional change, the other *N*-1 counties will receive more acres if $p > \alpha$, and fewer if $p < \alpha$ and the index value will change by:[27]

$$s_N(p-\alpha)\left(\sum_{i=1}^{N-1} s_{bi}y_{bi} - y_{bN}\right) = I_A^L - A_t / A_b,$$

---

[27] See Section A.3 for the derivation. Note that the conclusions apply to the more general case in which $n \le N$ counties all change at different rates since the "N^{th}" county and its associated area, growth percentage and yield in both periods could represent a weighted average of any arbitrary aggregation of counties.

where s_N is county *N*'s share of *total* area and s_{bi} is county *i*'s share of the acreage in the first *N*-1 counties. Notice that the second bracketed term is equal to the area weighted average yield of the first *N*-1 counties less the yield for the county whose area increased by α percent. Thus, If α > *p*, county *N* took more than its area weighted share of the new acres and the first bracketed term will be negative. If county *N*'s yield is the same as the area weighted average of the other counties, the index will not change relative to that for the equi-proportional change; in this case if its yield is higher (lower) than average, the bracketed term will be negative (positive) and the index will increase (decrease).

For a general change in the area vector, I_A^L will always differ by $I_A^L - A_t/A_b$ relative to its value under an equi-proportional area change of the same total magnitude. If the index value is larger (smaller) than A_t/A_b, the change in relative area shares (i.e., the footprint of agriculture) caused the index to increase (decrease). Therefore, the effect of the relative spatial reallocation can be summarized by:

$$\frac{I_A^L - A_t/A_b}{(1+p)} = \frac{I_A^L - A_t/A_b}{A_t/A_b} = I_A^L \frac{A_b}{A_t} - 1 .$$

The latter quantity suggests an index of the effect of the relative spatial reallocation, the Laspeyres version of which is:

$$I_R^L = I_A^L \frac{A_b}{A_t} .$$

Notice that this is equivalent to scaling each of the year-*t* areas in the numerator so that they sum to A_b, thus maintaining the relative year-*t* spatial allocation while removing the effect of the change in overall area. A similar Paasche-type reallocation index can be defined by scaling the year-*b* areas so that they sum to A_t:

$$I_R^P = I_A^P \frac{A_b}{A_t}$$

While the Paasche reallocation index is also scaled by A_b/A_t, intuition is aided by keeping in mind that this is equivalent to scaling the denominator by A_t/A_b, thus scaling the base-year areas so that their sum equals the total area in the current year.

The spatial reallocation indexes answer slightly different questions from those above, namely:

- I_R^P gives the amount by which output increased due to the change in the *relative* spatial allocation between the base year and year *t*, evaluated at year-*t* yields and,
- I_R^L gives the amount by which output increased due to the change in the relative allocation, evaluated at base-year yields.

The county production and area data do not include observations for all county-years in the panel, either because data were withheld due to privacy concerns or because there was no production in the given county-year (and in some census years, the reason for the null observations was not reported). In the latter case, the null values for area and production can be replaced with zeros. When data were withheld to maintain the privacy of individual producers, it is assumed that the production in the county-year was accounted for by just one or a few farms, and therefore that the total production and area in the county were not very large. The area and production in these cases are also assumed to be zero; in the majority of such cases the effect on the calculated index will be minor.

Missing yield observations are somewhat more problematic. Consider, for example, calculation of an area index. It is possible that a county with no reported production in, say, 1889, markedly increased its area and output by 2007. A

Laspeyres area index that uses 1889 as its base year will have in its numerator the inner product of the base-year yield and the year-2007 areas. If some of the 2007 areas had no yield observation in 1889, using zero as the yield value will introduce a downward bias into the index. Had the crop been grown in the county in 1889, it is likely that the yield would have been positive, especially given that it was grown there in 2007 (otherwise the county would not affect the numerator of the index). To reduce the bias, it is helpful to estimate the yield that the county would have exhibited had it grown the crop. To resolve this issue, the inverse-distance weighted mean of the county's three nearest neighbors for which production and area were reported is used as an estimate of the county's yield in the period.[28]

The Paasche and Laspeyres values for a given variable will not, in general, be the same, and the differences between the indexes have received more attention in the literature than most topics of similar import. The inequality of the indexes is almost universally viewed as a deficiency among those whose goal is to determine the rate of inflation, and many solutions have been put forth. These objections usually arise because it is rarely obvious whether one should choose the base period or the current period to weight the index, and many of the proposed solutions are constructed to take account of both weighting vectors. For example, the Fisher index is calculated as the geometric mean of the Laspeyres and Paasche indexes for a given variable,[29] while the Marshall-Edgeworth index uses an arithmetic average of the weights. As shall be

---

[28] The distance between counties was determined using the county centroids.

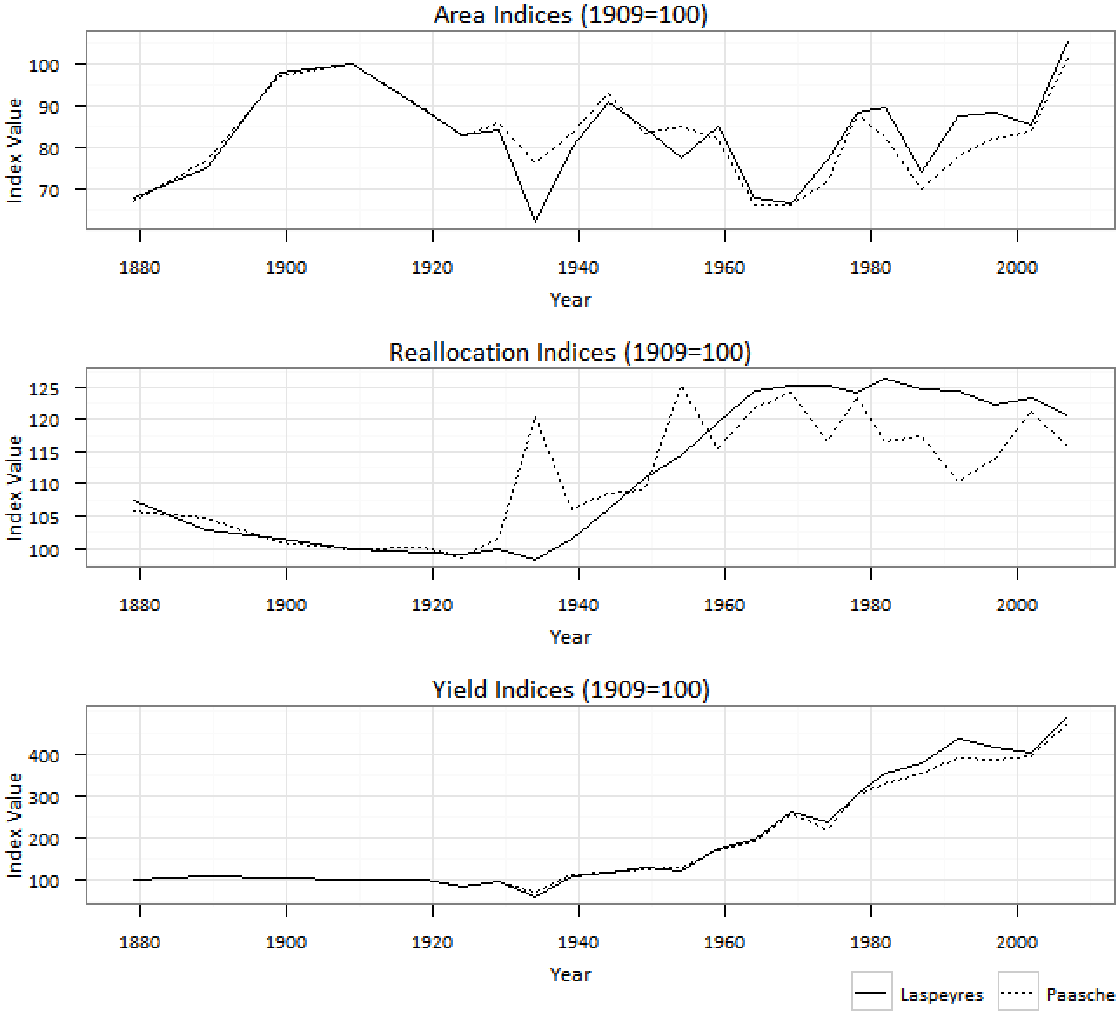
[29] While this was the specification originally suggested by Fisher (1921), in modern practice, the term "Fisher index" refers almost exclusively to the geometric mean of *chain-linked* versions of the indexes which have been given added economic interpretation as a consequence of this chaining approach (see Richter 1966 with regard to economic indexes in a production context and Boskin et al. 1998 for a consumption perspective).

demonstrated, the basic indexes will yield different information about the nature of production, and additional insight will be gleaned from the differences in the indexes—for the current purpose, it will be advantageous to keep the traditional weights and no attempt will be made to empirically reconcile disparities among the indexes.

Other formulations, such as the Young index use expenditure shares from a given period to weight prices; if applied to the present problem this approach would be equivalent to holding the production shares of the counties constant. Holding total area constant while allowing the shares to change has no known analogue in the economic index domain, likely because neither the sum of prices nor that of quantities of different goods would have a readily interpretable meaning, while the sum of areas does.

Figure 13 shows the calculated indexes, the values for which are provided in the Appendix (Table 11, p. 169). The solid lines in each panel represent Laspeyres indexes and the dashed lines represent Paasche indexes. 1909 was chosen as the base year for all of the indexes because more counties were represented in 1909 than in the earlier years, thus minimizing bias that might result from the yield estimation procedure described above (charts for the indexes using other base years are provided in the Corn Atlas). Before the base year, the area indexes (top panel) were generally increasing, as total area under corn production reached a peak in 1909. Over the same period, the yield indexes (bottom panel) both rose slightly then fell; both showed a net decrease by 1909, indicating that yields were falling in both the new areas (I_Y^P) and the old areas (I_Y^L). This trend is reflected very clearly in the reallocation indexes (center panel)—which reveal the extent of change in U.S. corn output attributable to a change in relative area from the 1909 footprint of corn production (Laspeyres) or the 2007 footprint (Paasche)—and both show a fairly substantial and steady decrease between 1879 and 1909.

Figure 13: Area, Yield and Reallocation Indexes, 1879-2007



Area Indices (1909=100)

Reallocation Indices (1909=100)

Yield Indices (1909=100)

Source: Calculated using data described in Appendix A.2.

Notes: The reader might take note of the differing scales in the various panels. The area and reallocation indexes range over about 30 to 40 percentage points while the yield indexes range over about 400 percentage points, causing the latter indexes to appear smoother than the others; the coefficient of variation of the yield indexes is actually much larger than either the area or the reallocation indexes.

After 1909, the area indexes both fell, reaching troughs in the Dust Bowl year 1934.[30] Much of the decrease can be explained by a decrease in total area under production (I_A^S=63). With such a sharp change in area, the reallocation indexes become particularly useful. In that year, a marked difference can be seen between the Laspeyres reallocation index value (98) and the value of the Paasche version (120). Thus, the "new" acres were less affected than the old acres, likely because production had already partially adjusted by 1934 (about four years into the Dust Bowl). Indeed, between 1929 and 1934, the Northern Plains reduced its corn acreage by over 16 million acres (91 percent) and the region's average yield dropped by nearly 67 percent.[31] It is also likely that crop failures accounted for some of the apparent beneficial reallocation of the crop as the figures here use *harvested* acres, not planted area as the denominator for yield calculations.

The Paasche reallocation index peaked sharply again in 1954, reaching a value of 125. As in the dust bowl, this was due to deleterious events that did not affect some of the highest-yielding areas: a heat wave covering the plains states, southern portions of the Corn Belt and the southeast. The corn acreage of


SUN BURNING ITS BRAND DEEP IN NATION'S FIELDS

(Chicago Tribune: 25 July 1954)

---

[30] The effects of the Dust Bowl were most pronounced in the Northern Plains region, which produced about 19 percent the nation's corn in 1929; the region's share of output shrunk to just over one percent in 1934. The period was characterized by drought which, when combined with poor soil conservation practices, caused widespread loss of topsoil. See Zeynep and Libecap (2004) for more information.

[31] The reader is invited to inspect the regional acreage, yield and production tables provided in the Appendix on pages 164-166. In addition, one may gain particular insight by comparing the Corn Atlas maps of various series for 1929 and 1934.

Missouri fell by nearly a quarter, while that of Kansas fell by over forty percent. Regarding Missouri, Westcott and Grady (2010) report that "[t]he damage to the corn was particularly devastating as hot, dry conditions continued through the crucial tasseling stage...(p. 4)." However, the writers of the 1954 Census note that the effect of the heat wave on the national yield was muted because of "…favorable weather conditions in most parts of the important corn producing States, the use of fertilizer and hybrid seed, and the improvement in tillage (Vol. 2, Part 7, p 580)." This story supports the divergence in the Paasche and Laspeyres reallocation indexes, and highlights the utility of keeping the indexes separate rather than trying to reconcile their differences.

The change in the relative spatial allocation of production between 1909 and 2007 accounted for about 16 percent of the change in output according to the Paasche version of the reallocation index, and by 21 percent according to the Laspeyres version. Most of this increase occurred during the period of rapid hybrid and fertilizer adoption (see Figure 3, p. 17). Spatial reallocation has not contributed much to the overall production of the nation during the four decades preceding the latest census—the indexes were more-or-less constant during the 1960s and 1970s, and generally decreased slightly thereafter.[32] Across the board growth in corn yields was the main reason that corn output grew by 103 percent in the 1960s and 1970s.[33] Of particular interest is the marked increase in national yield and area reported in the

---

[32] The relative stability of area shares from 1964-2002 can be seen in the animated area shares map provided in the Corn Atlas.

[33] This figure represents the output growth between census years 1959 and 1982. Over 99.5 percent of the counties that reported corn area and output in both of these years exhibited increased yield, and the total area under production decreased slightly.

2007 census over that reported in the 2002 census. If the index is calculated with a base year of 2002, it shows almost no change between the periods, indicating that increases in area and output were more-or-less proportional across the important corn-producing counties. As shall be seen in the following chapter, the weather of 2007 was also more suited to the production of corn than was the weather of 2002, accounting for at least some of the output increase.

Beginning in 1959, the Laspeyres reallocation index is always lower than the Paasche reallocation index. This occurred because of the widening yield distribution (see Figure 10 on p. 36)—the current yield advantage of the current-year area over the base-year area was decreasing relative to the yield advantage derived using base-year yields. However, more recent trends are narrowing the gap significantly.

2.3 *Conclusion*

The foregoing sections of this chapter highlighted most of the major events in U.S. corn production over the past 150 years. The adapted indexing methods provided a straightforward way to decompose changes in output, and to determine the extent to which changes in the footprint of production affect output. However, the indexes embed an implicit assumption that yields of each county are simple attributes of that county, leaving open the question of *why* yields vary over space and time. The discussion surrounding Figure 3 provided some insights by pointing out that yield is correlated with the adoption and use of certain technologies, such as fertilizer and improved varieties. Further, the analysis of changes in index numbers required reference to the weather to explain some of the values—for example, the spatial reallocation indexes would have been difficult to interpret absent knowledge of the dust bowl, the 1954 summer heat wave and the improvement in growing conditions in 2007. The following chapter will develop a more systematic approach for understanding how technology adoption and weather affect county yields. In combination with the insights developed in this chapter, the forthcoming analysis will lend insights into *how* spatial reallocation of growing areas affects output (versus the previous assessment which provided information on *how much* spatial reallocation affected output).

Chapter 3 Weather, Climate and Corn Yield

In the introduction, it was noted that most economic productivity assessments suffer from problems induced by (implicit or explicit) spatial and temporal aggregation. These problems are similar to those generated by left out or misspecified variables: for example, insofar as a spatially aggregated variable fails to represent the full range of the variation exhibited within each spatial sub-unit, the variable may be considered misspecified. Alternatively, one might consider that the aggregated variable is serving as a proxy for a number of left out local variables. Griliches (1957) considered the effects of misspecified production functions, and succinctly described the effects of not fully adjusting for the quality of land and labor. He demonstrated that production functions with left out or non quality-adjusted variables will at best generate biased parameter estimates. Alston, Babcock and Pardey (2010b) showed that productivity growth rates are biased when one or more inputs are left out of the analysis—so long as the left out inputs are not "…growing at the same rates as their included counterparts (p. 455)." Neither result is surprising given the standard treatment of omitted variables in introductory econometrics texts, nor is the problem new to economists. Jarrett (1957) noted that "[s]ince the number of possible variables which may be relevant in production decisions is very large some form of aggregation is necessary (p. 70)." In an insight he attributed to Plaxico (1955), Jarrett further noted that "…if inputs within the aggregate combine in a manner different from that specified in the production function fitted then bias will result (p. 70)."

Jarrett was certainly correct—the number of inputs to agricultural production does require aggregation if a more complete accounting of inputs is to be attempted. Indeed, simply enumerating the inputs to production would be an intractable exercise, much less properly modeling them. Economists have taken Jarrett's advice and endowed their productivity models with various indexes and indicators of agroecological suitability to account for differences in land quality and other factors.[34] As documented by Peterson (1986), during the decades preceding the mid-1980s economists widely used land prices to adjust for quality differences, an idea he attributed to the dissertations of Chicago students Hoover (1961) and Boyne (1962). Peterson took issue with the assumption that land prices were attributable to the utility of the land in farming and, as a result, concluded that they were not good indicators of agricultural land quality. Further analysis revealed that about two-thirds of the variation in agricultural land prices can be explained by population density, with the implication being that at most one-third of the variation in land prices derives from suitability for agricultural production.[35] However, it is still not clear how the various attributes of location influence yield.

Accounting for differences in weather variables across space and time is an important first step in assessing differences and changes in yields. However, one must

---

[34] It is worth keeping in mind that notions of "land quality" often include natural services such as rainfall and sunlight in addition to the quality of the land itself. In such cases, a more accurate term might be "location quality."

[35] Peterson was not interested in productivity per se, but in land prices. Thus, he states that "…land quality is a long-run rather than a short-run concept (p. 814)" and therefore that it was appropriate to use long-run average precipitation in his models. In the context of land prices, long-run average effects are important since prices are formed around buyers' and sellers' expectations. Such an argument cannot be applied to seasonal/annual productivity assessments.

take care in how the various variables are taken into account. Over 130 years ago, the writers of the 1880 Census of Agriculture highlighted this fact in reference to corn, noting that:

> "[t]he distribution of the crop according to climatic conditions depends upon certain peculiarities of climate rather than on averages of either temperature and rainfall. The mean annual temperature is of much less importance than is the summer temperature, and this, in turn, is less important even than the manner in which this summer temperature is distributed *(Vol 3, p. 92)*."

Economists often focus on temperature and precipitation when correcting for weather differences, and it is worth briefly exploring these variables in more detail. For most crops, corn included, there is a complex relationship between growth and temperature. In controlled environments, lower temperatures result in higher yields because the length of the grain-filling period is inversely proportional to the temperature (Hunter et al. 1977). However, temperature does affect the rate at which grain grows (Badu-Apraku et al. 1983)[36] so that when other variables, namely radiation, are held constant, the decrease in the grain filling period induced by high temperatures directly reduces yield. The situation becomes more complicated in natural environments where higher temperatures often result from higher solar radiation (and vice versa), so that the direct effect of temperature is muted (Muchow 1990).[37] This is not intended to imply that temperature is unimportant, but rather

---

[36] This and the previous citation are after Muchow (1990).

[37] The fact that temperature and moisture alone are not sufficient for plant growth is obvious to those who grow plants indoors. While the indoor temperature is more-or-less kept constant, indoor plant growth is influenced by the amount of direct sunlight received by a plant.

that one must be circumspect whenever temperature is used directly to explain crop yields.

As noted, economists also tend to consider precipitation (usually as total seasonal or annual precipitation) as independent variables in explaining yield differences over space and time (see Section 3.1.2 for examples). As with temperature, the situation is complicated by localized biological realities. Consider a simple example in which crop yield at a location is "corrected" for temperature and precipitation. Setting aside the discussion of temperature above, there are two primary difficulties with such a model. First, soil moisture, not precipitation, directly affects crop growth as most moisture is absorbed through the roots. This implies that soil hydrology should ideally be taken into account as, for example, there will be no marginal impact of precipitation if the soil is saturated (that is, soil moisture is a stock variable sourced by precipitation). Second, temperature is an important variable in determining the rate of evapotranspiration and therefore affects the relationship between moisture and yield. As a result, one should expect the partial effect of precipitation on yield to also be a function of temperature, and so models that (implicitly or explicitly) assume separability of these variables might be misspecified.

The above caveats are mentioned largely because they help clarify the indirect and complex role of weather variables in determining crop yields. The next section includes a brief review of the literature on the relationship between crop yields and the weather. When considering the studies, it is worth keeping in mind that the interactions between crops and weather are more complex than any model could possibly represent. Thus, researchers must make difficult decisions based on data availability, available resources, and the state of knowledge about the production function of the plant when modeling weather-yield relationships.

3.1 *Review of the Literature*

This section will briefly review the vast literature on the interaction between crop yields and the weather. The literature on the subject is so extensive that some selectivity had to be applied. While a study of any crop could potentially provide insights, those that focus on corn were generally chosen over those that do not, with the hope that the results of the model developed in this report can be informed by those of other studies. Research with a geographic scope including the United States was preferred for similar reasons. One goal of the review was to identify the various streams of research that have fed into the ways agricultural economists currently think about the weather. In the papers selected from the early-to-mid 1900s, there is more focus on the literature from the fields of agronomy, meteorology and geography. Starting in about 1950, when agricultural economists started to intensively think about the weather, the focus shifts to the economics literature, although no important studies have been intentionally left out.

3.1.1 Studies Descended from Smith's Work

One of the earliest statistical attempts to investigate the impacts of weather on crop yields using cross-sectional data was provided by Smith (1904) in the 1903 *Yearbook of Agriculture*. By plotting total average annual rainfall for 1888-1902 against each year's average yield for several Corn Belt states, the author concluded that "…if one knows the precipitation during the month of July over the great corn-producing district, he can estimate the yield for the season very closely (p. 217)." He further found that June and August precipitation were individually not very predictive, but that the sum of June and July rainfall was more highly correlated with yield, and that the sum of rainfall over all three months was even slightly better. Smith (1914) extended his earlier results using correlation coefficients between rainfall in various months with corn yield, again finding that for the major corn producing states, July rainfall has a larger effect on yield than precipitation in any other month. The later analysis was, however, somewhat more nuanced, allowing for the additional conclusion that precipitation in the ten days following tasseling "...has an almost dominating effect upon the yield of corn (p. 87)."

Smith's results hint that one might simply use July rainfall to predict yield, at least in the Corn Belt. To assess this possibility, the analysis was repeated using the long-run production and weather datasets constructed for this study.[38] The county output and precipitation data for the eight states in Smith's analysis (Ohio, Indiana, Illinois, Iowa, Nebraska, Kansas, Missouri and Kentucky) were aggregated to the state level.[39] Since yields increased markedly over the 119 year period, yields were detrended using a second order polynomial ($\bar{R}^2 = 0.95$) and deviations from this trend were considered instead of raw yields. As would be expected from Smith's result, neither June nor August precipitation was very highly correlated with yield deviations ($\rho = 0.04$ and - $\rho = -0.36$, respectively) while July precipitation was highly correlated

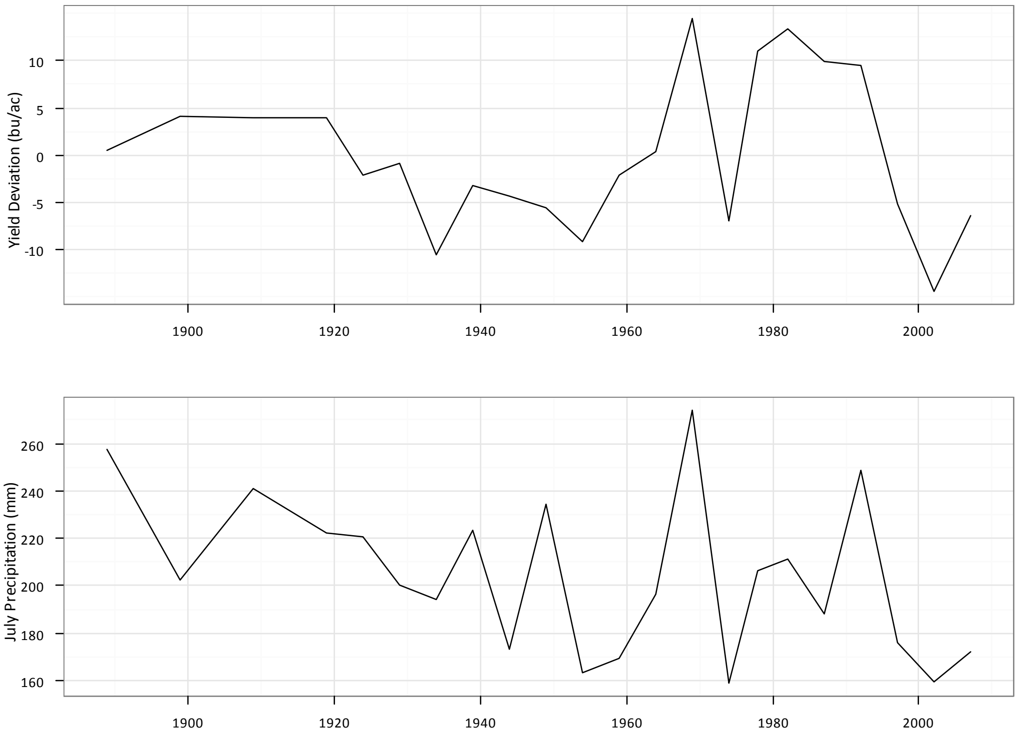
---

[38] For the present purpose it is sufficient to note that the weather dataset developed for this study includes monthly precipitation and temperature data for each county in the United States for 1879-2007. The dataset is discussed in more detail in Section 4.1 (below).

[39] County acreage was used to derive the weighted average precipitation over the area for each month.

with yield ($\rho = 0.75$). The result was similar when all states were included, yielding correlation coefficients of 0.07, 0.71 and -0.42 for June, July and August precipitation, respectively.

Figure14: July Precipitation and Corn Yield for Several States



Sources: Derived using the data described in Sections A.2 and 4.1.

Notes: The yield (top panel) and July precipitation (bottom panel) are for the eight states included in Smith's (1914) analysis (Ohio, Indiana, Illinois, Iowa, Nebraska, Kansas, Missouri and Kentucky). The yield (top panel) is the total production of the states divided by the total harvested area. The July precipitation is the average precipitation of the states weighted by their harvested corn area.

Wallace (1920) objected to Smith's conclusion, noting that while July rainfall is an important determinant of yield in several Corn Belt states, it is not the most important variable in others.[40] To show this, Wallace first derived a linear yield trend

---

[40] Other early researchers also found that July rainfall did not have a positive association with yield in some areas, for example Wilson (1914) for New York state and Wolfe (1925) for Blacksburg, Virginia.

for each of several Corn Belt states for the period 1891-1919 along with the correlation coefficient between yield deviations and deviations from mean monthly precipitation and temperatures for May, June, July and August. For each state, he then selected the three variables with the highest correlation coefficients and performed state-by-state linear regressions of yield deviations on those variables. He found, for example, that May and July temperatures, along with August rainfall were all more highly correlated with Iowa yield deviations than was July rainfall. While the author wisely noted that the most important monthly weather variables in determining yields may differ by state, he did not propose a mechanism through which these differences might arise.

Wallace went on to investigate, in some detail, the relationship between weather and yield in Iowa, taking note of the differences in agroclimatology prevalent in different parts of the state. He concluded that "[e]ach State is a specific problem in itself, and the probabilities are that each county in a State is a specific problem (p. 445)."

Charles Brooks (editor of the Monthly Weather Review) added a discussion section that outlined suggestions for an idealized study. These included:

- taking account of corn planting dates, "…taking the weather not by calendar months but by periods following the mean date when corn was planted in the region each year (p. 446)"
- the use of weekly instead of monthly mean weather data

The former author expresses a favorable view of the Corn Belt studies, but finds that "…the relationship between the summer rainfall and the yield of corn in New York is lacking entirely (p. 108)."

- consideration of the weather during the days following tasseling (per Smith's 1914 conclusions), and
- consideration of the rate at which the crop develops.

He noted in passing that such a study would likely "…require an impossible amount of labor for one person."

Kincer and Mattice (1928) extended Wallace's approach in an assessment of the effects of weather on wheat yields in North Dakota. In addition to mean monthly temperatures and rainfall, additional monthly weather variables such as the number of cloudy days and relative humidity were also considered. The authors used correlation coefficients to choose variables for a regression of yield on five weather factors. The Kincer and Mattice approach was later used by Mattice (1931) in a similar assessment of state-level corn yields for the Corn Belt states, developing a different equation for each state. Of particular interest here is that a different set of variables was chosen for each state.

The first use of a reasonably large panel of county-level data to assess the effect of weather on corn was by Rose (1936a), who studied the May through August precipitation and temperature effects on yield of 55 Corn Belt counties. The length of the time series varied by state, from 9 to 24 years, all ending in 1932. Rose posited that earlier attempts to study the effects of weather on corn yields were confounded by the use of state-level data. Using maps of correlation coefficients between county yield and monthly temperature and rainfall, the author delineated areas of the Corn Belt that had significantly positive or negative correlations between each variable and corn yield. He concluded that "[n]o one climatic factor gives significant correlations for all parts of the Corn Belt… (p. 100)," lending evidence to Wallace's earlier claim that the problem of correlating yield must take account of the local climatology. At about the same time, Rose (1936b) assessed the correlation between various climatic variables within the Corn Belt, using a 20 year series of monthly data from a number of weather stations. The study "…reveals that rainfall and accumulated temperatures, of the same month and successive months, correlate to significant extents in parts of the American Corn Belt (p. 79)." The findings provided an early warning about the potential for multicollinearity between monthly weather variables (although the author did not note that such collinearity may be problematic).

David and Harrell (1942) addressed both spatial and temporal aggregation problems by assessing yields from fields for which nearby weather station data were available, using rainfall and temperature variables defined for five-day periods. While this approach is not feasible in the present study, it is worth considering some of the authors' conclusions. While the authors noted that rainfall is likely to affect corn yields only insofar as it affects soil moisture, they did not endeavor to account for soil hydrology. They did, however, account for differential effects of rainfall and temperature deviations occurring at different times for a given location, finding that timing of events is indeed important, and that these effects vary by location.

A number of researchers have used regression techniques to estimate the effects of temperature and rainfall on corn yields for a single state or location. A few notable studies will be highlighted insofar as they provide insights of use to the current effort. Huddleston (1955) was perhaps the first study to use a temperature-rainfall interaction term in a yield-weather regression, finding that July rainfall and temperature had negative individual effects on Iowa yields while their interaction had a positive effect.[41] Runge and Odell (1958) used yield and weather data for a single farm in Illinois, dividing the season into a number of periods as small as two days. Among other conclusions, they found that "…corn yields are influenced much more by weather conditions immediately before and during anthesis than during any other parts of the growing season (p. 453)," and further that above normal rainfall during this period and the weeks preceding it are of particular importance.

One of the more influential regression studies was by Thompson (1969a), who regressed deviations in yield from a linear time trend, the deviation of several weather variables from their mean and the square of that deviation using data from five Corn Belt states. He included accumulated precipitation from September through June with the justification that this variable "…is an indicator of the reserve supply of soil moisture at the time when the crop starts its 'grand period' of growth in July (p. 454);" other variables included the monthly temperatures for June through August and rainfall for July and August. Thus, the model, or at least its justification, embeds Smith's conclusion about the importance of July rainfall from some 65 years prior. Of primary interest is that Thompson detrended each state's yield data separately, using two linear "technology" trends, one for 1930-1960 and another, apparently steeper, trend for 1960-1967. While all of the weather variables were significant, the trend accounted for over 83 percent of the total variation.

---

[41] It is worth noting that Huddleston's goal was to showcase the "inverse matrix" approach to regression, and wasn't necessarily intended as a robust assessment of the relationship between weather and corn.

Katz (1977) provided an important critique of Thompson's approach as employed in his 1969 paper (along with his similar studies on wheat (1970) and soybeans (1969b)), and other purely statistical models aimed at estimating the effects of weather on yield. Katz was writing in the context of the then recently issued report on potential impacts of climate change that employed regression models by Thompson and others to predict changes in (among other things) agricultural production due to predicted changes in climate. Katz raised two primary objections. First, he noted that the purely statistical "black box" approaches used in much of the literature may be incorrectly specified insofar as there is little or no assurance or concern that the models represent physiological realities. For example, few studies employed biologically-based logic to assess the direction and magnitude of the parameters they estimated. Along the same lines, Katz claimed that "…it is well known that the relationship between crop yields and a given climatic variable … is nonlinear (p. 89)," while many of the studies assumed that, for example, temperature and rainfall affect yield in proportion to their deviation from their mean. While such models may indeed explain within-sample variation, it may be inappropriate to apply them to other cases. Katz's second major objection was that many of the independent variables are highly correlated, producing the familiar problems associated with multicollinearity.

Nelson and Dale (1978a) implemented "Thompson approach" models for panels of Indiana county data, along with a model utilizing an "energy-crop growth" (ECG) index. The exact structure of the ECG index is not of particular relevance and it is sufficient to note that it endeavors to estimate plant-intercepted radiation, which is then corrected for moisture-stress.[42] Of primary interest is that the authors tested the use of time trends to account for technology in Thompson-type models against alternative models in which the amount of nitrogen applied was used to proxy for technology levels. Their general finding was that models in which nitrogen application rates are used as a technology proxy are superior to those that employ a time trend, and further that this is particularly relevant for studies in which the impacts of climate change are estimated. This conclusion was verified by a companion paper by the same authors (1978b), in which they bemoaned the disparate results obtained from the various statistical models, and suggested a methodology for comparing the predictions of various models.

3.1.2 Economists Discover the Weather

In addition to the lines of research descended from Smith's (1904 and 1914) work, a second stream of literature developed when economists became keenly interested in the effect of weather on yields beginning in the 1950s. Famously, in *The Economic Organization of Agriculture*, Schultz noted: "…variations in yield are in fact large; they are caused primarily by weather (1953, p. 178)." Cochrane responded rather dismissively to this notion, noting that while "[a] poor growing season may break an individual farmer, … the relative importance of the variable, weather, diminishes as the unit of inquiry moves from the individual farm operation to the

---

[42] The ECG index uses the daily amount of solar radiation and an estimated leaf area index (leaf size) to estimate the radiation intercepted by the leaf as a primary determinant of plant growth. The intercepted radiation is then reduced by relative evapotransipiration (that is, $^{ET}/_{PET}$), which correlates negatively with moisture stress. See Nelson and Dale (1978b) and the references therein for the exact specification.

national aggregate (1953, p. 282)."[43] While he offered no evidence, it does seem plausible that, on average, the national production consequences of weather effects will be more muted than the local effects, although it is unlikely that these local effects would exactly net each other out. More importantly, Cochrane (and Shultz) did not recognize that agriculture was on the move. Had this fact been clear, Cochrane might have assigned more importance to the weather, especially when taking a long-run view, noting that, for example, corn production was moving out of the warm Southeastern portion of the country and into the cooler Northern Plains and Lake States.

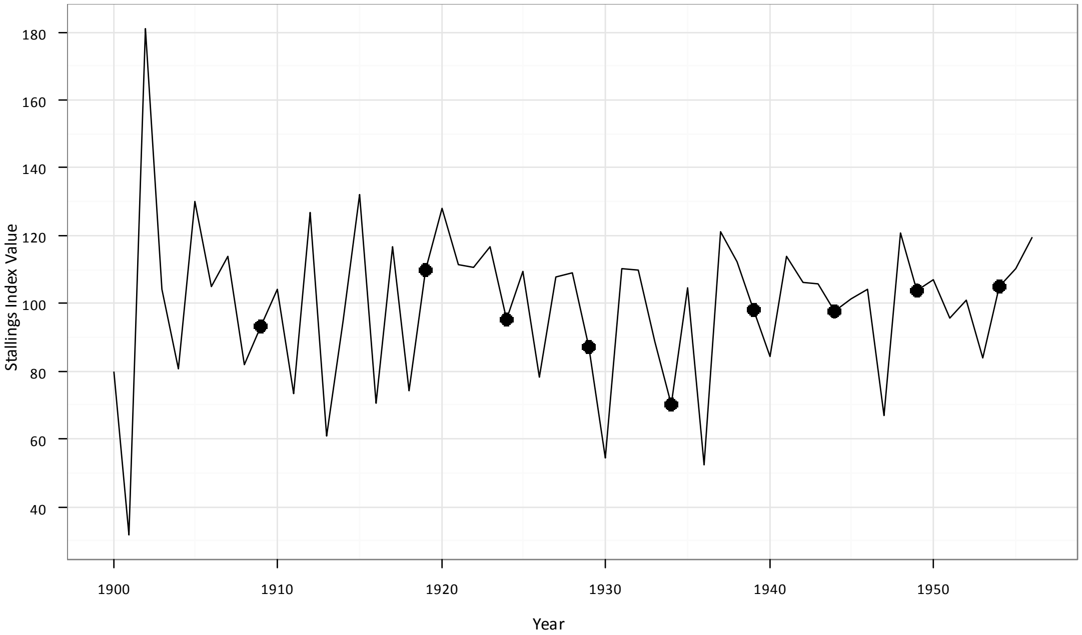
In any case, *Agriculture in an Unstable Economy* brought weather to the attention of mainstream agricultural economists. Stallings (1960) developed a long-run annual series of national weather indexes for U.S. crop production. Stallings' general approach was to use experimental data in which non-weather variables were assumed to be controlled and attributed yield departures from a time trend to weather. Other factors that might cause yields to depart from the trend were assumed to be positively correlated with the weather (e.g., insect damage and soil moisture), distributed such that they were accounted for by the time trend (e.g., fertilizer) or normally distributed. These indexes were calculated for individual crops at individual

---

[43] The article cited here was titled "Professor Schultz Discovers the Weather," and it is clear that the choice of title was not intended to be flattering. Cochrane seems to have had a special gift for writing critiques of books authored by high-profile authors; Schultz was neither the first nor the last to receive rather acrid rebuts from Cochrane who, for example later put forth that Kenneth Boulding was, at least as regards agricultural matters, a "nincompoop" (Levins 2003, p. 36). But, Prof. Schultz also made no secret of his opinion of Cochrane's work (Schultz 1951), although the two were later to become friends.

locations, then combined into a single national index by calculating the production-weighted average of the index. The results of Stallings' calculations for corn are presented in Figure 15, where the agricultural census years are indicated by a dot. Stallings assessed the index by regressing yield deviations from an 11-year moving average on the index values, finding that national yields deviated by 0.12 bushels for each percentage point change in his index (R^2=0.61).

Figure 15: Stallings' Corn Weather Index Values



Source: Created using data provided by Stallings (1960, Table 1, p. 183).

Notes: The line represents Stallings' weather index values and the dots show the years in the current dataset (the years in which an agricultural census was taken).

As summarized by Stallings (1961), the indexes were utilized in a number of studies to accomplish somewhat different goals. Cromarty (1957 and 1959)[44] and Griliches (1960) used the indexes in place of direct weather variables in assessing

---

[44] Cromarty used indexes provided in Stallings' (1958) Ph.D. dissertation.

agricultural supply and the economic structure of the agricultural economy. Clawson (1959) developed a very similar methodology for constructing weather indexes based on deviations from yield trends to adjust for the effects of weather on annual yield and output which he applied to correct for the weather effects in an assessment of the impacts of technology. In a third type of use, Hathaway (1959) used the Stalling indexes to explain changes in output, concluding that, in his assessment, the indexes "…support the hypothesis that factors other than weather are needed to explain the changes in crop yields (p. 492)," a conclusion that will be heeded in the current assessment.

In reference to Stallings' work, Plaxico (1961) stated that "Stallings is somewhat unique among economists in that he has done *more* than talk about the weather (p. 1160)." Plaxico went on to express a generally favorable opinion of Stallings' work, noting, for example, that Griliches had found that the index values account for "…a substantial fraction of the explained variance (Griliches 1960, p. 288)" when estimating aggregate U.S. farm supply functions. However, Plaxico noted that the Stallings indexes are error terms from regressions of yield on time, and therefore would be expected to explain yield variation by design. Thus, the indexes essentially label unexplained variation as an explanatory variable, and it therefore isn't clear that much is accomplished in the process.

Shaw brought the two streams of literature back together in his 1964 article. In his summary of the literature, Shaw noted several problems associated with the use of regression models in yield assessments. Of particular interest here is his discussion of spatial aggregation problems, and some of that discussion is worth repeating since it is directly relevant to the present work. The main difficulty in using spatial aggregates of yield and weather variables is that crops do not have a linear response to weather, so that if a number of individual units are aggregated it cannot be assumed that average weather over the area is relevant. A simple example was provided to make the point clear. Shaw asked the reader to suppose that a state is comprised of two districts, and that an arbitrary weather variable causes yield to decrease by 25 percent when it is either one unit above or below some optimal value. Consider a case in which the weather variable in one of the districts is one unit below the optimum, and it is one unit above the optimum in the other district. On average, the state will appear to have optimal weather, while state yields will be reduced by 25 percent. A state level assessment would be unable to tease out the effect of weather in such a case, and would appear to ascribe sub-optimal yields to optimal weather conditions. The assessment to be conducted in this study will consider the non-linearity of weather effects and will reduce the problems of spatial aggregation by using county-level data.

There is an important interaction between weather and technology. For example, Shaw (1964) reported on an Iowa study in which corn was grown both using early 1930s technology[45] and using current technology. A weather index for the yields from each experiment for 1930-1959 was then created, and it is clear that the new technology affected the calculated index markedly. Thus, the difference in the index values was used to assess the effect of technology on yield. During the 1950-1959

---

[45] The author neither specifies exactly which technologies were held constant to create the early 1930s scenario nor which technologies were adopted later.

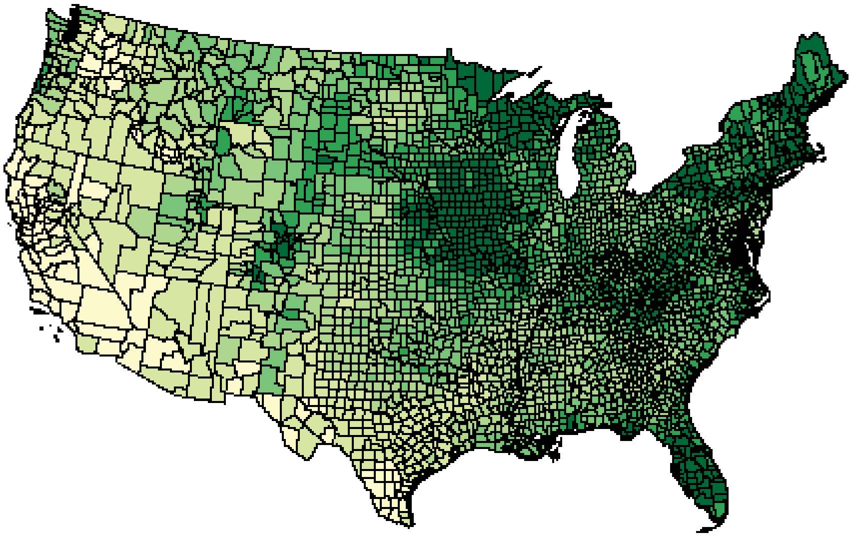
period, when current technology was most different from the technology of 1930, the indexes tend to be most different. This difference was most evident during periods of bad weather (e.g., the index was less than 100), when the index based on current technology exceeded the index based on old technology by 40 percent. While Shaw didn't make note of it, the conclusion makes it clear that technology must be directly accounted for instead of using time trends, a conclusion that would be confirmed over a decade later by Nelson and Dale (1978a) (as reported above).

There have been a number of attempts to combine temperature and precipitation variables into a single index. These indexes are justified by their similarity to the work of Köppen, and especially Thornwaite (1948), who classified the aridity of regions based on estimates of relative evapotranspiration—the aridity indicies can therefore be thought of as indicators of moisture stress. Oury (1965) described and implemented two aridity indexes: that of Ångström and another by de Martonne, both of which were formed as ratios of precipitation to simple increasing functions of temperature.[46] For example, for a given amount of precipitation, a 10°C increase in temperature will about halve the Ångström index value. To give the reader a sense of the Ångström index, values were calculated for May through September of 1982 and plotted in Figure 16, where darker shades indicate a higher index value. A map of the de Martonne values appears similar since, over the normal range of temperature, both indexes respond in the same direction to a change in temperature.

---

[46] The de Martonne index was $P/(T+10)$ and the Ångström index was $P/(1.07^T)$, where P is the total precipitation of a period in millimeters and T is the temperature in centigrade.

Figure 16: Ångström Aridity Index, 1982



Source: Created using data described in Section 4.1.
Notes: Darker counties were less arid as determined using the Ångström index.

As the indexes are conceived, precipitation will be of less benefit to the crop as the temperature increases. Oury tested two regressions, one of state-level yield on a linear time trend and the de Martonne index and the other using the Ångström index against a form using temperature and precipitation directly. Both of the indexes were more statistically significant in their respective regressions than were either temperature or rainfall in the check regression, although none of the regressions seemed to have any advantage in explaining variation in yield. These indexes have been employed in a number of economic studies, and especially in assessments of productivity (Carter and Zhang 1998; Hallam 1990; Thirtle and Bottomley 1988; Zhang and Carter 1997).

Doll (1967a) took a different approach to defining an index. He noted that "Oury, Stallings and Shaw reject the use of meteorological variables—primarily on the grounds that the functional relationship between these variables and yield is not known (p. 81)," and further that that the methods of Stallings and Shaw were impractical for estimating a panel of state-level indexes due to a paucity of experimental data in some states. Instead of using observed experimental yields to imply weather effects, Doll directly estimated the effect of precipitation on aggregate (industry) yield using an iterative approach that allowed inclusion of total precipitation during an arbitrary number of periods during the growing season. He then computed a yearly index by taking the ratio of predicted yield for a given year to the predicted yield under average conditions. Shaw (1967) objected, charging that by using a trend to account for technology, methods such as Doll's do not address the economists' primary goal: "…estimating the effects of controllable factors (p. 636)." Doll (1967b) made several remarks in response. First, he noted that there were insufficient experimental data to apply Shaw's approach in all states, and that the data that were available did not represent the average conditions of the state. Second, Doll noted that the appropriateness of a trend variable depends on what the researcher is trying to accomplish: if the goal is to explain yields, a trend is not an adequate correction for technology, while such a variable is sufficient if one wishes to assess the potential utility of a weather variable in explaining yield.

In the 1970s, economists started implementing models very similar to that of Thompson, despite Shaw's objection. For example, Swanson and Nyankori (1979) used various trend specifications along with cumulative September through April precipitation and monthly temperature and precipitation for May through August (the growing season in their area of interest). Herdt (1970) investigated models including various cumulative and individual monthly rainfall variables. Kaylen and

Koroma (1991) applied a Kalman filter algorithm to estimate the weather-yield relationship by calculating weather indexes derived for districts based on departures from normalized state-level weather.

Kaylen, Wade and Frank (1992) estimated corn yield equations for each of the U.S. corn regions (Figure 6, p. 26) using weather variables (calculated as deviations from their means), the square of weather variables, fertilizer prices and lagged crop prices. The authors made some attempt to account for different cropping seasons across the regions by considering a soil moisture variable calculated as the cumulative rainfall for the six months before the earliest planting month in the region (either April or May). Other monthly weather variables and their squares were also considered, and the final set of included variables was determined for each region using a subset regression procedure to select from 26 potential weather variables. As might be expected based on the previous literature, a different set of weather variables was selected for each region (except for the Southern Plains and Southeast, which both used only June precipitation and July temperature), and the signs and magnitudes of the estimated parameters differed for the same variables across regions.

Based on Thompson's results, Menz and Pardey (1983) used the log of national average nitrogen applications, an acreage weighted mean of July precipitation in several Corn Belt states and a time trend to explain annual variation in the national average corn yield for 1954-1980. Linear, square-root and log time trends were assessed, but in all cases they achieved an R^2 of about 0.95. That the R^2 values from these regressions were so high demands some attention. To assess this result, the

Menz and Pardey specification was applied to the county-level data compiled for this study, and the following regression was estimated for the census years in the current dataset:

$$y_{it} = \beta_0 + \beta_1 \ln(N_{it}) + P_{it} + T ,$$

where N_{it} is the level of nitrogen applied in county *i*, year *t*, P_{it} is the amount of June precipitation and *T* is the year. The parameter estimates were all positive and significant ($p < 0.01$), and the R^2 was 0.48. The variables were then aggregated to the national level, using acreage shares to weight nitrogen use and June precipitation, and the same regression was performed. The national regression also exhibited parameter estimates with the expected signs (although the estimated parameter on *ln*(*N*) was not statistically significant). The only difference between the two regressions was the level of aggregation. Nevertheless, the R^2 from the national regression was 0.99, more than double that of the county-level regression, underscoring the earlier cautions of Gehlke and Biehl (1934) and Robinson (1950) regarding scale effects.[47]

Offutt, Garcia and Pinar (1987) made a similar point after separately regressing yields from a single farm, crop districts and the U.S. overall on raw monthly weather variables, noting that "…the amount of variation explained by the trend and weather variables…increases with the level of aggregation (p.58)." Garcia et al. (1987) make the same point (using the same words), but also note that economists "…tend to introduce a certain circularity by defining bad weather as that which is associated with poor yields (p. 1096)." Nevertheless, the authors continue by

---

[47] The pattern continues as the level of aggregation increases. The R^2 for the regression on data aggregated to the state level was 0.65, and for data aggregated to the regional level, the R^2 was 0.74.

estimating weather-yield relationships using untransformed weather variables, thus implicitly assuming that bad (good) weather is that associated with low (high) yields.

Other economists have used proxy variables to account for weather indirectly. Alston et al. (2010a) used a USDA index of pasture moisture conditions to correct for weather effects at the state level. Houck and Gallagher (1976) used a similar pasture moisture index for several states, along with a temporal lag of the same variable. Interestingly, the estimated coefficients on the temporal lag of the weather variable seem to indicate that, in general, the previous year's moisture conditions are at least as important in determining corn yields as those of the current year.[48] Demir and Mahmud (2002) included a binary precipitation variable in a study of Turkish agriculture, where the variable was set to one if an area had above average rainfall.

Kaufmann and Snell (1997) provided the first paper in the economics literature to employ a dynamic crop model in an assessment of the impact of weather on yield. The authors used the phenology subroutine of the CERES MAIZE model along with data on planting dates, cultivars planted and daily temperature data to estimate the dates during which corn was, on average, in each of its phenological stages. Estimates of phenological dates were made for the census years 1969 through 1987 for the eight to ten largest corn producing counties in each of eight states.[49] The authors then calculated seven climate variables for each of eight phenological stages, including: total precipitation, minimum and maximum temperatures during the

---

[48] The authors presumed that the previous year's moisture conditions show "…the net impact on yield contributed by carryover from the previous year's soil moisture (p. 733)."

[49] The included states were: Illinois, Indiana, Iowa, Missouri, Minnesota, Michigan, Ohio and Wisconsin.

period, average maximum temperature, average temperature, and maximum daily rainfall. These terms, along with their squares then formed a superset of variables for consideration. Some variables were removed to remedy collinearity problems, while others were removed because they were not statistically significant. In the end, some fifteen climate variables were chosen for inclusion, along with the length of two of the phenological stages and other socio-economic variables. One of the more interesting variables was the change in the amount of land planted in the county, which was found to be negatively related to the county's yield.

The approach used by Kaufmann and Snell (1997) offers some important insights, namely that the "[t]he effect of climate on yield depends on the phenological stage of crop development rather than the chronology of the human calendar (p. 179)." Unfortunately the results are weakened by the process of selecting variables without apparent regard to the biology of the crop, so the objections made by Katz (1977) might still be directed at this study. Nevertheless, the consideration of biological time and the authors' point that it matters when intra-seasonal events occur was a major step forward.

3.1.3 Lessons from the Literature

The goal of the following chapter is to develop a tractable method that can account for the effect of weather on corn yields over a large geographic area and over a long period of time; this section will summarize some of the main points from the literature review in preparation for that work. Smith's (1904 and 1914) conclusion that July precipitation is the most important determinant of corn yields has rippled through the literature for nearly a century. But, the variable is only useful because July happens to coincide with the typical date of silking in the Corn Belt. Since most studies consider only the Corn Belt states, the variable may indeed work well on average; a regression of national yield may also find the variable useful given the concentration of corn production in the Corn Belt. But, such a procedure would be insufficient to explain county-level yields across the entire nation insofar as silking dates differ across space (and across seasons).

Over 90 years ago, Wallace (1920) concluded that one likely could not select raw weather variables that could directly explain yields over multiple states, and that each county might require a different set of variables. The study by Mattice (1931) confirmed this intuition since, in order to explain yields for several states, the author needed to use a different set of variables for each state. It seems therefore that regressing yields on raw weather variables is likely to produce unsatisfying results. It is also difficult to specify *a priori* what the expected signs or magnitudes of the estimated parameters should be, so the model results would be difficult to assess.

Katz's (1978) point that the "statistical black box" approach may yield spurious and uninterpretable results shall be heeded. A similar issue arises in the indexing approach of Stallings (1960)—Plaxico's (1961) objection that the indexes are essentially relabeling unexplained variation as an explanatory variable is difficult to set aside. Regression of yield on raw weather variables and creation of weather indexes from residuals suffer the same problem: they show the extent to which various factors correlate with yield but, they are unable to explain yields.

Many studies highlighted the importance of the level of spatial aggregation. Shaw (1964) most clearly articulated the point, noting that averaging of weather variables across space might lead a study astray, especially insofar as crops do not respond linearly to the weather. Several studies found that spatial aggregation increases the correlation among variables, as Gehlke and Biehl had warned in 1934. There is little doubt that if the goal is to study the relationship between weather and yields, it is best to conduct the analysis at the lowest possible level of spatial aggregation.

Chapter 4 Determining Whether Weather Matters

Samuelson (1947, p.316) noted that "…there is nothing sacred about the conventional boundaries of economics; if the [business] cycle were meteorological in origin, economists would branch out in that direction..." Weather is important to, among other things, agricultural production and productivity, and economists have not shied away from the subject. This chapter will push the boundaries of the literature described in the previous chapter to develop a new location- and time-specific "corn suitability indicator" for use in economic productivity assessments. Rather than viewing weather as a disturbance to the status quo that must be "corrected for" (or as simply part of a residual), meteorological variables will be conceived and treated on par with the usual management variables.

4.1 *A County-Level Long-Run Weather Dataset*

To enable the forthcoming assessment, it was necessary to develop a long-run database of weather at the county level. Minimum, maximum and mean monthly temperatures for 1901 through 2005 were obtained from the Climate Research Unit (CRU) at the University of East Anglia. The data were provided on a monthly 30 arc minute grid for the globe. Mitchell and Jones (2005) describe the methodology used in version 2 of the CRU data; no publication is currently available describing the version 3.0 data used in this assessment. Maximum and minimum temperatures for 2006 and 2007 are from NOAA's U.S. Daily Minimum Temperature Analysis, and were provided on a daily 30 arc minute grid for the United States (NOAA-CPC 2011 a, b and c). To maintain compatibility with the CRU data, the daily NOAA data were temporally aggregated by averaging the daily minimum and maximum temperatures for each month for each grid cell. Precipitation data for 1900 through 2007 were obtained from the Center for Climate Research at the University of Delaware (Matsuura and Wilmott 2009).

Gridded weather data were not available for the crop years before 1900. Data for these years were derived using weather station data from NOAA's Global Historical Climate Network (GHCN) (NOAA-NCDC 2007). Some work was necessary to convert the weather station data into gridded data. Broadly, the goal was to establish the relationship between the CRU grid data and the available weather stations. From a set of acceptable weather stations,[50] the four closest stations to a given cell were selected, and the cell values for 1901-2000 were regressed (to the median) on the four station values. The estimated parameters were then applied to the station values to derive the pre-1900 grids. The same procedure was applied separately for the minimum, average and maximum temperature.

It was somewhat more challenging to estimate reasonable precipitation grids since precipitation is much more spatially variable and does not tend to change along a gradient as does temperature. As a result, station values were much less reliable in estimating gridded precipitation. To overcome this, information was added by using the station and pixel deviations from their long-run mean values. The pixel

---

[50] Only U.S. weather stations with data for each month beginning January 1888 and ending December 2000 were included in the analysis, a total of 136 stations. The station density was highest in the eastern half of the country, including most of the major corn producing regions of the period. The reader might consider the low station density when interpreting the estimated weather values and the results that incorporate them. Results for census years after 1899 are not affected by this estimation procedure.

deviations were then estimated by inverse distance weighting the station precipitation deviations. Recognizing that the geography of precipitation is quite variable, the power applied to the inverted distance was selected for each month by performing a grid search over a range of possible values,[51] then choosing the one that achieved the minimum root mean squared error when performing the same weighting of the available data.

After assembling annual grids of temperature and precipitation, the data were aggregated to the county level by averaging the cell values in each county weighted by each cell's area in corn production. Thus, the weather data are specific to corn production, and cannot be properly applied to other crops (the aggregation procedures were similar to those described in Sections 2.2.1 and A.2). Finally, while gridded weather data were estimated for the 1879 crop year, there were not enough acceptable weather stations in the data to yield reliable results. Thus, that year will not be included in the forthcoming analysis.

4.2 *Movement-Induced "Effective" Climate Change*

The potential effects of climate change on agricultural production have been extensively studied. The results generally show that U.S. agricultural productivity will, on average, decline due to increased temperatures. Mendelsohn, Nordhaus and Shaw (1994) pointed out that most studies fail to account for farmers' ability to adjust to changing situations, implicitly employing a "…dumb farmer scenario (p. 753)."

---

[51] The range of weighting values was allowed to vary from one to six in steps of 0.25. The selected values were: 2.5 for July, August, September and November, 2.75 for October, March, May and June; 3.0 for December and February, and 3.25 for January and April.

The authors partially addressed this deficiency by directly estimating the effect of weather on farm land rents (and revenue), thus allowing land to be repurposed as climate regimes change. The results showed smaller climate change impacts than have generally been predicted. Unfortunately, the assessment was essentially non-spatial in its consideration of climate change: a spatially uniform average temperature and precipitation change was assumed. By most accounts, the largest temperature and precipitation changes will be concentrated in certain areas, and some of the largest changes are expected in areas that are not major agricultural areas. It is therefore difficult to interpret their results.[52]

Lobell and Asner (2003) assessed yield trends of a selection of counties for the period 1982-1998 by estimating a regression of the form $\Delta Y = m + r\Delta T$, where Y represents the county yield of either corn or soybeans and T represents the average of daily June through August temperatures for the county. Temperatures were interpolated to the county level from a 2.5 degree grid.[53] The counties chosen for inclusion in the regression were those that exhibited a negative correlation between yield and the temperature variable. Temperatures fell over the period in the majority

---

[52] Mendelsohn, Nordhaus and Shaw (1994) have been criticized on a number of other grounds, most of which are not particularly relevant to the present discussion. For example, see Darwin (1999) and Quiggin and Horowitz (1999), and the corresponding replies by Mendelsohn (1999a and b).

[53] Cells in 2.5° grid are about 77,000 km^2 at the equator; by comparison, the *largest* county in the contiguous United States has an area of about 52,000 km^2 (with the average being about 2,500 km^2). The authors also investigated the effect of precipitation changes, finding no significant relationship between precipitation and yield. Given the discussion in Section 4.1, it seems unlikely that interpolation of such a large grid would produce reliable county-level precipitation estimates, much less precipitation estimates that are relevant to the cropping areas of the counties. The authors' conclusion regarding precipitation might have been different given higher resolution data.

of the counties selected for the regression, and therefore Lobel and Asner find that temperature changes have had a positive impact on crop yields in those counties. The use of detrended yields and raw temperature values for a subset of the season has a long history, and the objections of Katz (1977) are applicable to this study—it is not at all clear if one can apply the Lobel and Asner results to other counties, or even to the same counties if the temperature deviates far from its mean, and exclusion of precipitation changes as an explanatory variable implicitly assumes that precipitation is either unimportant or that it is invariant.

It is not clear what the result would have been had all counties been included in the analysis. Gu (2003) charged that Lobel and Asner had predetermined the result by their inclusion of only counties that showed a negative yield response to temperature. Lobell and Asner (2003) responded, noting that the result is largely the same if the procedure is applied to the entire United States, and that the selection of counties was intended to avoid biasing the result by including counties with different climate trends.

Lobell and Field (2007) assessed the relationship between climate change and yield of several crops globally. The authors disaggregated 1961-2002 annual country-level crop production data to a half-degree grid using the 1992 crop geography estimated by Leff, Ramankutty and Foley (2004). Since the Leff, Ramankutty and Foley data are provided for only a single year, Lobell and Field's estimates do not account for any spatial change *within* countries over the period. To derive the average temperature and rainfall for a crop, the authors used a "global growing season" for the crop, defined as the period during which precipitation and temperature were most highly correlated with yield.[54] Thus, for example, the global growing season for corn was determined to be July and August for all years and countries (in contrast to the finding in this assessment, where it will be shown that the growing season is different even within a country during a single year). The authors' methodology in assessing the relationship between weather and yield relies on a regression similar to that used by Lobell and Asner (2003), and is therefore similar to the early studies critiqued by Katz (1977).

In the case of corn in the United States, some climate change impacts may be offset by changes in the spatial concentration of production. The top charts in Figure 17, panels a, b and c show, respectively, the area-weighted average March, June and September temperatures for the 1909 U.S. corn production area (solid line), evaluated at the current year's temperatures. The dashed line shows the weighted average temperatures for the area of the current year.[55] Thus, the extent to which the solid line is above the dashed line shows how much changes in spatial concentration since 1909 have resulted in decreased average corn growing temperatures, as quantified by

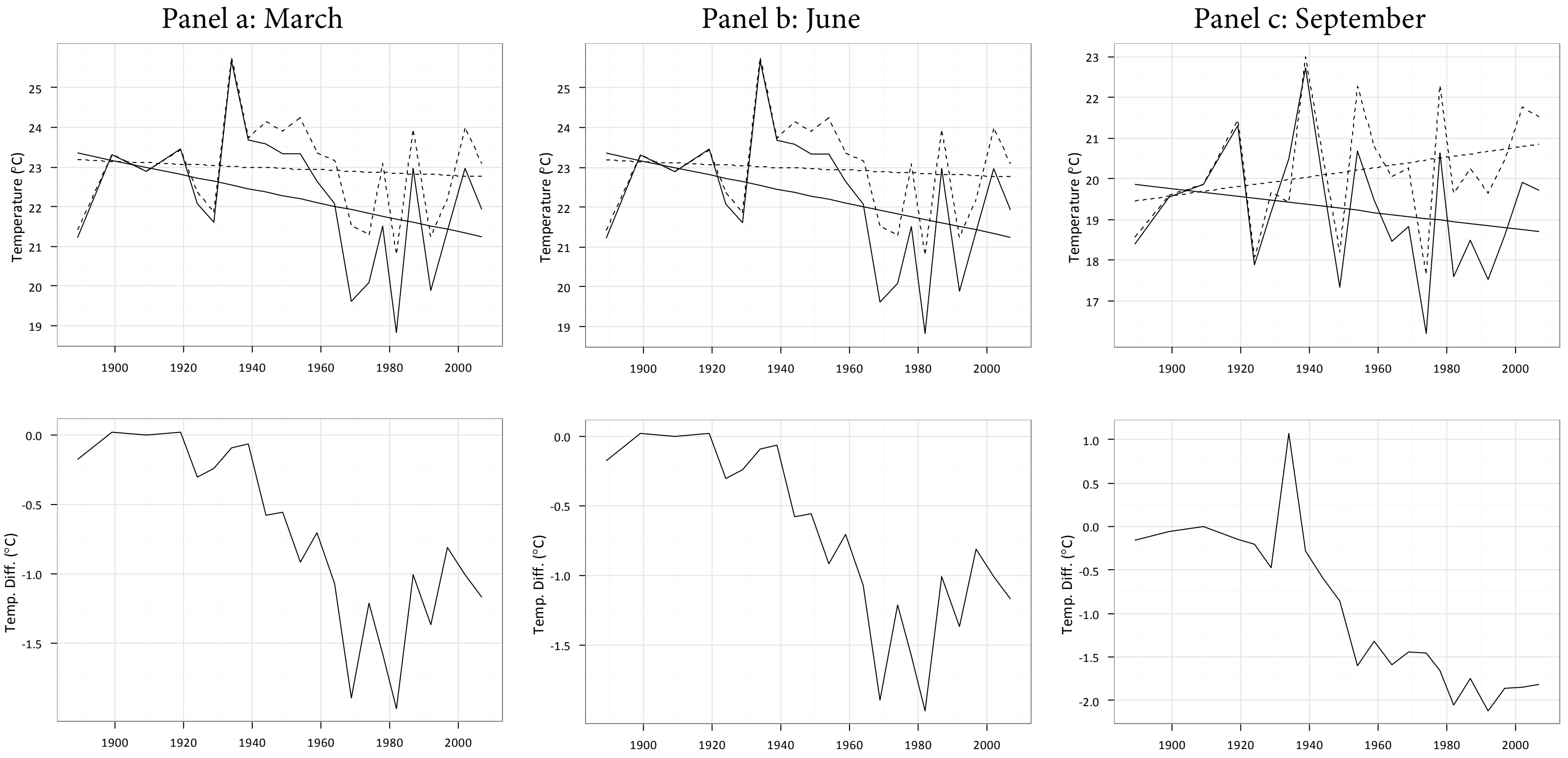
---

[54] The authors "…defined an effective 'global growing season' for each crop based on the contiguous months within the growing seasons for the major growing regions…that produced the highest model R^2 (pp. 1-2)." However, the model from which this R^2 is derived is not clear from the context—here it is assumed to be the yield-weather model defined later in the paper. While the authors determined that their results were not very sensitive to the choice of the growing season, the use of a single global growing season renders the results difficult to interpret. For example, July and August, the "global growing season" for corn, are often important months in the major corn producing regions of the world's largest producers, China and the United States. Thus, the assessment is similar to, and is subject to the same criticisms as, the earlier referenced U.S. studies that applied data for important months in the Corn Belt to the entire country. That July and August are winter months in the southern hemisphere is particularly concerning.

[55] The "current year" is allowed to change. So, for example, the value for 1934 uses the 1934 area, while the value for 2007 uses the area of 2007.

the bottom chart in each of the panels. Depending on the year, the spatial movement has resulted in a 3-4°C drop in average March growing temperatures, a 1-1.25°C drop in June temperatures, and a nearly 2°C drop in September temperatures. Mendelsohn et al. assert that the average anticipated temperature increase due to climate change will be about 2.8°C, so it appears that the change in the footprint of production may substantially, if not fully, offset the temperature effects of climate change.

Figure 17: Average Temperature Changes Due to Spatial Movement



Panel a: March

Panel b: June

Panel c: September

Source: Derived using the data described in Sections A.2 and 4.1.

Notes: The top chart in each panel shows the average temperature for the 1909 area (solid line) and the current-year area (dashed line). The bottom chart in each panel shows the absolute temperature difference between the 1909 corn area and the current-year corn area

4.3 *The Effect of Weather and Technology on Yield*

The previous section of this chapter showed that, on average, monthly temperatures in corn growing areas have decreased due to relative changes in where corn is grown. However, the literature gives little guidance on the potential impacts of these changes. For example, although the average June temperature under which the average corn plant grows has decreased, the impact of a given monthly temperature change is different at different locations. To sort out the effect of temperature and moisture changes it is necessary to consider how the crop interacts with the weather. In addition, one reason why the impacts of calendar-based weather variables differ across locations is because the growing season differs across space. The start of the growing season (the planting date) is chosen by farmers, and the length of the growing season varies according to the local weather in each year. This section proposes a set of methods for determining the effect of weather on corn based on the biology of the crop along with assumptions about how the weather affects farmers' planting decisions.

4.3.1 The CLIMEX Model[56]

The CLIMEX model (Sutherst and Maywald 1985) and the commercial modeling software in which the model is implemented (Sutherst et al. 2007) are widely used to assess the potential for occurrence of pests and diseases. CLIMEX is a simplified computer model that infers the response of a species to climate from its geographical distribution and seasonal patterns of growth and mortality in different

---

[56] This section draws on Beddow et al. (2010b).

locations. In the CLIMEX modeling framework, the aim is to define the regions that a species could potentially occupy by characterizing the response of the species to various climatic factors.

The model has two primary outputs: a Growth Index and an Ecoclimatic Index. The annual Growth Index (GI) is a summary of the potential for a given organism to thrive at a location, absent any stressors (heat, cold, wet, and dry). Conceptually, the annual and weekly Growth Indexes summarize the species' growth response to temperature and moisture. A second output, the Ecoclimatic Index, represents the ability of a species to establish at a given location.

Both indexes refer to the species' population responses to local climate conditions and the species' phenology. Ecoclimatic Index values are calculated as the product of the annual Growth Index and indexes of the species' response to various stressors. Most commonly, the Growth Index is used to determine the extent to which a region is suitable for an organism within a favorable season, while the Ecoclimatic Index indicates the ability of an organism to establish in the region (that is, the ability of the organism to survive climatic extremes, usually during overwintering or oversummering).

The CLIMEX model suggests a particularly attractive and practical methodology for estimating the effect of weather on crop yields.[57] Namely, by specifying the ranges

---

[57] Other modeling platforms were also carefully considered, including ECOCROP and DSSAT. ECOCROP (Hijmans et al. 2005) is a simple mechanistic growth model supplemented by a database of parameters of the same name (FAO 2011). The ECOCROP modeling approach is similar to that of CLIMEX. However, CLIMEX has been more widely applied (and therefore tested and reviewed) and incorporates somewhat more detail (for example, ECOCROP relies on rainfall, not soil moisture). DSSAT (Jones et al. 1998) implements a variety of crop models including the CERES Maize model

of temperature and moisture that are most suitable to the growth of corn, one could interpret the annual growth index as an indicator of the suitability of a given climate in a location-year for growing corn. Since corn is planted each year, and therefore is not expected to overwinter, the Ecoclimatic Index is of less interest. Concepts from the CLIMEX modeling approach will be employed in developing a model of corn yields.

4.3.2 An Intra-Seasonally Dynamic Corn Suitability Indicator

Phenology

Corn goes through a number of distinct phenological stages, beginning with pre-emergence and ending at physical maturity.[58] Broadly, these stages can be divided into the vegetative stages and the reproductive stages. The reproductive stage begins when the silks (stigmas) form and continues through physiological maturity.

The phenological model begins by using a simple rule to estimate the planting date. In order to assure that soil is thawed, planting cannot occur until at least ten growing degree days have accumulated (using a base of 10°C). In many areas, this provides a reasonable approximation of the planting date, but in some areas this would result in planting dates much earlier than are observed. To correct for this,

---

(Jones and Kiniry 1996) and IXIM (Lizaso et al. 2011) for corn. Although DSSAT is intended for field-level assessments, the system has been applied at higher levels of spatial aggregation (see Pardey et al. 2008 and the references therein), although with some objection. However, the "minimum data set" requirements needed to run DSSAT (see Hunt and Boote 1998) exceed the data available for a long-run, national assessment. Other crop models (WOFOST, APSIM) suffer similar issues.

[58] See Neild and Newman (1986) for a short introduction to corn's phenological stages, or Hicks and Thomison (2004) for a more detailed introduction.

planting is not allowed until day 61 of a Julian calendar, which is the earliest date at which corn is usually planted in the United States (see USDA-NASS 1997). In order to avoid soil compaction and getting their machinery bogged down in the field, farmers do not typically plant when the soil is fully saturated or when there is standing water on the field, so the model assumes that planting does not occur at a location until the estimated fractional soil moisture is below 100 percent, resulting in later estimated planting dates in years that have significant rainfall, or when there is a large snow pack to melt.[59] The model assumes that emergence occurs about 69 degree-days (base 10°C) after planting, or 30 calendar days, whichever is shorter.[60]

Dates for silking and physiological maturity are also estimated using a degree-day model, employing a beta function (Figure 18) to characterize the daily growth response to mean daily temperature, using parameters suggested by Streck et al. (2008). The beta function, *f*(), accepts a temperature argument, *t*, and returns $f(t) \in [0,1]$, thus providing a scaling factor for the plant's growth response that is equal to zero at a given minimum and maximum temperature, and equal to one at a given optimal temperature. The plant's phenological response to temperature is

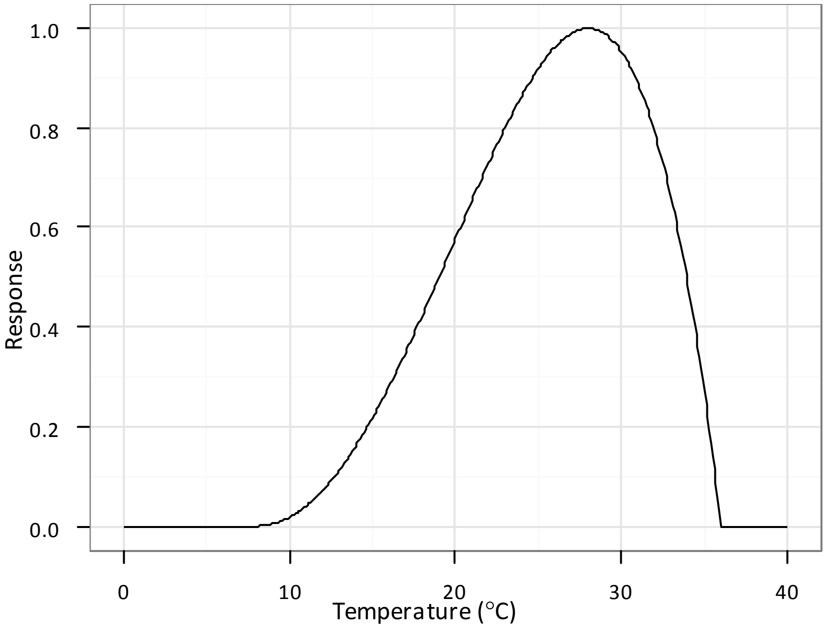
---

[59] Artificial drainage has certainly affected corn production. For example, Beauchamp (1987) notes that in its natural condition, much of the fertile land in important portions of several Corn Belt states "…was either swamp or frequently too wet to farm (p. 32)." Artificial drainage helps farmers avoid delays in planting and harvesting that result from saturated surface soil layers (Fausey, Doering and Palmer 1987). However, few estimates of the extent and location of drainage are available—census data for 1920, 1930, 1969 and 1974 report about 40 to 60 million drained acres over all crops (with no discernible trend), but these estimates are not considered reliable (Jaynes and James 2007). Absent long-run, county-level data on drainage on corn acres, this element could not be incorporated into the present model.

[60] This calculation actually used 125 Fahrenheit degree days (with a base of 50°F).

represented by a growth rate, $r < 0.1$, so that $r \cdot f(t) \leq 0.1$ is the daily growth response. The vegetative state is completed when $\sum r \cdot f(t) = 1$, at which point the accumulated degree days are reset and the reproductive state begins. The algorithm begins with the assumption that a slow-growing variety is planted (r = 0.0254), and continues increasing the growth rate until a plant can complete a season at the location.

Figure 18: Phenological Response to Temperature



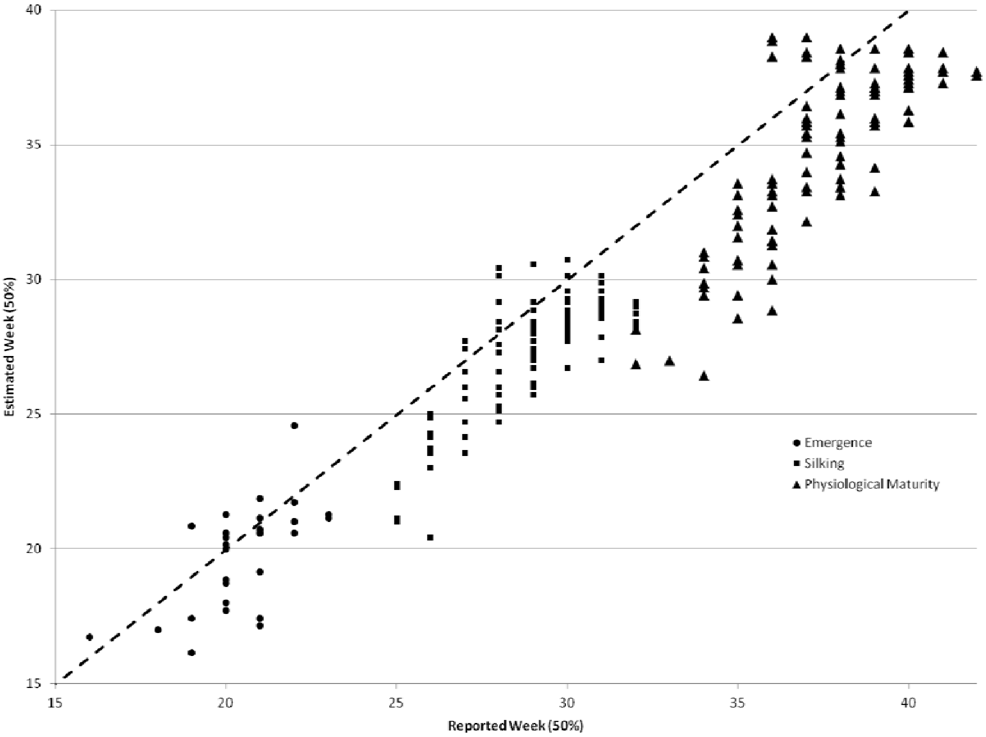
Source: This figure reproduces a portion of Figure 1 in Streck et al. (2008, p. 451).

Notes: The curve shows the daily developmental response of corn to daily temperature. A minimum temperature of 8°C, an optimal temperature of 28°C and a maximum temperature of 36°C were used to derive the growth response curve.

USDA-NASS (2011b) provides dates at which corn emerged, silked and reached physiological maturity for several states. Silking and physiological maturity dates are available beginning in 1981 and emergence dates are available beginning in 1999. The data represent the week during which half of the planted acres in the state reached each physiological milestone. To assess the physiological model presented above, the date at which half of the counties in each state reached each phenological stage was estimated for each census year for which comparable USDA crop progress reports were available. While these values are not directly comparable with the USDA reports, a comparison of the series should at least exhibit a high degree of correlation. Figure 19 shows the reported week of emergence (circles), silking (squares) and physical maturity (triangles) by state for census years 1982 through 2007, plotted against the week estimated by the model. Each point represents the week during which a given physiological event occurred in a given state-year. The figure reveals that the estimated dates are highly correlated with reported dates, but that the model, on average, predicts that phenological events occur earlier than the dates reported by USDA.

It is not clear whether the discrepancy between estimated and reported dates should be of concern. The state-level USDA data reveal that many disparate states are reported to have phenological events occurring during the same week, raising some questions about the source data. Perhaps more importantly, it is not clear whether one should expect the date at which half of the acres in a state are beginning a given phenological state to be similar to the date at which half of the counties (weighted by acreage) have silked. The estimates appear to be reasonable, especially since they were derived directly from county-level weather without any information on actual planting dates (as is necessary for consistent treatment of data extending back over a century).

Figure 19: Estimated and Reported Phenological Dates by State, 1982-2007

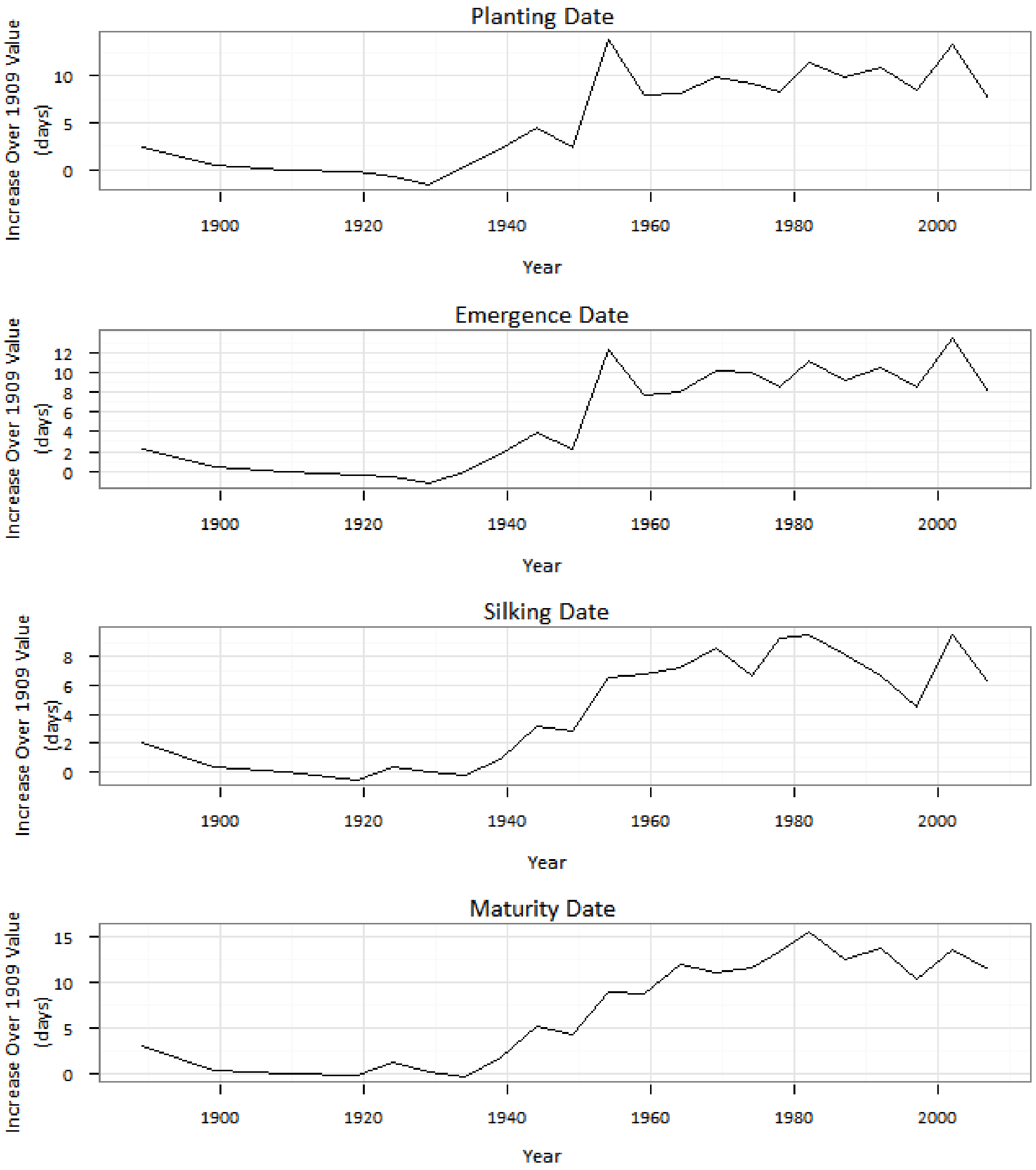


Source: The estimated week is derived using the data described in Sections A.2 and 4.1. The reported week is from USDA-NASS (2011b).

Notes: The estimated week is the Julian week number during which corn planted in fifty percent of the counties in a state were estimated to have reached the indicated phenological stage. The reported week is the week during which fifty percent of the acres in a state were reported to have reached the indicated stage.

The results show that the changing spatial footprint of corn production has affected the average date at which the nation's corn crop is in the various phenological stages. Since weather varies from year-to-year, it is difficult to directly assess changes in phenology over time. To get at the pure effect of spatial movement, the 1909 weather was applied to the spatial footprint of production during each census year. The result of these calculations is presented in Figure 20, which shows that all phenological stages are delayed relative to when they would have occurred had the 1909 footprint of production not changed: planting, emergence and silking now occur, on average, about a week later than in 1909, and physiological maturity is now delayed by ten days to two weeks.

Figure 20: The Effect of Spatial Movement on Average Phenology



Source: Calculated using the data described in Sections A.2 and 4.1.

Notes: The lines represent the estimated area-weighted average Julian day on which the indicated phenological event occurred in the current area less the Julian day on which the event occurred in the 1909 area (always estimated using the weather for the current year). Thus, the values can be interpreted as the number of days by which the phenological event was delayed due to the spatial reallocation of the corn area that occurred between the current year and 1909.

Soil Moisture Index

To account for moisture, the CLIMEX model uses a simple soil-moisture model based on that of Fitzpatrick and Nix (1969). In the model, soil moisture is allowed to increase up to the soil's moisture holding capacity and moisture decreases according to an evapotranspiration coefficient. While the model does enable a simple accounting of rainfall, local soil and climatic conditions are not otherwise taken into account—all soils are assumed to have a moisture holding capacity of 100mm, and evapo-transipiration is held constant. Nevertheless, the use of a soil moisture sink, instead of using precipitation (and perhaps temperature) variables to directly estimate the impact of moisture on an organism is a substantial improvement.

In this assessment, monthly soil moisture is estimated for each location using a slightly modified version of the Thornthwaite (1948) model presented by McCabe and Markstrom (2007).[61] The model begins by partitioning precipitation into rainfall and snow, where all precipitation falls as rain when the mean monthly temperature is above 3.3°C, all precipitation falls as snow for temperatures below -10°C, and rain and snow are mixed between these temperatures (see the citations in McCabe and Markstrom for justification of these temperature thresholds). Snowfall is accumulated in a snow sink until the mean monthly temperature is above -10°C,

---

[61] The CGIAR Consortium for Spatial Information (CGIAR-CSI) developed a soil water balance dataset (Trabucco and Zomer 2010). The soil moisture model used to develop the Trabucco and Zomer dataset was tested for this assessment, but the McCabe and Markstrom's model was ultimately selected because of its ability to account for snow storage, an important consideration in many U.S. corn growing regions. In most circumstances, the modeled results were similar, however the McCabe and Markstrom model generated higher estimates of early-season soil moisture in locations that had significant snowfall, resulting in more realistic delayed planting date estimates in many areas of the Northern Plains and Corn Belt.

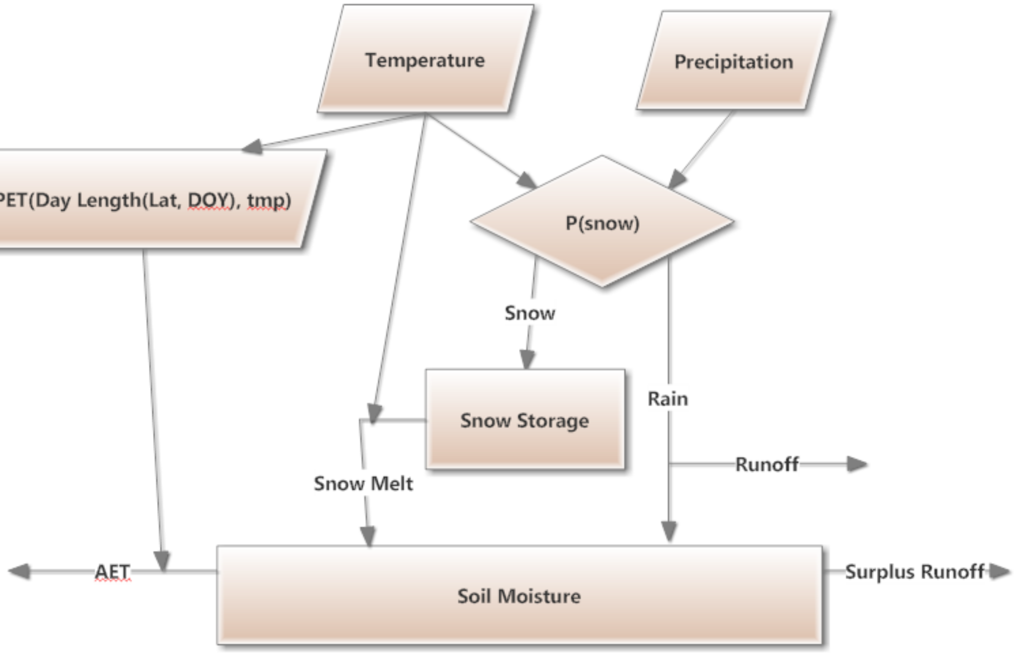
when snow begins to melt at a rate determined by the difference between the mean monthly temperature and the snowfall temperature threshold. To account for insulation effects, the model does not allow more than 50 percent of the snow held in the snow sink to melt in any given month.[62]

Snow melt and rainfall feed into the soil moisture sink up to a soil moisture storage capacity, set equal to 150mm for all locations as suggested by the authors. Water is removed from the soil moisture sink by three primary mechanisms (actual) evapotranspiration, surplus runoff and direct runoff. Direct runoff is roughly estimated as a constant five percent of rainfall for all locations, and surplus runoff is the amount of snow melt and net rainfall (after runoff) in excess of the soil's moisture holding capacity. Actual evapo-transipiration (AET) is a fraction of estimated potential evapo-transipiration (PET), constructed such that there is less actual evapo-transipiration as a fraction of potential evapo-transipiration as soils become drier.

---

[62] As a result of the percentage limit on snow melting, the snow storage sink never completely empties. However, this has little effect on the soil moisture estimated by the model.

Figure 21: Stylized Diagram of the Water Balance Model

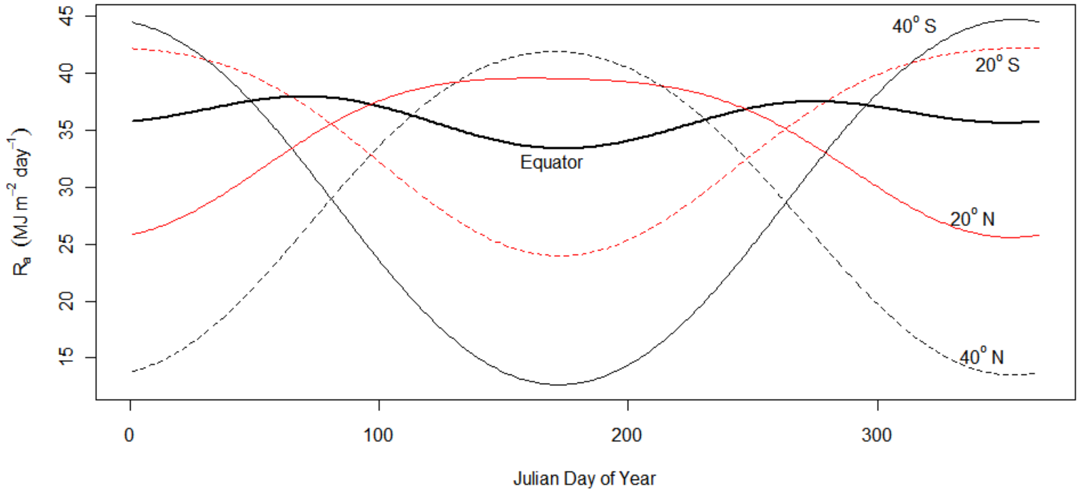


Source: Modification of Figure 1 in McCabe and Markstrom (2007).

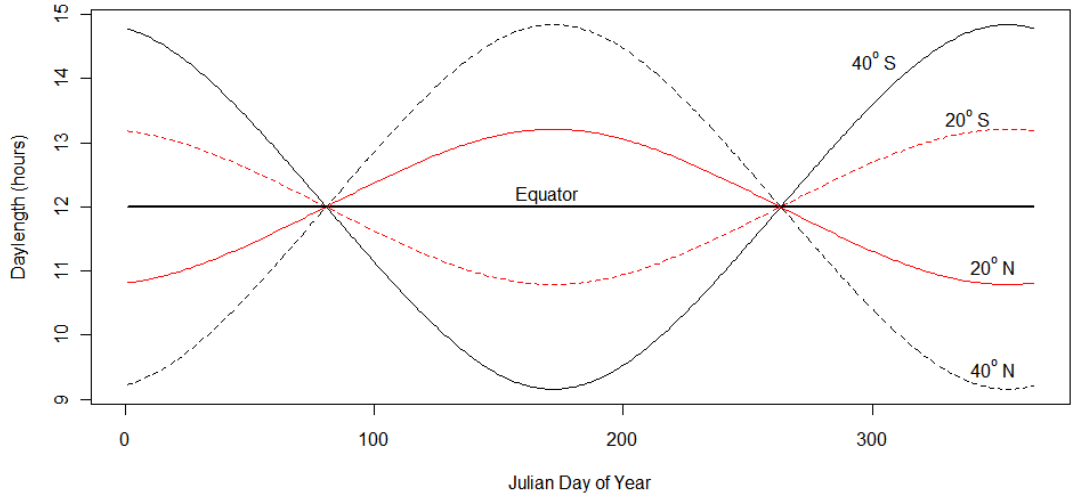
Potential evapo-transipiration is calculated using the Hamon equation (Hamon 1961), which estimates PET from the number of daylight hours (derived using the day of year and the latitude) and temperature. Another common specification, the Hargreaves equation, estimates PET from estimated extraterrestrial radiation. The latter formula tended to give higher PET estimates than the Hamon formula, although the values were highly correlated ($\rho = 0.95, p < 0.01$). However, there was no reason to suppose that either formula was preferable, so the Hamon specification used by McCabe and Markstrom (2007) was implemented in the present model, using a formula for day length estimation provided by Allan et al. (1998). Figure 22 shows the estimated solar radiation (panel a) and day length (panel b) that were calculated as intermediate steps in the two PET specifications.

Figure 22: Solar Radiation and Daylength by Latitude and Day of Year

Panel a: Solar Radiation



Panel b: Daylength



Source: Panels a and b are reproductions of Figures 13 and 14, respectively, in Allen et al. (1998) and were created using formulae provided therein.

The soil moisture values are calculated for each location using a daily time series of splined monthly precipitation values and linearly interpolated monthly minimum and maximum temperatures. The soil moisture model was calibrated by running it for the year preceding each census-year so that the January starting values would be reasonable.

It is not expected that soil moisture should correlate directly with crop yields as there are various dynamic effects. Thus, the daily soil moisture values were used to calculate a moisture index value for each phenological state in each location-year. In the pre-emergence stage, too much soil moisture is often cited as being more problematic than low moisture. However, since planting dates are delayed until the soil moisture is less than 100 percent, this problem is already taken into account and soil moisture values are not used for this phenological stage.

There is little information on which to base the transformation of the soil moisture variable into a moisture index value. Morgan, Biere and Kanemasu (1980) estimate a piecewise linear function for corn's response to soil moisture in the reproductive stages using data from Kansas. While the data on which the estimates are based is limited, these values seem reasonable in light of a number of extension publications and the practical knowledge of farmers. The index values estimated by this function equal zero when soil moisture equals zero, but increase quickly until soil moisture is at one-third capacity, such that the index equals 80 at that point. The rate of increase in the index is much slower after one-third soil moisture is reached, although it increases somewhat after the soil is two-thirds saturated. In light of diminishing returns, the latter increase may seem unlikely, but the soil does give up its moisture much more easily when it is near saturation than when it is not.

Various parameterizations were attempted before the Morgan, Biere and Kanemasu function was chosen. Reference to numerous extension publications did not shed much light on the appropriate parameters as these usually refer to rainfall and not soil moisture. A grid search over possible soil moisture values was conducted, and it would be possible to select seemingly reasonable parameters that would appear to explain more of the variation in yields than does the chosen function. There was no evidence that high levels of soil moisture correlate with reduced yields, indeed the contrary appears to be true. Thus, all of the best-fitting transformations increase with soil moisture and therefore increase in a weakly monotonic fashion with the selected transformation. Absent solid evidence that a different transformation should be used, one runs the risk of repeating the errors of the Stallings-type weather indexes—namely, artificially renaming unexplained variation as explained variation. Thus, despite their limitations, the Morgan, Biere and Kanemasu parameters were applied to the soil moisture in the vegetative, silking and reproductive stages to derive daily soil moisture indexes.

The resulting index values were then averaged over each phenological stage for each county-year as an indicator of the average conditions during that stage. While it was tempting to attempt a different aggregation (e.g., account for daily dynamics and cumulative effects during each stage), this did not seem justifiable since the daily moisture and temperature data were estimated from monthly data. Given more temporally disaggregated weather data, other aggregations might be possible, but no such data are available for the entire period under analysis.

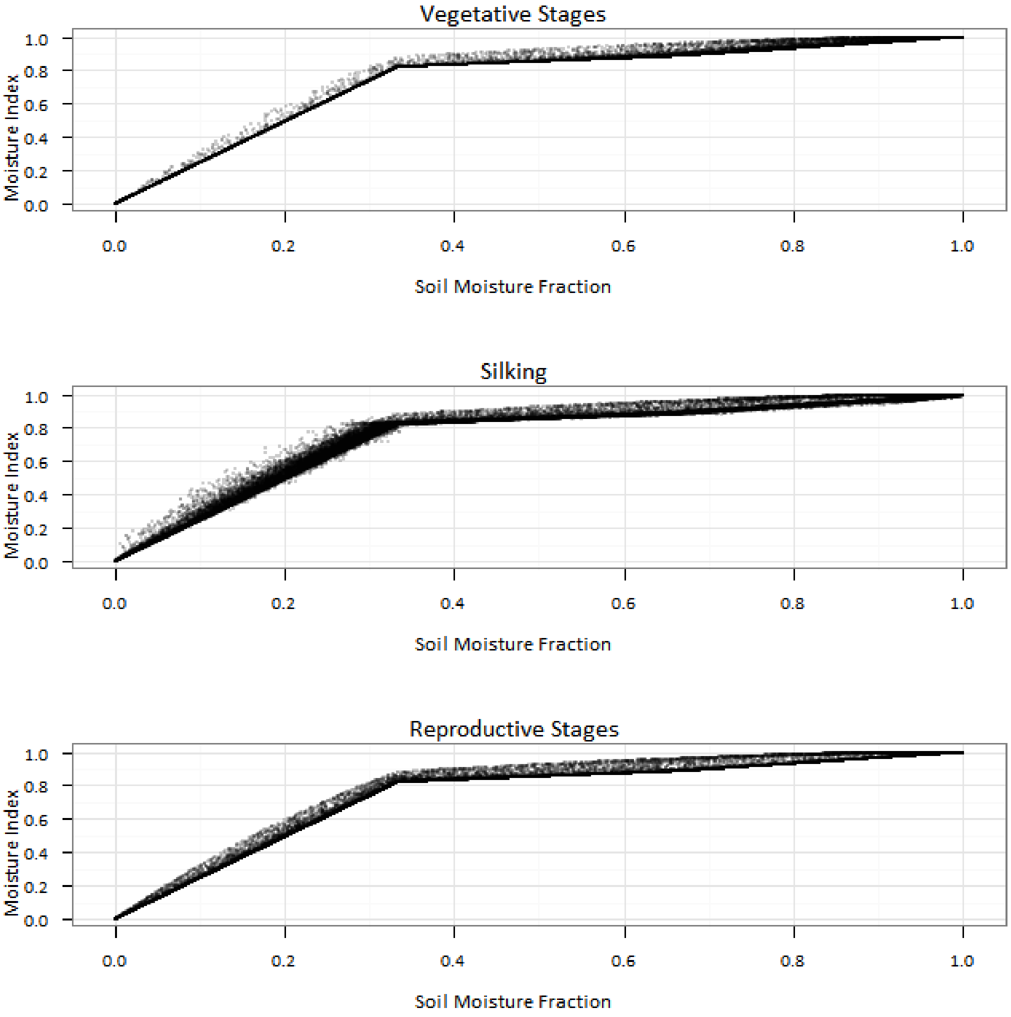
It was further assumed that irrigated acres are maintained at a soil moisture that maximizes the moisture index. For each county i in year t and phenological stage p, the total moisture index was derived from the fraction of acres that were irrigated (*irrig*) and the soil moisture index for non-irrigated acres as $MI_{ipt}^{Tot} = irrig_{it} + (1 - irrig)SMI_{ipt}^{Non}$. This is unrealistic given that the specification used implies that the maximum soil moisture index is achieved when the soil is fully saturated. The economically optimal irrigation level is likely lower, and even where irrigation water is free, there are costs associated with pumping it.[63] However, this assumption is likely to have little impact given that the index value changes very little after the soil is at a third of its capacity.

The calculated soil moisture index values are shown in Figure 23. Each dot represents a county-year, and the dots are semi-transparent so that darker portions of the chart indicate that more counties are represented in the same location of the chart. Although there are three linear segments of the soil moisture curve, some points lie above the curve due to the correction for irrigated acres.

---

[63] Assessment of the water price responsiveness of farm-level irrigation use is complicated by the multitude of responses available to farmers. For example, in response to an increase in the cost of pumping or irrigation water, land may be reallocated to less water-demanding crops (Moore et al. 1994). Further, use of irrigation water is thought to be more sensitive to the price of the crop than to the cost of irrigation water.

Figure 23: Soil Moisture Index Values



Vegetative Stages

Silking

Reproductive Stages

Source: Derived using the data described in Sections A.2 and 4.1.

Notes: Each point represents the calculated moisture index for a phenological stage in a given county-year. The specification of the moisture index is tri-linear. However, some points are not on the linear segments due to adjustments for irrigation.

Temperature Suitability

The suitability of the temperature regime for each county-year was assessed separately from the phenological response to temperature. The plant's growth response to temperature was modeled using the estimated daily high and low temperature. Following the CLIMEX methodology, the daily temperature cycle is represented by a sine curve, mapped so that the minimum temperature occurs at midnight and the maximum temperature occurs at noon. Degree days are accumulated by integrating the sine curve between (potentially non-binding) lower and upper temperature thresholds using formulae provided by Baskerville and Emin (1969) and Zalom et al. (1983).

The temperature module requires four growth response parameters, T_0, T_1, T_2 and T_3. The lower and upper temperature limits for plant growth are defined by T_0 and T_3, respectively, while T_1 and T_2 define the optimal temperature range. The temperature index is then derived for each day using a sine curve fitted to the estimated maximum and minimum daily temperatures (see Sutherst and Maywald 1985 for more details).

Temperature indexes are specified separately for each phenological stage. A grid search over plausible temperature parameter sets was conducted to find the set of temperature parameters that yielded location- and year-specific temperature indexes that were most highly correlated with yield. This assessment permitted the optimal temperature parameters to vary across years to test whether varietal change had systematically affected optimal growing temperatures. While optimal parameter values differed from year-to-year, they did not vary in any systematic way, giving little justification for the use of different parameter values for each year. The only noticeable trend was that the maximum achievable correlation between the indexes and crop yields has decreased over time, likely because corn varieties have developed largely in a way that ameliorates the negative effects of bad weather (Shaw 1964).

Corn growing degree days are usually calculated using a base of 10°C and an upper limit of 30°C, and those parameters were defined as the optimal range of the temperature index computed here. Contrary to the usual degree-day concept, temperatures above the optimal range are considered to be detrimental, so that the temperature index decreases by ten percentage points for each degree above the optimal range, until the daily index value equals zero at a temperature of 40°C. Below the optimal range, the index decreases by 20 percentage points for each degree under 10°C. As with the moisture index, the temperature index was calculated on a daily basis, then averaged for each phenological stage.

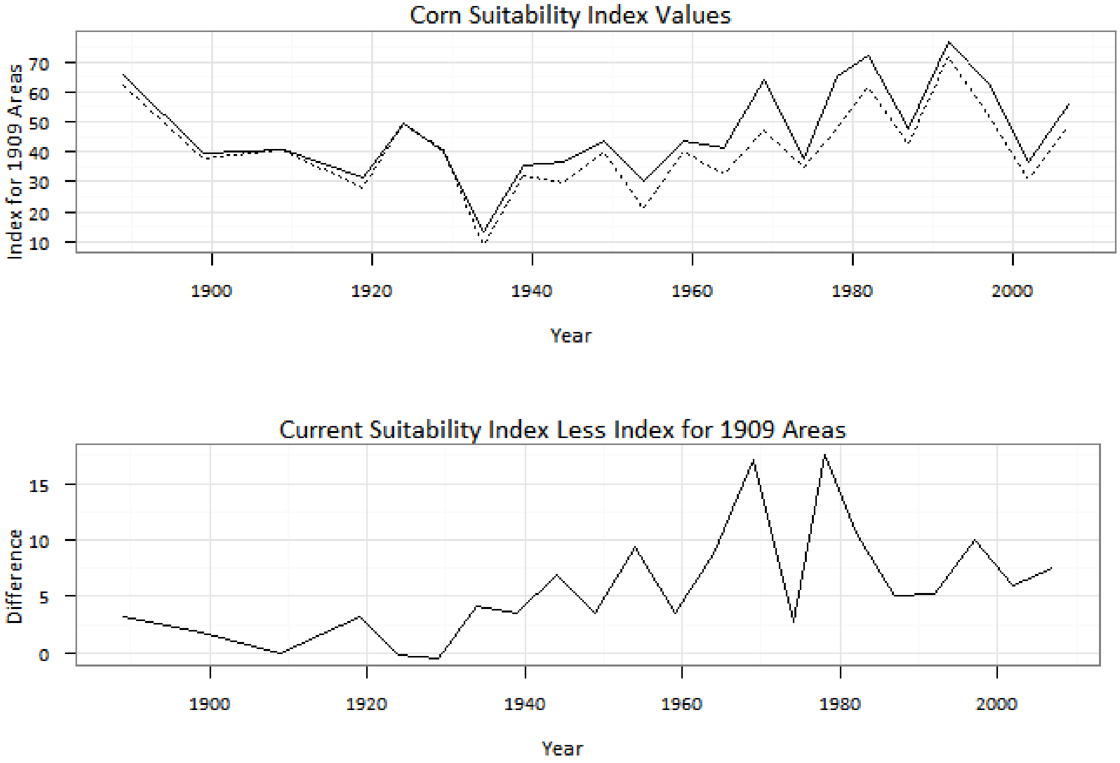
Climate Indicator

In-line with the previous research on the subject, it is desirable to calculate a location and time-specific indicator of corn suitability. However, it was not immediately clear how the various moisture and temperature indexes should be combined into an indicator of crop suitability. In keeping with the methodology of the CLIMEX model, the moisture and temperature indexes for each phenological state were multiplied to generate a single index per stage. Thus, for example, the effect of high temperatures can be somewhat ameliorated by good moisture conditions, but the deleterious effect of high temperatures is magnified when there is little moisture.

Corn growth is a dynamic process, and there is likely to be interaction between each stage of development. While there is evidence that moisture stress and temperature responses are additive across major phenological stages (Denmead and Shaw 1960; Schlenker and Roberts 2008), and the crop can recover from short periods of temperature stress, it is unlikely that the effects will be additive across long phenological stages. For example, it is unlikely that damage wrought by poor weather during the silking stage could be repaired by good weather during the reproductive stage. The index values for each phenological stage were therefore multiplied to create a single index of weather suitability for each location-year. When interpreting the climate indicator, it is important to keep in mind that it represents non-linear yield responses to temperature and moisture (even when used in an otherwise linear statistical assessment), thus avoiding many of the problems highlighted by Shaw (1964).

The top panel of Figure 24 presents the suitability values for each year, weighted by the current area (solid line) and the 1909 area (dashed line). While the indexes are positively correlated, the current suitability index is almost always higher than it would be if the crop had been grown in its 1909 footprint. By this metric, the spatial movement over the past century has improved corn growing conditions. The bottom panel of the figure shows the difference between the index values, where positive values indicate that the current-year's crop area had a better suitability than did the 1909 area. The values are positive with the exception of three years. In 1909, the difference is zero by design, and the 1909 area had a very slightly higher suitability score than did the current area in 1924 and 1929. The census years 1969 and 1978 are particularly evident in the bottom panel of the figure. Both years saw better than average weather (as indicated by the CSI) in the area of 1909 and in the current area. However, the current area exhibited CSI values that were about 40 percent above the long-run mean of that area, while the 1909 area had CSI values that were about 15 percent higher than normal.

Figure 24: Corn Suitability Index, 1889-2007



Source: Derived using the data described in Sections A.2 and 4.1.

Notes: The top panel shows the area-weighted CSI values for the current corn harvested area (solid line) and for the 1909 area (dashed line), always calculated using the current year's weather. The bottom panel represents the increase in the weighted average CSI values due to the spatial reallocation of corn area that occurred between 1909 and current year.

4.3.3 Accounting for Technology-Cum-Management: Covariates

It is clear from the literature review that when one wishes to understand the processes underlying yield, some accounting of technology must be attempted, but that a time trend is not appropriate. It is not important that hybrid varieties were not available to farmers in 1900, only that they did not use them. Therefore a farmer in 1900 can be compared with a farmer in 1940 who has adopted a hybrid variety, and the latter need not be viewed as benefiting from some mysterious gift of time. Instead he adopted a hybrid variety and, as a result, improved his yields while his predecessor did not. From a technology adoption standpoint, the latter farmer is viewed as a more intensive manager of his crop (or simply as using more advanced technology). While the management or technological "covariates" are conceptualized separately from the weather for the purposes of this study, it is worth keeping in mind that farmers are not, in the long run, unresponsive to weather. The "dumb farmer" assumption implicit in some of the more basic production models does not hold—farmers respond to the weather and climate by altering their crop choice and growing methods (Ruttan 2002).

That the various technology and use variables are highly correlated is of some interest, both from a statistical and a conceptual point of view. The statistical problems of correlated independent variables are well-known, and whether those problems manifest themselves shall be considered when statistical models are applied. From a conceptual point of view, the adoption of various technologies can, as described above, be viewed as indicators of the level of management, and their collinearity only provides evidence that intensive managers with respect to, say, fertilizer also tend to intensively adopt other technologies as well.

This study considers four of the most important technology variables that affect corn yields: irrigation, hybrid varieties, commercial nitrogen use and adoption of genetically modified (GM) corn. Each of these variables are discussed in more detail in the sections that follow.

Nitrogen Use

State-level data on the nitrogen used on corn are available from the USDA, ARS (1957) for 1954, from Ibach, et al. (1964) for 1959 and from the USDA, ERS for the years spanning 1964-2010 (2011b).[64] The state data were fairly sparse, and a number of values needed to be estimated. First, missing application rates for states were estimated from the data of other states using the fourth power of the inverse of their distance from each other as the weighting. Remaining null data points were then filled in by assuming that the growth rate in fertilizer use for any state was constant between any two periods for which data were available. Reported data were used instead of estimated data whenever such data were available.

Whether these estimates are reasonable can be assessed by comparing the values that would have been estimated for a given year with reported values. The procedure seems to have performed well for most states, for example see the estimates for Minnesota provided in Figure 25, Panel a. Values represented by a cross are

---

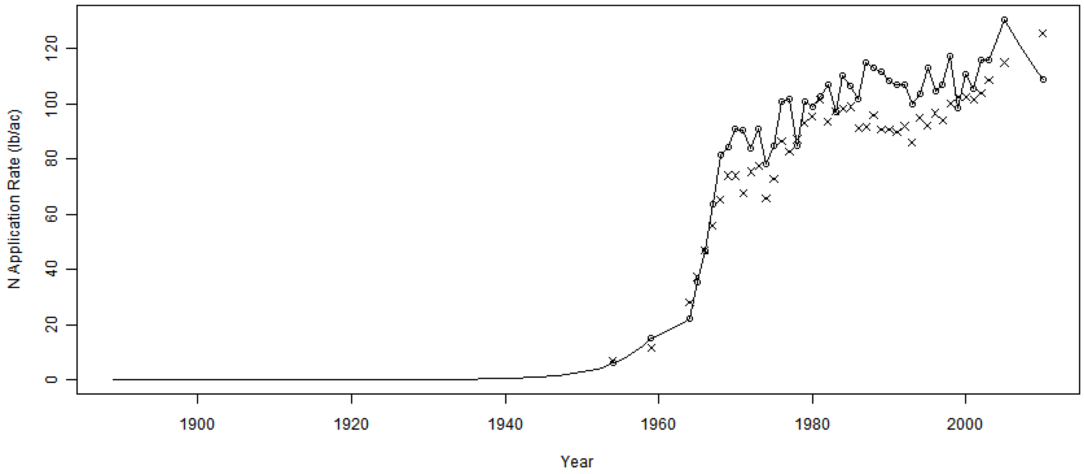
[64] It was confirmed with Mr. Robert Ebel of the USDA's ERS that there are no other known panels of fertilizer use on corn (personal communication, 16 June, 2011).

estimated values, while values represented by a circle are reported values; the line indicates the set of values used in the analysis. In Minnesota, as in most other states, the majority of the estimates that most deviate from their observed values occur in the later years when data were available. Estimates in some states, however, were not as convincing. For example, in the years after about 1970 the estimates for Nebraska (Panel b) would not have been very accurate had observed data not been available. However, for the earlier years, the Nebraska estimates seem reasonable, but slightly biased towards slower growth in commercial nitrogen use. Such figures for all states are provided in the Corn Atlas.

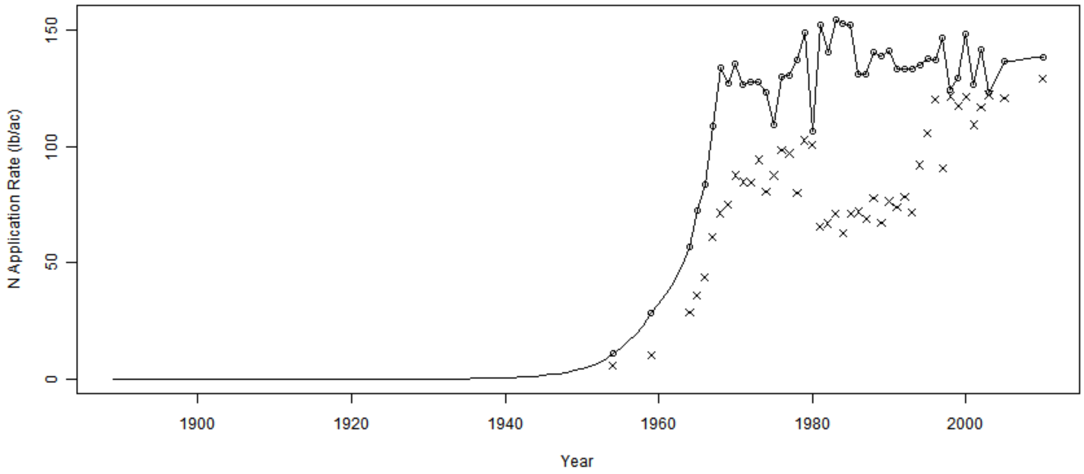
Various other methods were tried and rejected. The most promising rejected method required assuming that the amount or rate of nitrogen applied to corn in a state maintained either a constant or a time-trended ratio to the state's overall nitrogen fertilizer use. Unfortunately, no stable relationship could be found, even in specifications that accounted for the prices of corn and fertilizer. It was hoped that this procedure would work because county-level *total* fertilizer use is available for most of the census years. Had a relationship between overall nitrogen application and the rate of use on corn been found, it might have been possible to estimate the nitrogen used on corn for each county. While such estimates were derived, they were not convincing and it was feared that they would mislead readers about the actual spatial accuracy of this variable.

Figure 25: Estimated Nitrogen Application Rates on Corn, MN and NE

Panel a: Estimated Minnesota Nitrogen Application Rates



Panel b: Estimated Nebraska Nitrogen Application Rates



Source: Estimated based on data from USDA, ARS (1957), Ibach, et al. (1964) and USDA-ERS (2011b).

Notes: Points represented by a cross are estimated values, while values represented by a circle are reported values; the line indicates the set of values used in the analysis

Hybrid Adoption

State-level hybrid corn adoption data are available from the USDA's Agricultural Statistics (various years) for 1933 through 1960. Of the 1,344 state-year combinations, 218 observations were nulls, and the nulls were only found at the beginning of each state's series, when adoption of hybrids was almost nil. Almost all of the states had fully or nearly fully adopted hybrid varieties by 1960, although some states lagged significantly (for example, Arizona, with 42.5 percent adoption and Wyoming with 63.0 percent). These series were extended back to 1920 (when the first hybrid variety was commercially released) and forward to 2007 by assuming that the growth rates at the early end of each state's series extended backward to 1920, and that the growth rate around 1960 extended forward to 2007. Once a state achieved full adoption, it was assumed to stay at that level. While the manipulations to the source data were very minor, each state's estimated adoption curve is provided in the Corn Atlas for the inspection of interested readers.

Irrigation

County-level irrigated acreage of corn was provided in the various censuses starting in 1949, when only 0.6 percent of corn acres were irrigated. By 1959, about two percent of U.S. corn acres were irrigated, but with substantial variation among counties. At the high-end were the counties in the Pacific and Mountain regions, in which 91 percent and 63 percent of corn acres were irrigated, respectively. These regions irrigated the highest proportion of their corn acres in all years for which irrigation data were reported (followed by the Northern and Southern plains, which neither adopted irrigation as quickly nor as completely). Given the implied growth rates, it is assumed that no corn was irrigated prior to 1949, although it is recognized that the irrigation rate every year was actually slightly positive.

GM Corn

The data for adoption of genetically modified corn[65] are from the USDA-ERS from 2000 through 2010 (USDA-ERS 2011a). Proprietary data from Doane were used to calculate 1996-1999 adoption shares (see Pardey, Alston and Ruttan 2010). These data appear to cover nearly the entire history of GM corn adoption, and GM adoption is taken to be zero in all states before 1996.

Summary of Variables

A summary of each of the variables to be used in the subsequent analysis is presented in Table 2. The dataset includes observations for county-years in which a yield was reported, for a total of 60,079 observations. There are no null values for any variable in the remaining observations—a zero for any variable is assumed to be equal to or approximately equal to zero as discussed in the sub-sections above.

---

[65] Data are available separately for herbicide tolerant (HT), *Bacillus thuringiensis* (Bt) and "stacked gene" varieties. No distinction between the three is made in this assessment.

Table 2: Summary of Variables

					Summary of Non-Zero Entries						
Variable	Period	N	T	Zeros	Min	Q1	Q2	Q3	Max	Mean	Sd
Yield (bu/ac)	1889-1919	11,942	4	0	1.5	14.8	22.1	31.3	79.4	23.8	10.9
	1924-1944	14,676	5	0	0.7	13.0	20.4	30.5	76.1	22.5	12.1
	1949-1978	19,539	7	0	2.0	28.3	45.8	66.8	178.5	49.1	26.2
	1982-2007	13,922	6	0	2.0	76.8	98.6	122.2	254.4	100.2	33.8
	1889-2007	60,079	22	0	0.7	20.0	36.2	72.1	254.4	49.4	38.1
CSI (0-1)	1889-1919	11,942	4	83	0.00	0.10	0.42	0.70	1.00	0.42	0.32
	1924-1944	14,676	5	260	0.00	0.02	0.23	0.63	1.00	0.33	0.31
	1949-1978	19,539	7	189	0.00	0.15	0.42	0.66	1.00	0.42	0.30
	1982-2007	13,922	6	34	0.00	0.32	0.54	0.74	1.00	0.53	0.27
	1889-2007	60,079	22	566	0.00	0.13	0.42	0.68	1.00	0.43	0.31
Nitrogen Application (lb/ac)	1889-1919	11,942	4	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	1924-1944	14,676	5	0	0.0	0.1	0.2	0.8	5.7	0.7	0.9
	1949-1978	19,539	7	0	0.2	14.4	56.9	103.0	158.4	60.7	46.4
	1978-2007	13,922	6	0	47.6	125.9	136.8	146.1	163.4	132.5	20.5
	1889-2007	60,079	22	0	0.0	0.1	10.0	117.8	163.4	50.6	59.2
GMO Adoption (pct. of acres)	1889-1919	11,942	4	11,942							
	1924-1944	14,676	5	14,676							
	1949-1978	19,539	7	19,539							
	1982-2007	13,922	6	7,068	0.5	9.0	32.7	69.3	93.0	37.7	27.3
	1889-2007	60,079	22	53,225	0.5	9.0	32.7	69.3	93.0	37.7	27.3
Hybrid Adoption (pct. of acres)	1889-1919	11,942	4	11,942							
	1924-1944	14,676	5	0	0.1	0.1	0.3	4.0	99.8	10.7	23.8
	1949-1978	19,539	7	0	3.5	89.0	99.5	100.0	100.0	89.7	19.5
	1982-2007	13,922	6	0	100.0	100.0	100.0	100.0	100.0	100.0	0.0
	1889-2007	60,079	22	11,942	0.1	10.0	99.5	100.0	100.0	68.6	42.6
Irrigation (pct. of acres)	1889-1919	11,942	4	11,942							
	1924-1944	14,676	5	14,676							
	1949-1978	19,539	7	12,849	0.0	0.4	3.8	69.1	100.0	29.8	39.5
	1982-2007	13,922	6	6,952	0.0	1.8	8.9	59.2	100.0	30.1	36.8
	1889-2007	60,079	22	46,419	0.0	0.9	6.6	63.5	100.0	29.9	38.1

Source: The data sources of the variables are described in Sections 2.2.1, 4.1, 4.3.3 and A.2.

Notes: Summary statistics are shown for non-zero entries only. The minimum may appear to equal zero due to rounding. Nitrogen application never equals zero, although the application rate approaches zero according to the estimated adoption curve (see the section "Nitrogen Use" for more information).

4.3.4 Climate Variability and Corn Yield

To begin considering the effect of weather on yield, a regression of the following form was estimated:

$$Y_{it} = \beta_0 + \beta_{GI} CSI_{it} + \beta_{fert} FERT_{it} + \beta_{gmo} GMO_{it} + \beta_{hybrid} HYBRID_{it} + \beta_{irrig} IRRIG_{it}$$

where:

- Y_{it} is the corn yield in county *i*, year *t*, measured in bu/ac;
- CSI_{it} is the corn suitability index value;
- $FERT_{it}$ is the level of fertilizer use in lbs/ac;
- GMO_{it} is the percentage GMO adoption;
- $HYBRID_{it}$ is the percentage hybrid adoption; and,
- $IRRIG_{it}$ is the percentage of acres that were irrigated.

Each of the variables is summarized in Table 2 (above). All of the estimated parameters are positive and, based on their t-values, all of the parameters are highly significant ($p < 0.01$), as was the regression itself. Unfortunately, such a model is likely to have significantly spatially autocorrelated error terms, and the significance of the estimated parameters may be questionable. It is therefore prudent to consider whether the residuals of this regression are autocorrelated before continuing. Before spatial autocorrelation can be examined, it is first necessary to specify the nature of the spatial relationship as shall be done in the following section.

Creating a Spatial Weighting Matrix

One common way to characterize a spatial relationship among observations in a cross section is to specify a spatial weighting matrix. The procedure begins by constructing a contiguity matrix that defines the (assumed) ways in which observations are related across space. The procedure is familiar to econometricians as very similar methods are used to model correlation among observations in general, and especially among observations that are temporally autocorrelated.

For a single-year cross section of spatial units, a contiguity matrix (C^t) is usually an $N \times N$ matrix in which element C_{ij}^t equals one if the i^{th} region shares a border with the j^{th} region and zero if regions i and j are not connected. It is customary to set the diagonal elements equal to zero as well so that $C_{ii}^t = 0$. Then, a spatial weighting matrix, W^t, is created by row-normalizing C^t so that $\sum_{j=1}^{N} C_{ij}^t = 1 \;\; \forall i$. Any n-vector, x, can then be spatially lagged by calculating the dot-product $x^{-} = W^t \cdot x$. In this formulation, each element i of x^{-}, specifically x_i^{-}, is the simple average of the x values for connected regions, namely $x_i^{-} = C_i^t \cdot x / C_i^t 1_N$, where C_i^t is the row-vector representing the i^{th} row of C^t.

The spatial weights matrix is only slightly more complex for a balanced panel of data. Since the contiguity matrix is the same for every year, a global contiguity matrix for data ordered by region then by time can be created by the Kronecker product $C^t \otimes I_{N \cdot T}$, where N is the number of spatial units and T is the number of years in the sample. The resulting contiguity matrix will be a sparse block diagonal matrix with C^t repeated on the diagonal, namely:

$$C = C^t \otimes I_{NT} = \begin{pmatrix} C^t & 0_N & \cdots & 0_N \\ 0_N & C^t & 0_N & \vdots \\ \vdots & \cdots & \ddots & 0_N \\ 0_N & \cdots & 0_N & C^t \end{pmatrix}$$

Since C^t is repeated on the diagonal, row normalizing C will yield the spatial weighting matrix W (which, under these conditions, is equivalent to $W^t \otimes I_{NT}$). Notice that C is a purely spatial weights matrix with no spatiotemporal lag processes. Such processes could be modeled using the off-diagonal matrices in C, for example a single year temporal lag could be added by creating a tridiagonal matrix with C^t on the diagonal and N-dimensional identity matrices on the first off-diagonals (in which case $W \neq W^t \otimes I_{NT}$). More complex spatiotemporal autoregressive processes could be modeled similarly.

A few complexities arise when the panel is unbalanced. Most importantly, the contiguity matrix is not necessarily the same for all years since there may be no observation for one or more of a region's neighbors in a given year. One option is to simply delete the rows and columns of C for which there is no observation then form W by row normalizing C. Unfortunately, this caused many county-years to have no spatial lag, and many more to have a spatial lag determined by only one neighbor. To assure that each county-year had a spatial lag, the four nearest neighbors with an observation (i.e., "data neighbors") were found for each county-year based on the county centroids. This resulted in some counties with neighbors located relatively far away, so the contiguity matrix was not specified with binary weights. Instead, the inverse of the distance between the county centroids was used as the weight ($C_{ij}^t = (D_{ij})^{-1}$) for year-*t* data neighbors, where D_{ij} is the distance from the centroid of county *i* to that of county *j*. After forming the contiguity matrix for each year, the global contiguity matrix was formed as a block diagonal using the matrix direct sum:[66] $C = \bigoplus_{t=1}^{T} C^t$, where C has $\sum_{t=1}^{T} N_t$ rows and columns. The resulting contiguity matrix looks much like the one from the balanced-panel case, except that the off-diagonal zeros matrices are not necessarily square, but are of dimension $N_{i,t} \times N_{j,t+1}$.

Spatial Autocorrelation

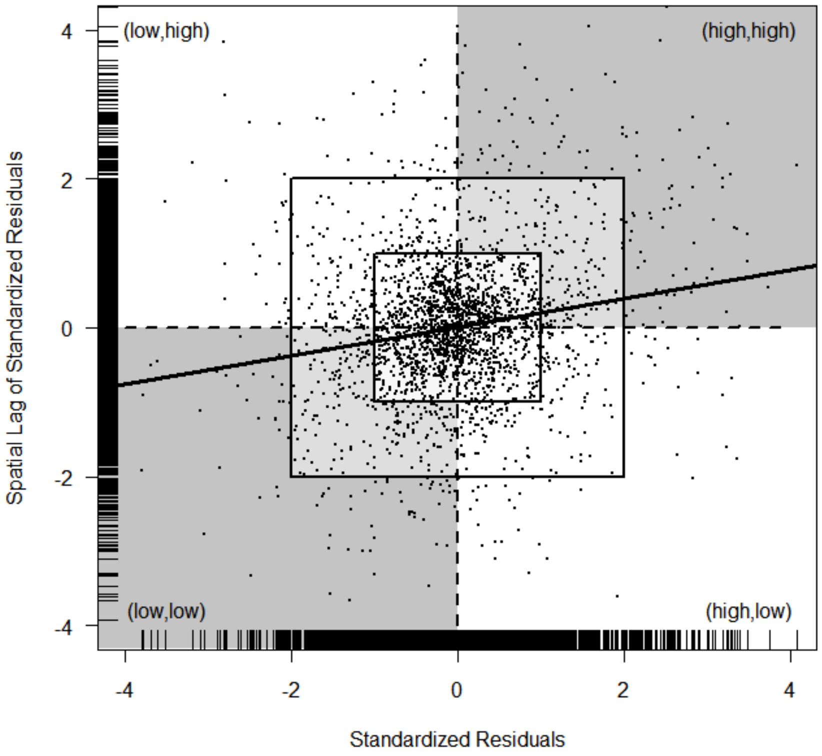
With the spatial weights matrix in-hand, it is now possible to examine whether the residuals of the regression are spatially correlated. A visual assessment can be made by plotting the standardized residuals against their spatial lag, as has been done in Figure 26. As the spatial weighting matrix has been constructed here, each point represents the error term of an observation on the horizontal axis, and the inverse distance weighted mean residual of that point's spatial neighbors on the vertical axis. Points in the first and third quadrants of the figure are either both above or below the average, and represent positive spatial autocorrelation. As the current dataset has over 60,000 points, only 5,000 randomly selected points are plotted in order to avoid visual clutter. A linear regression of the spatially lagged residuals on the residuals (e.g., $W\varepsilon = \beta_0 + \beta_1\varepsilon$) was then fitted (using all of the observations), and the result is represented by the solid line. The positive slope of the line indicates that there is positive spatial autocorrelation in the data.[67]

---

[66] The matrix direct sum forms a sparse block diagonal matrix from a list of input matrices. For example, $A \oplus B = \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}$, where A and B are arbitrary square matrices and 0s represent appropriately sized matrices of zeros. This should not be confused with the Kronecker sum which uses the same notation.

[67] The general form of this figure was suggested by Ward and Gleditsch (2008).

Figure 26: Moran Scatterplot of Residuals



Source: The general form of this figure was suggested by Ward and Gleditsch (2008).[68] The values were calculated using the data described in Table 2.

Notes: This figure summarizes the standardized residuals from the regression specified on p. 116. To avoid over-plotting, only 5,000 randomly selected residuals are shown (of the over 60,000 available data points). The line represents a simple regression (through the origin) of the spatially lagged residual on the residual. The "rug" on each axis represents the point density in that dimension. The slope of the line was 0.178 (standard error = 0.004).

There are a number of more formal tests for spatial autocorrelation. The most common of these is the (global) Moran's (1950) I statistic:

$$I = \frac{n\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(y_i - \bar{y})(y_j - \bar{y})}{S_0 \sum_{i=1}^{n}(y_i - \bar{y})^2}, \text{ where } S_0 = \sum_{i}^{n}\sum_{j}^{n} W_{ij} .$$

---

[68] Ward and Gleditsch (2008) also provided code to produce their version of the figure. That code was updated and modified to produce this figure.

The statistic is asymptotically normal with an expected value of -1/(N-1) (Ward and Gleditsch 2008). Using this formula and the spatial weighting matrix defined above generates a Moran's I that is significantly different from its expected value ($p < 0.01$). A similar statistic, Geary's C is specified as:

$$C = \frac{n-1}{2S_0} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} W_{ij}(x_i - x_j)^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2}.$$

The expected value of a random pattern is one, and the statistic ranges from zero to two (Getis 2010). The errors from the model above yield $C = 0.899$ ($p < 0.01$).

Impacts of Weather on Corn Yield

There are numerous other methods for assessing spatial autocorrelation. In the present case, however, it is clear that the errors of the current model are autocorrelated. Among econometricians there are two basic model forms that are used to account for spatial autocorrelation, one with a spatially lagged dependent variable (a so-called SAR model) and another called the spatial error model (SEM).

The spatially autoregressive (SAR) model takes a form similar to that used to account for first order temporal autoregressive processes:

$$y = \rho Wy + X\beta + \epsilon$$

where $\epsilon \sim N(0, \sigma^2 I)$. The expression Wy creates a spatial lag of the dependent variable, and therefore ρ indicates the level of spatial dependence in the dependent variable, given $X\beta$. In this model, $\hat{\beta}$ does not estimate the marginal effect of X on y; some manipulation reveals:

$$E(y) = (I - \rho W)^{-1} X\beta$$

Therefore, for $\rho \neq 0$ a change in *X* for any observation changes the marginal effect of β for all other observations that are spatial neighbors of *any* order to that observation.[69] While the SAR model could be estimated using OLS, the inherent feedback mechanism causes the resulting parameter estimates to be biased (see LeSage and Pace 2010).

To justify a SAR model, one would need to propose a mechanism through which the dependent variable at one location might *directly* affect the dependent variable of its neighbors. In the case of crop yields, it seems implausible that higher yields at one location (county) would cause higher yields in other locations (counties). Nevertheless, there does appear to be autocorrelation in the error that is not sufficiently accounted for by the independent variables. It is therefore reasonable to assert that there is some (unspecified) spatial process that affects yields and is not accounted for in the OLS model. Some potential spatially correlated variables that are not sufficiently accounted for in the independent variables include:

- pest and disease impacts
- soil types (the soil moisture model implicitly assumes homogenous soils)
- daily weather patterns (e.g., periods of high or low rainfall or temperature that do not reveal themselves in the monthly data)

---

[69] See Ward and Gleditsch (2008) and and LeSage and Pace (2010) for concise and accessible descriptions of parameter interpretation in SAR models.

- farming practice (other than those included, and in part reflecting spatial differences in human capital and management skills[70]) and extension service recommended practices
- other weather variables, such as humidity
- water pricing and pumping costs (e.g., well depth), and
- especially in early years, use of farm machinery and implements.

All of these (and many other) variables could generate error autocorrelation, but would not be properly modeled in the lag of the dependent variable.

The spatial errors model (SEM) allows for separation of the error into a spatially autoregressive portion and a non-spatial portion and is thus a better fit for the problem being addressed in this thesis. The usual specification is:

$$y = X\beta + \lambda W\xi + \epsilon,$$

where $\epsilon \sim N(0, \sigma^2 I)$. Here, ξ represents the spatial component of the error term and ϵ represents the non-spatial component. The log-likelihood function for the SEM is fairly simple, but certain empirical difficulties arise in maximizing it. Various algorithms are available to reduce the computational requirements. This study uses a method proposed by Smirnov and Anselin (2009) as implemented in Roger Bivand's R package spdep using code provided by Prof. Anselin.

Again, all of the estimated parameters are positive and highly significant. The estimated value for λ is significant, indicating that there was indeed spatial autocorrelation in the errors of the OLS model (Table 3). Taking account of this autocorrelation caused the estimated parameter for nitrogen to decrease by about a

---

[70] Alston et al. (2010, Chapter 3) report large variations across U.S. states in the composition (age, education, hired vs. family status) of agricultural labor.

quarter, and that for hybrid adoption to increase by over 75 percent. The parameter on the climate suitability indicator changed very little. All estimated parameters remained statistically significant.

Table 3: OLS and SEM Results

Variable	OLS			SEM		
	$\hat{\beta}$	$SE(\hat{\beta})$	t Value	$\hat{\beta}$	$SE(\hat{\beta})$	z Value
Intercept	10.840	0.158	68.710	10.412	0.219	47.649
CSI	28.420	0.270	105.450	26.951	0.267	101.070
NITRO	0.319	0.002	133.000	0.241	0.003	76.894
GMO	0.512	0.006	86.680	0.537	0.009	58.218
HYBRID	0.116	0.003	40.540	0.206	0.004	54.318
IRRIG	0.270	0.004	70.310	0.269	0.004	73.061
$\hat{\lambda}$				0.407	0.008	49.677
df		60,073			60,073	
R^2		0.728			0.739	
Log likelihood		-264,882			-263,261	

Notes: The R^2 for the SEM is the Nagelkerke R^2. To maintain scaling of the results, CSI is a decimal percent, with a potential range of zero to one.

The results imply that the suitability indicator has some ability to account for weather. Namely, when all of the management cum technology variables are equal to zero, as would be the case in the earlier years, yields would range from about 10 to 40 bushels per acre. In the period before 1910, the mean area-weighted CSI value was 0.579, implying that the mean yield would be about 26 bushels per acre; the actual area weighted mean yield of that period was about 27.6 bushels per acre. Since all of the technology variables have positive estimated coefficients, as technology increased over time, the relative importance of weather decreased.

However, it is likely that the weather interacts with some of the technology variables. For example, the impact of hybrids, nitrogen and GMOs would be expected to have a positive interaction with the weather,[71] while irrigation would have less impact under years with good weather. To assess this possibility, the model was re-estimated with some interaction terms, and the results are presented in Table 4. The OLS results are presented in the left half of the table and are included simply for comparison. The results of the SEM are presented in the right columns of the table.

There is a version of the Breusch-Pagan test for heteroskedasticity that is modified to account for λ in spatial error models (Anselin 1988, pp. 121-123). Testing the SEM version of the model for heteroskedasticity yields a statistic that, at any reasonable level of certainty, rejects the null hypothesis. Thus, it is not clear that the standard errors estimated for the SEM are correct. Fortunately, robust standard errors can be estimated in a way that is similar to that used in OLS. Namely, the residuals ($\hat{\epsilon}$) from a regression through the origin of Wy on $\hat{\lambda}WX$ are used to form White's *hc0*.[72] The resulting robust standard errors and corrected z-scores are reported in Table 4.

---

[71] The reader is reminded that the non-weather covariates included in this model are intended to measure the intensity of management or the state of technology. Insofar as technologically more advanced managers are more likely to adopt and use these technologies, the yield effect of the technology covariates is expected to be positive. From a bio-economic viewpoint, hybrid adoption, irrigation and nitrogen application all increase the cost of production; thus, few assumptions are required to conclude that the technologies would not be adopted if they did not either increase yields or decrease risk. The effects of genetically modified corn are less obvious. Herbicide tolerant (HT) corn has little effect on yield, but reduces overall production costs, while Bt corn reduces both yield losses due to insects and insecticide-related costs (see Qaim 2009 and the references therein for a summary). Bt and stacked gene varieties are therefore expected to have a yield-enhancing effect, while HT varieties are expected to increase yield indirectly.

[72] $hc0 = \left((\hat{\lambda}WX)'(\hat{\lambda}WX)\right)^{-1} \hat{\lambda}WX \; diag(\varepsilon)\hat{\lambda}WX \left((\hat{\lambda}WX)'(\hat{\lambda}WX)\right)^{-1}$. This method was suggested by Roger Bivand.

Table 4: Regression Results with Interactions

Variable	OLS			SEM		
	$\hat{\beta}$	$SE(\hat{\beta})$	t Value	$\hat{\beta}$	$SE(\hat{\beta})$	z Value
Intercept	15.87	0.18	87.3	14.97	0.11	130.8
CSI	14.73	0.38	39.1	14.50	0.20	72.1
NITRO	0.25	0.00	55.8	0.19	0.01	35.2
GMO	0.47	0.01	40.6	0.47	0.02	20.6
HYBRID	0.03	0.01	6.0	0.11	0.00	27.7
IRRIG	0.55	0.01	78.2	0.56	0.01	49.0
CSI*NITRO	0.13	0.01	15.7	0.10	0.01	11.2
CSI*GMO	0.15	0.02	6.8	0.19	0.04	5.1
GIS*HYBRID	0.24	0.01	25.6	0.23	0.01	31.5
GSI*IRRIG	-0.54	0.01	-47.5	-0.54	0.02	-29.6
$\hat{\lambda}$				0.39	0.01	47.2
R^2		0.73			0.76	
Log likelihood		-262,374			-261,262	

Notes: The R^2 for the SEM is the Nagelkerke R^2, and the standard errors are adjusted for heteroskedasticity. To maintain readability of the results, CSI is a decimal percent, with a potential range of zero to one (without this correction, the parameters on the interacted terms become very small).

The model gives a pseudo-R^2 of about 0.76 and all of the coefficients are highly significant. The estimated spatial autocorrelation correction ($\hat{\lambda}$) is of about the same magnitude and significance as it is in the model without interaction terms. In this model, the direct effect of the suitability indicator (CSI) is reduced, so that an increase in the indicator of 0.1 points will, on average, be expected to increase yields by 1.45 bushels per acre.[73] Of particular note is that full adoption of irrigation is

---

[73] It is worth keeping in mind that the corn suitability index is unitless. This is not practically different from other variables, such as rainfall or temperature, insofar as these exhibit no parametrically linear relationship to yield. The advantage of the CSI is that, unlike temperature or rainfall, the indicator increases as the climate becomes more hospitable for corn over its entire range of values. Thus, the CSI has a natural interpretation in a regression analysis, while temperature and rainfall do not.

estimated to increase the yield of a county by about 55 bushels per acre, but that under perfect weather (CSI=1), this effect is almost entirely cancelled out. This result is surprisingly accurate, considering that a CSI equal to one can only be achieved if the moisture index during each phenological stage were also equal to one. It also implies that nitrogen use, GMO adoption and hybrid adoption provide sufficient proxies for management without consideration of irrigated acres.

The regression specification does not include a time trend: production responses are not modeled as benefiting from an unspecified set of technologies that are correlated with the year. This approach has a number of benefits. First, insofar as adoption profiles differ across space, a time trend will be more or less accurate depending on where production takes place. By directly considering the spatial and temporal profile of technology adoption, the differing effect of time across space need not be considered (thus avoiding the complexities of deriving and interpreting an unbalanced spatial panel model). Second, events such as the 1973-1974 oil embargo (see Section 2.1) are not likely to be accurately modeled by a time trend, but do affect the use of technology, underscoring the importance of considering the variables directly. Finally, it will be possible to directly compare typical management regimes without specific regard to time: consistent with production theory, it only matters which technologies were applied, not when the crop was produced.

To explore this result further, consider three management regimes, representing the average technology adoption and use of 1899-1909, 1959-1969 and 1997-2007. Under the 1899-1909 scenario, the marginal effect of climate suitability is 14.5 bushels per acre, or 54 percent of average yield. As technology adoption and use increases, the relative marginal effect of climate decreases, such that in the 1997-2007 scenario, the marginal effect of climate is 39 percent of average yield (53.7 bushels per acre).

Consider further three weather scenarios: a poor weather year (CSI = 0.1), a moderate weather year (CSI = 0.5) and a good weather year (CSI = 1.0). The expected yield under each technology scenario is presented in the left portion of Table 5. Based on these results, a good weather year under the 1899-1909 scenario increases yield by about 80 percent over the expected yield under poor weather; good weather increases the yield in intensively managed areas (typified by the 1997-2007 scenario) by about 54 percent. The expected yields under the poor, moderate and good weather scenarios roughly correspond to the first, second and third quartiles of the yield distributions of the respective periods (right part of Table 5), indicating that the expected yields of each period are reasonable, even though a time trend was not used to derive the values.

Table 5: Expected Yield and Yield Quartiles

Expected Yield				Yield Quartiles			
	Management Scenario				Period		
Weather	1899-1909	1959-1969	1997-2007	Quartile	1899-1909	1959-1969	1997-2007
		bu/ac				*bu/ac*	
Poor	16.4	31.8	83.9	First	14.9	23.9	81.6
Moderate	22.2	46.1	104.2	Second	22.1	38.2	105.3
Good	29.5	64.0	129.5	Third	31.1	54.7	130.9

Notes: Poor, Moderate and Good weather scenarios were calculated using a CSI of 0.1, 0.5 and 1.0, respectively. The technology adoption scenarios represent the area-weighted average technology adoption and use for the indicated periods.

4.4 *Conclusion*

The discussion and analyses presented in Chapter 2 showed that the geographic footprint of U.S. corn production has changed over time, and methods developed at the end of that chapter were used to show the nation's output of corn has been positively affected by these movements. This chapter continued the analysis by investigating the ways in which location might affect yields, and hence output. A review of the literature revealed that applying statistical models to raw monthly weather variables was unlikely to give convincing or readily interpretable results, namely because models should ideally be informed by the biology of the crop. Crop responses to weather are dynamic within seasons—it matters when weather events occur during the lifecycle of the crop, and crop responses are often cumulative.

Inspired by ecological modeling techniques, an intra-seasonally dynamic model of corn yield was created. In keeping with the traditions of the economics literature, the model provides a single indicator of the suitability of the weather for a given location-year—the "corn suitability index." While similar in name, this index has little else in common with the weather indexes of Stallings and his successors. For example, the index is defined based on the biology of the crop, and is not as likely to capture spurious correlations among variables. Thus, whether the index is useful can be tested, namely the index is expected to have a positive estimated coefficient when regressed on yield, and it should show reasonable interactions with technology variables.

Whether the index gives reasonable results was tested in a regression model. As anticipated, the estimated sign on the index was positive, and as technology adoption and use increases, the relative effect of weather decreases. Good weather almost completely nullifies any benefit of irrigation, implying that the moisture component of the index likely does a good job of representing the impact of rainfall.

At the same time, the results imply that adoption and use of other technologies allows farmers to take more advantage of good weather, and to partially ameliorate the effects of poor weather. It appears then that the index might offer a reasonable way for researchers to account for weather in productivity and other models. Nevertheless, the index and the models used to assess it do have certain limitations; these will be discussed in Section 5.2.

Chapter 5 Conclusion

This thesis reported on an investigation into the relationship between the footprint of corn production and corn productivity. A number of different methods and procedures were used in the investigation, each attacking the problem from a different angle. This conclusion will bring these lines of inquiry together in the next sub-section. In the final subsection of this chapter, a program of work for extending and improving the methods used in this study will be presented.

5.1 *General Comments*

The results reported in Chapter 2 leave little doubt that the footprint of U.S. corn production is dynamic, and that the changes in where corn is produced have had important consequences for the nation's food, feed and biofuels production. The economic index approach developed in the second chapter gives direct measures of the impact of yield, area and the spatial allocation of production on the nation's corn output. Very few assumptions are required to interpret the results, and the assumptions required are at least as reasonable and practical as almost any in economics. The indexing results are about as transparent as any indicator could be. This attribute is important if the results are to inform policy decisions since those charged with making the decisions must understand and believe the evidence they are presented.

While Chapter 2 investigated *how much* location might have affected the nation's output of corn, Chapters 3 and 4 continued the investigation by considering *how* location influences production. Chapter 3 reviewed past work and Chapter 4 reported on new work conducted as part of this study. An initial assessment showed that the monthly temperatures under which corn is grown have decreased because corn production has, on average, moved into cooler climates. While the present goal was not to assess climate change, the results have important implications for those who endeavor to estimate the impacts of global warming. The results clearly show that if credible predictions of the impacts of climate change on the world's food production are to be achieved, researchers will need to think about agriculture as a spatially dynamic process. Insofar as climate researchers fail to account for farmers' ability to adjust to changing conditions, they will overestimate the impacts of climate change.

The problem of assessing climate change is compounded because there is no scientific consensus on the impacts of weather on the yield of corn or other crops, especially at the broad spatial scales that are relevant for thinking about consequences of climate for global food supplies. The literature can be used to support almost any desired conclusion, for example, one could cite credible studies to support the notion that climate change will decrease land values and will at the same time increase yields, or that an increase in July temperatures will both increase and decrease the production of corn. The statistical approach to determining the relationship between yield and weather variables has been plagued by such problems for more than a century. Parameter estimates from such models are extremely sensitive to location, and it is difficult to interpret what they mean or how far one can extrapolate from the result. Worse, however, is that absent any information on the biology of the plant, one cannot assess whether the estimated parameters are reasonable. It is notable that almost none of the regression-based studies reviewed made any *a priori* predictions on the signs of the estimated parameters, and tended to interpret the results only in terms of their statistical fit.

Data availability limits what can be done to correct this. Certainly, there are insufficient data on varieties, technology and weather to attack the problem with a comprehensive field-level crop model. However, that does not mean that biology must be ignored. The crop model created in this study introduces some agronomic realities into the estimation, yielding a spatially (and temporally) sensitive indicator of suitability that is meaningful. The effect of a given change in temperature or rainfall is based on the biology of the crop, not any peculiarities of the area under assessment. Thus, the model is transparent and, while one might disagree with the way it is specified, it is at least open to criticism (and thus improvement) on biological grounds. Using this model, it was shown that the climate in which U.S. corn is grown is more suitable now than it would be had the footprint of production not changed.

This study presented new ways of conceptualizing production and productivity that are consistent with how economists think about production—namely, the inputs to the production process are important, the year in which production takes place and the location of production are not. For example, studies that rely on deviations from temperature or rainfall trends (either explicitly, or implicitly by using these variables in regressions) give primary importance to *where* production takes place—so much so that it is difficult to understand how the results of such analyses apply to, for example, an entire nation. That the footprint of production is in constant flux means that such procedures are likely inappropriate when used to assess changes over time. But, weather is important to agricultural production and it must be taken into account if the productivity effects of technology adoption are to be assessed, particularly in the context of climate change. As shown, there are important interactions between weather and technology adoption.

Modeling weather as an input to production rather than as a nuisance that must be corrected for has many conceptual and practical advantages for economists. For example, regression coefficients on temperature, rainfall and their deviations have little meaning, especially when applied over space and time, while the corn suitability index has a ready interpretation for all locations and time periods, especially when its interaction with technology adoption is taken into account. These methods offer new ways to consider how technological interventions might, for example, be used to remediate the effects of climate change. Further, the CSI is agnostic about space and time—only the weather that occurs in a given crop year is important; the long-run "average" weather at a location is not. If indeed long-run average weather (i.e., climate) is changing, and the "dumb farmer" scenario does not hold, then economic analysis will be greatly enhanced by direct consideration of weather in ways that are not dependent on the ever-changing climate and geography of production.

5.2 *A Program of Future Work*

There are a number of possible ways in which the results presented here could be extended and improved. Broadly, these relate to improving the ability of the model to account for yield differences and applying the results in other lines of inquiry. Inclusion of more information on the likely varieties used at each place and time could be helpful, especially in predicting the phenological responses of the plant. Further, some methodology should be devised to disaggregate the state-level input data to the county level, or to secure more county-level data directly. Using data from the state-level in a county-level assessment both decreases the accuracy of the model and induces spatial autocorrelation. In addition to developing more spatially disaggregated data on the technology adoption and use variables, the regression results reported above imply that there are significant interactions between the weather and technology. Rather than viewing technology independently from the weather, some accuracy may be gained by embedding these interactions directly in the crop suitability index. For example, it is unlikely that fertilizer application rates interact linearly with the crop suitability index, but rather that the variable has a non-linear interaction with moisture.

While not much can be done to improve the historical weather data, the crop model used here would likely benefit from some accounting of differences in soils. Since soil types are more-or-less constant across time, current data could be used to estimate soil moisture holding capacity and percolation rates. This was attempted, but the author was not sufficiently confident in his understanding of hydrology to use the results in place of the proven CLIMEX (Sutherst and Maywald 1985) and CGIAR-CSI (Trabucco and Zomer 2010) approaches, which do not consider the soil type. The effort to estimate county-level soil moisture capacity is documented in Section A.5.

A second promising line of inquiry relates to climate change. This thesis investigated how much changes in the footprint of production have changed output, and how those movements have affected yields. But, the question of what drives that movement is left unanswered. To assess the impacts of a projected change in climate, it is important to know what farmers might do in response to the change—applying the expected climate to the existing area clearly has its limitations. This study reveals that the landscape of agriculture can change quite quickly and quite markedly (for example, consider the substantial change in the location of corn production between 2002 and 2007), but it is not yet possible to predict such spatial reallocation with any accuracy.

To develop his weather indexes, Stallings (1960) assumed that certain indirect effects of weather (such as insect damage) were positively correlated with the weather, and therefore that his indexes included certain indirect effects of weather on yield. By comparison, the corn suitability index developed here is intended to represent only the direct effect of weather. Damage due to disease, for example, is a function of spore availability (say in the case of southern maize rust), and the climatic factors that foster disease development are likely to be different from those that tend to increase yield. Similarly, the climate suitability of plant pests (e.g., corn stem borer) might not always be coincident with the climate attributes that optimize plant growth. Nevertheless, any correlation between the suitability indicator and yield may be either muted or increased by pest and disease effects, depending on how the developmental responses of these organisms are similar to or different from that of the host plant.

The HarvestChoice project has recently started a program of work to characterize the climatic responses of a number of important pests and diseases using methods similar to those used to develop the corn suitability indicator used in this study (see Beddow et al. 2010a). While the HarvestChoice estimates are intended to provide indicators of the potential for a given pest or disease to thrive at a given location on average, the models developed by the project should allow for establishment of reasonable upper bounds on the yield effects of important pests and diseases. As a result of these and similar efforts, it will soon be possible to be more precise about some of the potential *indirect* effects of weather in addition to the direct weather effects assessed in this study.

The regression model used to assess the weather index was purposely specified in the most parsimonious way possible. Many of the previous weather-yield regression models suffered from the appearance of data-dredging, and such appearances were intentionally avoided in this assessment. That the estimated coefficient on the suitability index had the appropriate sign and was statistically significant in the first incarnation of the regression model also required that little be done to further develop the model. Nevertheless, the results do present some difficulties. For example, the estimated parameters imply that adoption of technology alone is sufficient to give relatively high yields. But, even in the later years when the crop was produced using relatively high technology, yields were sometimes much lower than would be implied by the results (see Table 2).

Extremely poor weather can certainly cause low yields, or even a yield of zero if the crop is abandoned. To get a handle on this effect, more data on abandonment are needed. The yields used in this study were calculated based on *harvested* acres since data on the number of acres planted were not consistently available. The yield of a county's harvested acres represents the upper bound of the county's actual yield, and yields calculated in this way will be especially biased whenever the weather is extremely poor (i.e., when many acres are simply abandoned or failed to produce a measurable grain harvest). Thus, the coefficients on the suitability indicator and the interactions with the indicator can only be interpreted conditional on weather that is good enough to produce a crop that is worth harvesting. Future work aimed at establishing reasonable county-level estimates of abandonment would allow for a more direct interpretation of the meaning of the suitability index.

Bibliography

Allen, R.G., L.S. Pereira, D. Raes and M. Smith. 1998. *Crop Evapotranspiration: Guidelines for Computing Crop Water Requirements*. FAO Irrigation and Drainage Paper 56. Rome: Food and Agriculture Organization of the United Nations.

Alston, J.M., M.A. Anderson, J.S. James and P.G. Pardey. 2010a. *Persistence Pays: U.S. Agricultural Productivity Growth and the Benefits from Public R&D Spending*. New York, NY: Springer.

Alston, J.M., B.A. Babcock and P.G. Pardey. 2010b. "Summary, Synthesis, and Conclusion." Chapter in J.M. Alston, B.A. Babcock and P.G. Pardey, eds., *Shifting Patterns of Agricultural Production and Productivity Worldwide*. Ames, IA: Center for Agricultural and Rural Development.

Alston, J.M., J.M. Beddow and P.G. Pardey. 2010. "Global Patterns of Crop Yields and Other Partial Productivity Measures and Prices." In J.M. Alston, B.A. Babcock and P.G. Pardey, eds., *Shifting Patterns of Agricultural Production and Productivity Worldwide*. Ames, IA: Center for Agricultural and Rural Development.

Anderson, J.A. and P.B. Hazell. 1984. *Variability in Grain Yields*. Baltimore: Johns Hopkins University Press.

Anselin, L. 1988. *Spatial Econometrics Methods and Models*. Boston: Kluwer.

Badu-Apraku, B., R.B. Hunter and M. Tollenaar. 1983. "Effect of Temperature During Grain Filling on Whole Plant and Grain Yield in Maize (*Zea mays* L.)." *Canadian Journal of Plant Science* 63:357-363.

Bai, J. and P. Perron. 1998. "Estimating and Testing Linear Models with Multiple Structural Changes." *Econometrica* 66(1):47-78.

Baskerville, G.L. and P. Emin. 1969. "Rapid Estimation of Heat Accumulation from Maximum and Minimum Temperatures." *Ecology* 50(3):514-517.

Beauchamp, K.H. 1987. "A History of Drainage and Drainage Methods." Chapter in G.A. Pavelis, ed., *Farm Drainage in the United States. History, Status, and Prospects."* Miscellaneous Publication No. 1455. Washington, DC: United States Department of Agriculture, Economic Research Service.

Beddow, J.M., D. Kriticos, P.G. Pardey and R.W. Sutherst. 2010a. "Potential Global Crop Pest Distributions Using CLIMEX: HarvestChoice Applications." Working Paper. Saint Paul and Washington, DC: HarvestChoice.

Beddow, J.M., P.G. Pardey, J. Koo and S. Wood. 2010b. "The Changing Landscape of Global Agriculture." Chapter in J.M. Alston, B.A. Babcock and P.G. Pardey, eds., *Shifting Patterns of Agricultural Production and Productivity Worldwide*. Ames, IA: Center for Agricultural and Rural Development.

Boskin, M.J., E.R. Dulberger, R.J. Gordon, Z. Griliches and D. Jorgenson. 1998. "Consumer Prices, the Consumer Price Index, and the Cost of Living." *Journal of Economic Perspectives* 12(1): 3-26.

Boyne, D.H. 1962. "Changes in the Real Wealth Position of Owners Agricultural Assets, 1940-1960." Ph.D. Dissertation, University of Chicago, Chicago.

Bray, J.O. and P. Watkins. 1964. "Technical Change in Corn Production in the United States, 1870-1960." *Journal of Farm Economics* 46(4):751-765.

Brown, L.R. 1967. "The World Outlook for Conventional Agriculture." *Science* 158:604-611.

Calderini, D.F. and G.A. Slafer. 1998. “Changes in Yield and Yield Stability in Wheat During the 20th Century.” *Field Crops Research* 57:335-347.

Carter, C.A. and B. Zhang. 1998. "The Weather Factor and Variability in China's Grain Supply." *Journal of Comparative Economics* 26(3):529-543.

Chan-Kang, C., P.G. Pardey, S. Wood, J. Roseboom and M. Cremers. 1999. "Reassessing Productivity Growth in African Agriculture." Paper presented at the American Agricultural Economics Association annual meetings, Nashville.

Clawson, M. 1959. "Reflections on Gross Farm Output." *Journal of Farm Economics* 41(2):247-256.

Cochrane, W.W. 1953. "Professor Schultz Discovers the Weather." *Journal of Farm Economics* 35(2):280-283.

Cromarty, W.A. 1957. "Economic Structure in American Agriculture." Ph.D. Dissertation, Michigan State University, Lansing.

Cromarty, W.A. 1959. "An Econometric Model for United States Agriculture." *Journal of the American Statistical Association* 54(287):556-574.

Darwin, R. 1999. "The Impact of Global Warming on Agriculture: A Ricardian Analysis: Comment." *American Economic Review* 89(4):1049-1052.

Davis, F.E. and G.D. Harrell. 1942. *Relation of Weather and Its Distribution to Corn Yields*. Technical Bulletin No. 806. Washington, DC: United States Department of Agriculture.

Demir, N. and S.F. Mahmud. 2002. "Agro-Climatic Conditions and Regional Technical Inefficiencies in Agriculture." *Canadian Journal of Agricultural Economics* 50(3):269-280.

Denmead, O.T. and R.H. Shaw. 1960. "The Effects of Soil Moisture Stress at Different Stages of Growth on the Development and Yield of Corn." *Agronomy Journal* 52(5):272-274.

Deschênes, Olivier and Michael Greenstone. 2007. "The Economic Impacts of Climate Change: Evidence from Agricultural Output and Random Fluctuations in Weather." *American Economic Review* 97(1):354-385.

Diamond, J. 1999. *Guns, Germs and Steel: The Fates of Human Societies*. New York: W.W. Norton & Company.

Doebley, J.F. 2001. "George Beadle's Other Hypothesis: One-Gene, One-Trait." *Genetics* 158: 487-493.

Doll, J.P. 1967a. "An Analytical Technique for Estimating Weather Indexes from Meteorological Measurements." *Journal of Farm Economics* 49(1):79-88.

Doll, J.P. 1967b. "Perspective in Estimating the Effects of Weather: Reply." *Journal of Farm Economics* 49(4):938-940.

Douglas, W. 1760. *A Summary, Historical and Political, of the First Planting, Progressive Improvements, and Present State of the British Settlements in North America*. London: for R. and J. Dodsley.

Eichers, T., P.A. Andrilenas and T.W. Anderson. 1978. *Farmers' Use of Pesticides in 1976*. Agricultural Economics Report No. 418. Washington, DC: United States Department of Agriculture, Economics, Statistics, and Cooperatives Service.

Evans, L.T. 1993. *Crop Evolution, Adaptation and Yield*. Cambridge: Cambridge University Press.

FAO (Food and Agriculture Organization of the United Nations). 2011. ECOCROP Database. Rome: Food and Agriculture Organization. Obtained October 2011 from ecocrop.fao.org.

Fausey, N.R., E.J. Doering and M.L. Palmer. 1987. "Purpose and Benefits of Drainage." Chapter in G.A. Pavelis, ed., *Farm Drainage in the United States. History, Status, and Prospects."* Miscellaneous Publication No. 1455. Washington, DC: United States Department of Agriculture, Economic Research Service.

Fisher, I. 1921. "The Best Form of Index Number." *Quarterly Publications of the American Statistical Association* 17(133):533-537.

Fitzpatrick, E.A. and H.A. Nix. 1969. "A Model for Simulating Soil Water Regime in Alternating Fallow-Crop Systems." *Agricultural Meteorology* 6(5):303-319.

Fussel, B. 1992. *The Story of Corn*. Albuquerque: University of New Mexico Press.

Garcia, P., S.E. Offutt, M. Pinar and S.A. Changnon. 1987. "Corn Yield Behavior: Effects of Technological Advance and Weather-Conditions." *Journal of Climate and Applied Meteorology* 26(9):1092-1102.

Gehlke, C.E. and K. Biehl. 1934. "Certain Effects of Grouping Upon the Size of the Correlation Coefficient in Census Tract Material." *Journal of the American Statistical Association* 29(185):169-170.

Getis, A. 2010. "Spatial Autocorrelation." Chapter in A. Getis and M.M. Fischer, eds., *Handbook of Applied Spatial Analysis.* New York: Springer-Verlag.

Griliches, Z. 1957. "Specification Bias in Estimates of Production Functions." *Journal of Farm Economics* 39(1):8-20.

Griliches, Z. 1960. "Estimates of the Aggregate US Farm Supply Function." *Journal of Farm Economics* 42(2):282-293.

Gu, L. 2003. "Comment on Climate and Management Contributions to Recent Trends in U.S. Agricultural Yields." *Science* 300:1505b.

Hallam, D. 1990. "Agricultural Research Expenditures and Agricultural Productivity Change." *Journal of Agricultural Economics* 41(3):434-439.

Hamon, W.R. 1961. "Estimating Potential Evapotranspiration." *Proceedings of the American Society of Civil Engineers* 87:107-120.

Hansen, Z.K. and G.D. Libecap. 2004. "Small Farms, Externalities, aand the Dust Bowl of the 1930s." *Journal of Political Economy 112(3): 665-694.*

Hathaway, D.E. 1959. "Agriculture in an Unstable Economy Revisited." *Journal of Farm Economics* 41(3):487-499.

Hazell, P.B. 1989. "Changing Patterns of Variability in World Cereal Production." Chapter in J.R. Anderson and P.B. Hazell, eds., *Variability in Grain Yields*. Baltimore: Johns Hopkins University Press.

Herdt, R.W. 1970. "A Disaggregate Approach to Aggregate Supply." *American Journal of Agricultural Economics* 52(4):512-520.

Hicks, D.R. and P.R. Thomison. 2004. "Corn Management." Chapter in C.W. Smith, J. Betran and E.C.A. Runge, eds., *Corn: Origin, History, Technology and Production*. Hoboken, NJ: John Wiley and Sons.

Hijmans, R.J., L. Guarino, A. Jarvis, R. O'Brien, P. Mathur, C. Bussink, M. Cruz, I. Barrantes and E. Rojas. 2005. *DIVA-GIS Version 5.2 Manual*. Obtained November 2011 from apps.cimmyt.org.

Hirsch, H.G. 1942. "Crop Yield Index Numbers." M.S. Thesis, University of Minnesota, Minneapolis.

Hirsch, H.G. 1943. "Crop Yield Index Numbers." *Journal of Farm Economics* 25(3):583-598.

Hoover, D.M. 1961. "A Study of Agricultural Land Prices in the United States, 1911-1958." Ph.D. Dissertation, University of Chicago, Chicago.

Houck, J.P. and P.W. Gallagher. 1976. "The Price Responsiveness of U.S. Corn Yields." *American Journal of Agricultural Economics* 58(4):731-734.

Huddleston, H.F. 1955. "An Inverted Matrix Approach for Determining Crop-Weather Regression Equations." *Biometrics* 11(2):231-236.

Hunt, L.A. and K.J. Boote. 1998. "Data for Model Operation, Calibration and Evaluation." Chapter in G.Y. Tsuji, G. Hoogenboom and P.K. Thornton, eds. , *Understanding Options for Agricultural Production*. Norwell, MA: Kluwer.

Hunter, R.B., M. Tollenaar and C.M. Breuer. 1977. "Effects of Photoperiod and Temperature on Vegaetative and Reproductive Growth of a Maize (*Zea mays*) Hypric." *Canadian Journal of Plant Science* 57:1127-1133.

Ibach, D.B. and J.R. Adams. 1967. *Fertilizer Use in the United States by Crops and Areas, 1964 Estimates*. Statistical Bulletin No. 408. Washington, DC: United States Deptartment. of Agriculture, Economic Research Service and Statistical Reporting Service.

Ibach, D.B., J.R. Adams and E.I. Fox. 1964. *Commercial Fertilizer Used on Crops and Pasture in the United States, 1959 Estimates*. Statistical Bulletin No. 348. Washington, DC: United States Department of Agriculture, Economic Research Service and Agricultural Research Service.

Jarrett, F.G. 1957. "Resource Productivities and Production Functions." *Australian Journal of Agricultural Economics* 1:65-76.

Jaynes, D.B. and D.E. James. 2007. "The Extent of Farm Drainage in the United States." Paper presented at the Soil and Water Conservation Society annual meeting, Tampa. Obtained January 2012 from www.ars.usda.gov.

Jelinski, D. and J. Wu. 1996. "The Modifiable Areal Unit Problem and Implications for Landscape Ecology." *Landscape Ecology* 11(3):129-140.

Jones, C.A. and J.R. Kiniry. 1986. *CERES-Maize: a Simulation Model of Maize Growth and Development*. College Station, TX: Texas A&M University Press.

Jones, J.W., G.Y. Tsuji, G. Hoogenboom, L.A. Hunt, P.K. Thornton, P.W. Wilkens, D.T. Imamura, W.T. Bowen and U. Singh. 1998. "Decision Support System for Agrotechnology Transfer: DSSAT v3." Chapter in G.Y. Tsuji, G. Hoogenboom and P.K. Thornton, eds. , *Understanding Options for Agricultural Production*. Norwell, MA: Kluwer.

Katz, R.W. 1977. "Assessing the Impact of Climate Change on Food Production." *Climatic Change* 1(1):85-96.

Kaufmann, R.K. and S.E. Snell. 1997. "A Biophysical Model of Corn Yield: Integrating Climatic and Social Determinants." *American Journal of Agricultural Economics* 79(1):178-190.

Kaylen, M.S. and S.S. Koroma. 1991. "Trend, Weather Variables, and the Distribution of U.S. Corn Yields." *Review of Agricultural Economics* 13(2):249-258.

Kaylen, M.S., J.W. Wade and D.B. Frank. 1992. "Stochastic Trend, Weather and US Corn Yield Variability." *Applied Economics* 24: 513-518.

Kim, S. 1998. "Economic Integration and Convergence: U.S. Regions, 1840-1987." *Journal of Economic History* 58(3):659-589.

Kincer, J.B. and W.A. Mattice. 1928. "Statistical Correlations of Weather Influence on Crop Yields." *Monthly Weather Review* 56(2):53-57.

Kirkby, A.T. 1973. *The Use of Land and Water Resources in the Past and Present Valley of Oaxaca, Mexico*. Memoirs of the Museum of Anthropology No. 5. Ann Arbor, MI: University of Michigan.

Leff, B., N. Ramankutty and J.A. Foley. 2004. "Geographic Distribution of Major Crops Across the World." *Global Biogeochemical Cycles* 18:1-27.

LeSage, J.P. and R.K. Pace. 2010. "Spatial Econometric Models." Chapter in A. Getis and M.M. Fischer, eds., *Handbook of Applied Spatial Analysis*. New York: Springer-Verlag.

Levins, R.A. and J.K. Galbraith. 2003. *Willard Cochrane and the American Family Farm*. Lincoln, NE: University of Nebraska Press.

Lin, B.H., M. Padgitt, L. Bull, H. Delvo, D. Shank and H. Taylor. 1995. *Pesticide and Fertilizer Use and Trends in U.S. Agriculture*. Agricultural Economic Report No. 717. Washington, DC: United States Department of Agriculture, Economic Research Service.

Lizaso, J.I., K.J. Boote, J.W. Jones, C.H. Porter, L. Echarte, M.E. Westgate and G. Sonohat. 2011. "CSM-IXIM: A New Maize Simulation Model for DSSAT Version 4.5." *Agronomy Journal* 103(3):766-779.

Lobell, D.B. and G.P. Ansner. 2003. "Climate and Management Contributions to Recent Trends in U.S. Agricultural Yields." *Science* 299:1032.

Lobell, D.P. and G.P. Asner. 2003. "Response to Comment on Climate and Management Contributions to Recent Trends in U.S. Agricultural Yields." *Science* 300:1505c.

Lobell, D.B. and C.B. Field. 2007. “Global Scale Climate-Crop Yield Relationships and the Impacts of Recent Warming.” *Environmental Research Letters* 2:1-7.

Luttrell, C.B. and R.A. Gilbert. 1976. "Crop Yields: Random, Cyclical, or Bunchy?" *American Journal of Agricultural Economics* 58(3):521-531.

Matsuura, K. and C.J. Wilmott. 2009. *Terrestrial Precipitation: 1990-2008 Global Monthly Time Series (v. 2.01)*. Newark: Center for Climate Research. Obtained September 2011 from climate.geog.udel.edu.

Mattice, W.A. 1931. "Weather and Corn Yields." *Monthly Weather Review* 59(3):105-112.

Maizel, M., R.D. White, S. Gage, L. Osborne, R. Root, S. Stitt and G. Muehlbach. 1999. “Historical Interrelationships Between Population Settlement and Farmland in the Conterminous United States, 1790 to 1992.” Chapter in T. Sisk, ed., *Perspectives on the Land Use History of North America: A Context for Understanding our Changing Environment.* Biological Science Report USGS/BRD/BSR-1998-0003. United States Geological Survey, Biological Resoures Division.

Massetti, E. and R. Mendelsohn. 2011. “The Impact of Climate Change on US Agriculture: a Repeated Cross-Sectional Ricardian Analysis.” Chapter in A. Dinar and R. Mendelsohn, eds., *Handbook on Climate Change and Agriculture.* Northhampton, MA: Edward Elgar.

McCabe, G.J. and S.L. Markstrom. 2007. *A Monthly Water-Balance Model Driven by a Graphical User Interface*. Open-File Report No. 2007-1088. Reston, VA: U.S. Department of the Interior, U.S. Geological Survey.

Mehring, A.L., J.R. Adams and K.D. Jacob. 1957. *Statistics on Fertilizers and Liming Materials in the United States*. Statistical Bulletin No. 191. Washington, DC: United States Department of Agriculture, Agricultural Research Service.

Mendelsohn, R. and W.D. Nordhaus. 1999a. "The Impact of Global Warming on Agriculture: A Ricardian Analysis: Reply." *American Economic Review* 89(4):1046-1048.

Mendelsohn, R. and W.D. Nordhaus. 1999b. "The Impact of Global Warming on Agriculture: A Ricardian Analysis: Reply (Darwin)." *American Economic Review* 89(4):1053-1055.

Mendelsohn, R., W.D. Nordhaus and D. Shaw. 1994. "The Impact of Global Warming on Agriculture: A Ricardian Analysis." *American Economic Review* 84(4):753-771.

Menz, K.M. and P.G. Pardey. 1983. "Technology and U.S. Corn Yields: Plateaus and Price Responsiveness." *American Journal of Agricultural Economics* 65(3):558-562.

Mitchell, T.D. and P.D. Jones. 2005. "An Improved Method of Constructing a Database of Monthly Climate Observations and Associated High-Resolution Grids." *International Journal of Climatology* 25:693-712.

Mitchell, D. 2008. “A Note on Rising Food Prices.” Policy Research Working Paper No. 4682. Washington, DC: World Bank.

Moore, M.R., N.R. Gollehon and M.B. Carey. 1994. "Multicrop Production Decisions in Western Irrigated Agriculture: The Role of Water Price." *American Journal of Agricultural Economics* 76(4):859-874.

Moran, P.A.P. 1950. "Notes on Continuous Stochastic Phenomena." *Biometrika* 37(1/2):17-23.

Morgan, T.H., A.W. Biere and E.T. Kanemasu. 1980. "A Dynamic Model of Corn Yield Response to Water." *Water Resources Research* 16(1):59-64.

Muchow, R.C. 1990. "Effect of High Temperature on Grain-growth in Field-grown Maize." *Field Crops Research* 23(2):145-158.

Neild, R.E. 1982. “Temperature and Rainfall Influences on the Phenology and Yield of Grain Sorghum and Maize: A Comparison.” *Agricultural Meterology* 27:79-88.

Neild, R.E. and J.E. Newman. 1986. “Growing Season Characteristics and Requirements in the Corn Belt.” National Corn Handbook NCH-40. West Lafayette, IN: Purdue University.

Nelson, W.L. and R.F. Dale. 1978a. "Effect of Trend or Technology Variables and Record Period on Prediction of Corn Yields with Weather Variables." *Journal of Applied Meteorology* 17(7):926-933.

Nelson, W.L. and R.F. Dale. 1978b. "A Methodology for Testing the Accuracy of Yield Predictions from Weather-Yield Regression Models for Corn." *Agronomy Journal* 70(5):734-740.

NOAA-CPC (Climate Prediction Center). 2011a. *U.S. Daily Maximum Temperature Analysis*. Camp Springs, MD: National Oceanic and Atmospheric Administration. Obtained March 2011 from ftp.cpc.ncep.noaa.gov.

NOAA-CPC (Climate Prediction Center). 2011b. *U.S. Daily Minimum Temperature Analysis*. Camp Springs, MD: National Oceanic and Atmospheric Administration. Obtained March 2011 from ftp.cpc.ncep.noaa.gov.

NOAA-CPC (Climate Prediction Center). 2011c. *Unified Global Gauge Daily Precipitation Analysis*. Camp Springs, MD: National Oceanic and Atmospheric Administration. Obtained March 2011 from ftp.cpc.ncep.noaa.gov.

NOAA-NCDC (National Climatic Data Center). 2007. *Global Historical Climate Network (GHCN)-Monthly Version 2*. Asheville, NC: National Oceanic and Atmospheric Administration, National Climate Data Center. Obtained December 2008 from ftp.ncdc.noaa.gov.

Offutt, S.E., P. Garcia and M. Pinar. 1987. "Technological Advance, Weather, and Crop Yield Behavior." *North Central Journal of Agricultural Economics* 9(1):49-63.

Olmstead, A.L. and P.W. Rhode. 2002. "The Red Queen and the Hard Reds: Productivity Growth in American Wheat, 1800-1940." *Journal of Economic History* 62(4):929-966.

Openshaw, S. and P. Taylor. 1979. "A Million or so Correlation Coefficients: Three experiments on the Modifiable Area Unit Problem." Chapter in N. Wrigley, ed., *Statistical Applications in the Spatial Sciences*. London: Pion.

Oury, B. 1965. "Allowing for Weather in Crop Production Model Building." *Journal of Farm Economics* 47(2):270-283.

Pardey, P.G., J.M. Beddow, J. Koo and S. Wood. 2008. "Global Agricultural Trends: Impact and Role for Australian Agricultural and Agronomic Research." Paper presented at the 14th annual Australian Society of Agronomy Conference, Adelaide.

Pardey, P.G., J.M. Alston and V.W. Ruttan. 2010. "The Economics of Innovation and Technical Change in Agriculture." Chapter in B.H. Hall and N. Rosenberg, eds., *Handbook of the Economics of Innovation*. New York: Elsevier.

Parker, W.N. and J. Klein. 1966. "Productivity Growth in Grain Production in the United States, 1840-60 and 1900-10." Chapter in D.S. Brady, ed., *Output, Employment, and Productivity in the United States after 1800*. New York: National Bureau of Economic Research.

Peterson, W. 1986. "Land Quality and Prices." *American Journal of Agricultural Economics* 68(4):812-819.

Plaxico, J.S. 1955. "Problems of Factor-Product Aggregation in Cobb-Douglas Value Productivity Analysis." *Journal of Farm Economics* 37(4):664-675.

Plaxico, J.S. 1961. "Discussion: A Measure of the Influence of Weather on Crop Production." *Journal of Farm Economics* 43(5):1160-1162.

Qaim, M. 2009. “The Economics of Genetically Modified Crops.” *Annual Review of Resource Economics* 1:665-693.

Quiggin, J. and J.K. Horowitz. 1999. "The Impact of Global Warming on Agriculture: A Ricardian Analysis: Comment." *American Economic Review* 89(4):1044-1045.

Reilly, J., F. Tubiello, B. McCarl, D. Abler, R. Darwin, K. Fuglie, S. Hollinger, C. Izaurralde, S. Jagtap, J. Jones, L. Mearns, D. Ojima, E. Paul, K. Paustain, S. Riha, N. Rosenberg, and C. Rosenzweig. 2003. “U.S. Agriculture and Climate Change: New Results.” *Climatic Change* 57: 43-59.

Reynolds, H.D. 1998. "The Modifiable Area Unit Problem: Empirical Analysis by Statistical Simulation." Ph.D. Dissertation, University of Toronto, Toronto.

Richter, M.K. 1966. “Invariance Axioms and Economic Indexes.” *Econometrica* 34(4):739-755.

Robinson, W.S. 1950. "Ecological Correlations and the Behavior of Individuals." *American Sociological Review* 15(3):351-357.

Rose, J.K. 1936a. "Corn Yield and Climate in the Corn Belt." *Geographical Review* 26(1):88-102.

Rose, J.K. 1936b. "Intercorrelations Between Climatic Variables in the Corn Belt." *Monthly Weather Review* 64(3):76-82.

Runge, E.C.A. and R.T. Odell. 1958. "The Relation Between Precipitation, Temperature and the Yield of Corn on the Agronomy South Farm, Urbana, Illinois." *Agronomy Journal* 50(8):448-454.

Ruttan, V.W. 2002. “Productivity Growth in World Agriculture: Sources and Constraints.” *Journal of Economic Perspectives* 16(4):161-184.

Samuelson, P.A. 1947. *Foundations of Economic Analysis*. Cambridge: Harvard University Press.

Schery, R.W. 1972. *Plants for Man*. Englewood Cliffs, NJ: Prentice-Hall.

Schlenker, W. and M. Roberts. 2008. "Estimating the Impact of Climate Change on Crop Yields: The Importance of Nonlinear Temperature Effects." Working Paper 13799. Cambridge: National Bureau of Economic Research.

Schultz, T.W. 1951. "Farm Price Gyrations by Cochrane." *Journal of Farm Economics* 33(4):540-544.

Schultz, T.W. 1953. *The Economic Organization of Agriculture*. New York: McGraw-Hill.

Shaw, L.H. 1964. "The Effect of Weather on Agricultural Output: A Look at Methodology." *Journal of Farm Economics* 46(1):218-230.

Shaw, L.H. 1967. "Perspective in Estimating the Effects of Weather: Comment." *Journal of Farm Economics* 49(4):935-937.

Smith, J.W. 1904. "Relation of Precipitation to Yield of Corn". In *United States Department of Agriculture Yearbook 1903*. Washington, DC: Government Printing Office.

Smith, J.W. 1914. "The Effect of Weather Upon the Yield of Corn." *Monthly Weather Review* 42(2):78-92.

Snyder, J.P. 1987. "Map Projections-A Working Manual." U.S. Geological Survey Professional Paper No. 1395. Washington, DC: U.S. Geological Survey.

Stallings, J.L. 1958. "Indexes of the Influence of Weather on Agricultural Output." Ph.D. Dissertation, Michigan State University,

Stallings, J.L. 1960. "Weather Indexes." *Journal of Farm Economics* 42(1):180-186.

Stallings, J.L. 1961. "A Measure of the Influence of Weather on Crop Production." *Journal of Farm Economics* 43(5):1153-1160.

Streck, N.A., I. Lago, L.F. Gabriel and F.K. Samboranha. 2008. "Simulating Maize Phenology as a Function of Air Temperature with a Linear and a Nonlinear Model." *Pesquisa Agropecuária Brasileira* 43:449-455.

Sutch, R.C. 20008. "Henry Agard Wallace, the Iowa Corn Yield Tests, and The Adoption of Hybrid Corn." Working Paper 14141. Cambridge: National Bureau of Economic Research.

Sutherst, R.W. and G.F. Maywald. 1985. "A Computerised System for Matching Climates in Ecology." *Agriculture, Ecosystems & Environment* 13(3-4):281-299.

Sutherst, R.W., G.F. Maywald and D. Kriticos. 2007. *Climex Version 3 User's Guide*. Melbourne: Hearne Scientific Software. Obtained January 2010 from www.hearne.com.au.

Swanson, E.R. and J.C. Nyankori. 1979. "Influence of Weather and Technology on Corn and Soybean Yield Trends." *Agricultural Meteorology* 20(4):327-342.

Thirtle, C. and P. Bottomley. 1988. "Is Publicly Funded Agricultural Research Excessive?" *Journal of Agricultural Economics* 39(1):99-111.

Thompson, L.M. 1969a. "Weather and Technology in the Production of Corn in the U.S. Corn Belt." *Agronomy Journal* 61(3):453-456.

Thompson, L.M. 1969b. "Weather and Technology in the Production of Soybeans in the Central United States." *Agronomy Journal* 62:232-236.

Thompson, L.M. 1970. "Weather and Technology in the Production of Wheat in the United States." *Journal of Soil and Water Conservation* 24:219-224.

Thornthwaite, C.W. 1948. "An Approach Toward a Rational Classification of Climate." *Geographical Review* 38(1):55-94.

Trabucco, A. and R.J. Zomer. 2010. *Global Soil Water Balance Geospatial Database*. CGIAR Consortium for Spatial Information. Obtained August 2010 from www.cgiar-csi.org.

Trostle, R. 2008. *Global Agricultural Supply and Demand: Factors Contributing to the Recent Increase in Food Commodity Prices*. Economic Research Service Report WRS-0801. Washington, DC: United States Department of Agriculture.

USDA-ARS (Agricultural Research Service). 1957. *Fertilizer Used on Crops and Pasture in the United States, 1954 Estimates*. StatisticalBulletin No. 216. Washington,DC: United States Department of Agriculture.

USDA-BAE (Bureau of Agricultural Economics). 1948. *Tobaccos of the United States*. Washington, DC: U.S. Government Printing Office.

USDA-ERS (Economic Research Service). 1994. *Agricultural Resources and Environmental Indicators*. Agricultural Handbook No. 705. Washington, DC: U.S. Dept. of Agriculture.

USDA. 1998. *Agriculture Fact Book 1998*. Washington, DC: United States Department of Agriculture, Office of Communications.

USDA-ERS (Economic Research Service). 2011a. *Adoption of Genetically Engineered Crops in the U.S.* Washington, DC: United States Department of Agriculture. Obtained January 2011 from www.ers.usda.gov.

USDA-ERS (Economic Research Service). 2011b. *Fertilizer Use and Price Data Files*. Washington, DC: United States Department of Agriculture. Obtained January 2011 from www.ers.usda.gov.

USDA-NASS (National Agricultural Statistics Service). 1997. *Usual Planting and Harvesting Dates for U.S. Field Crops*. Agricultural Handbook No. 628. Washington DC: United States Department of Agriculture.

USDA-NASS (National Agricultural Statistics Service). 2011a. *Agricultural Chemical Use Database*. Washington, DC: United States Department of Agriculture. Obtained January 2011 from www.pestmanagement.info.

USDA-NASS (National Agricultural Statistics Service). 2011b. *QuickStats*. Washington, DC: United States Department of Agriculture. Obtained on various dates from quickstats.nass.usda.gov.

USDA-NRCS (National Resources Conservation Service). 2011. *Soil Survey Geographic (SSURGO) Database*. Washington, DC: United States Department of Agriculture. Obtained July 2011from soils.usda.gov.

Vavilov, N.I. 1992. *Origin and Geography of Cultivated Plants*. Cambridge: Cambridge University Press.

Wallace, H.A. 1920. "Mathametical Inquiry into the Effects of Weather on Corn Yield in the Eight Corn Belt States." *Monthly Weather Review* 48(8):439-446.

Ward, M.D. and K.S. Gleditsch. 2008. *Spatial Regression Models*. Los Angeles: Sage.

Warren, G.F. 1998. "Spectacular Increases in Crop Yields in the United States in the Twentieth Century." *Weed Technology* 12(4):752-760.

Warrington, I.J. and E.T. Kanemasu. 1983. "Corn Growth Response to Temperature and Photoperiod I. Seedling Emergence, Tassel Initiation, and Anthesis." *Agronomy Journal* 75(5):749-754.

Westcott, N. and K. Grady. 2010. "Impacts of the 1954 Summer Heat Wave." Paper presented at the 18th American Meterological Society Conference on Applied Climatology, Atlanta.

Wilkes, G. 2004. "Corn, Strange and Marvelous: But Is a Definitive Origin Known?" Chapter in C.W. Smith, J. Betran and E.C.A. Runge, eds., *Corn: Origin, History, Technology and Production*. Hoboken, NJ: John Wiley and Sons.

Wilson, W.M. 1914. "The Climate of New York in Relation to Agriculture." *Cornell Countryman* 12:100-108.

Wolfe, T.K. 1925. "A Mathematical Inquiry into the Influence of the Amount and Distribution of Rainfall on the Yield of Corn." *Agronomy Journal* 17(6):356-362.

Working, E.J. 1940. "Crop-Yield Index Numbers." *Journal of Farm Economics* 22(4):701-713.

Wright, B.D. 2011. "The Economics of Grain Price Volatility." *Applied Economic Perspectives and Policy* 33 (1): 32-58.

You, L. and S. Wood. 2006. "An Entropy Approach to Spatial Disaggregation of Agricultural Production." *Agricultural Systems* 90:329-347.

You, L., Z. Guo, J. Koo, W. Ojo, K. Sebastian, M.T. Tenorio, S. Wood and U. Wood-Sichra. 2011. *Spatial Productio Allocation Model (SPAM) 2000 Version 3 Release 2*. Obtained March 2011 from www.mapspam.info.

Zalom, F.G., P.B. Goodell, L.T. Wilson, W.W. Barnett and W.J. Bentley. 1983. "Degree Days: The Calculation and Use of Heat Units in Pest Management." Leaflet No. 21373. Berkeley, CA: University of California Division of Agriculture and Natural Resources.

Zeileis, A., C. Kleiber, W. Krämer and K. Hornik. 2003. "Testing and Dating of Structural Changes in Practice." *Computational Statistics and Data Analysis* 44(1-2):109-123.

Zeynep, K.H. and G.D. Libecap. "Small Farms, Externalities, and the Dust Bowl of the 1930s." *Journal of Political Economy* 112(3):665-694.

Zhang, B. and C.A. Carter. 1997. "Reforms, the Weather, and Productivity Growth in China's Grain Sector." *American Journal of Agricultural Economics* 79(4):1266-1277.

Appendix

A.1 A Note About Growth Rates

The actual growth rate over any period can be calculated using only the beginning and ending data points. Under continuous compounding at a constant growth rate, the amount of a quantity, x, is defined by:

$$x_t = x_s e^{r(t-s)}; t > s\,, \tag{A.1.1}$$

where x_s and x_t are the amounts at times s and t, respectively, and r is a continuously compounded growth rate between the two periods. Thus, the growth rate in a variable x between any two periods can be calculated as:

$$r = \frac{\ln(x_t / x_s)}{t-s} \tag{A.1.2}$$

Some difficulties arise when one needs to describe the growth rate between two periods. One option is to apply a log difference approach by taking the natural logarithm of equation A.1.1, which calculated the period growth rate using only the beginning and ending points of the period. This approach calculates the growth rate for the period under the assumption that the growth rate was constant. However, the calculation discards all of the data in $\{x_{s+1} \ldots x_{t-1}\}$, and therefore may not represent the *typical* growth rate over the period.

Since $\ln(x_t / x_s) = \ln(x_t) - \ln(x_s)$, a convenient way to calculate period-to-period growth rates in a time series is to subtract the lag of the natural log of the series from the natural log of the series:

$$r_t = \ln(x_t / x_{t-1}) / (t - (t-1)) = \ln(x_t) - \ln(x_{t-1}) \;\; \forall t \tag{A.1.3}$$

Since the year-to-year growth rates are calculated individually via this formula, it is tempting to calculate the average growth rate for a time period by taking the mean of the annual growth rates. At first glance this seems to take account of the data that are discarded using the log difference of the endpoints, however the result still depends only on the endpoints—the average of a series of log difference growth rates is exactly equivalent to applying equation A.1.1. For example, consider the growth in a series $\{x_1, x_2, \ldots, x_t\}$. Applying equation A.1.3 and taking the average yields:

$$\frac{\ln(x_2)-\ln(x_1)+\ln(x_3)-\ln(x_2)+\ldots+\ln(x_t)-\ln(x_{t-1})}{t-1}.$$

All terms in the numerator except for the endpoints $\{x_1, x_t\}$ will be netted out, and the result will exactly equal that obtained from a natural logarithmic transformation of equation A.1.1. The options therefore are to either apply equation A.1.1 or to estimate a regression of the log of the variable on a time trend. Which is appropriate depends on the question at hand. If the concern is with the typical growth rate over the period or the extent to which year-to-year growth in a period differs from the overall trend, then the semi-log regression might provide more useful information. However, if one wishes to describe *how much* a variable increased over the period, the log-difference approach might be more appropriate.

A.2 Corn Area and Production Dataset

There are two primary sources for U.S. county-level agricultural production data: the various agricultural and general censuses, and survey data compiled by the National Agricultural Statistics Service (NASS) (see Table 6). While NASS provides annual data, they are not consistently provided for all counties. The NASS county-level crop data collection expanded from a pilot program in Nebraska in 1910 to include additional counties, but has never provided data for all counties. By about 1972, the NASS survey program had grown to encompass nearly 100 percent of both corn area and production. However, the number of counties in the dataset has been shrinking since the early 1970s such that only about 85 percent of U.S. corn area and output were accounted for by 2006. For more details see Figure 27, which shows the percentage of national corn harvested area and the number of counties represented in the NASS county-level data.

Table 6: Sources of Long-Run U.S. County Level Corn Data

Source	Frequency	Time Series	County Coverage
General Census	Decennial	1879-1919	All, plus most territories
Agricultural Census	Petennial	1924-2007	All
NASS	Annual	1910-2010	Sample, varies

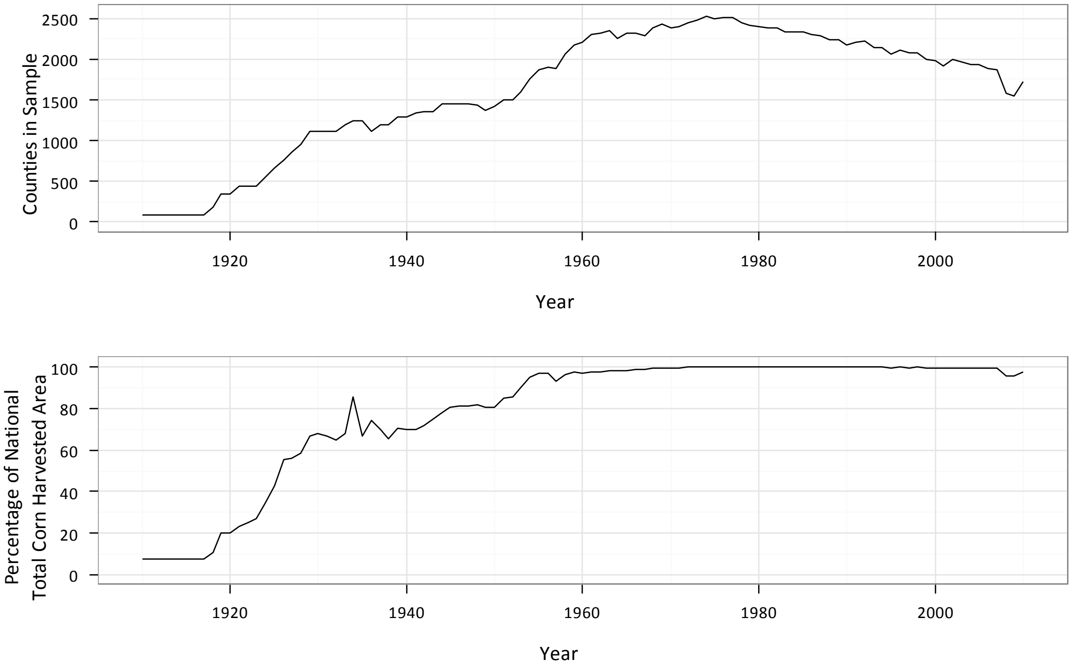
The NASS data are informative, but their coverage (especially before 1940) is not sufficient to conduct a national, spatially-explicit, long-run assessment. However, the census data do provide complete county-level data compiled at regular intervals.[74]

---

[74] The census data provide coverage for all but a small minority of counties which are left out of the data, generally to assure privacy of respondents.

Surprisingly, while the county-data have long been available in printed form, no long-run, county level datasets had been compiled when this work was initiated. The preparatory work for the present study required manual entry of data from census books, processing electronic census datasets (available in disparate formats) and geo-referencing the data to annual county vector files. The current set of county data has been geo-referenced onto maps corresponding as close as possible to the data years.[75] However, county boundaries changed (in many cases significantly) between 1880 (the first data year) and 2007. While the data are complete, they do not form a consistent panel since the units of analysis differ over time. While not a problem per se, this limits the analytical procedures that can be employed.

Figure 27: NASS County-Level Survey Data Coverage



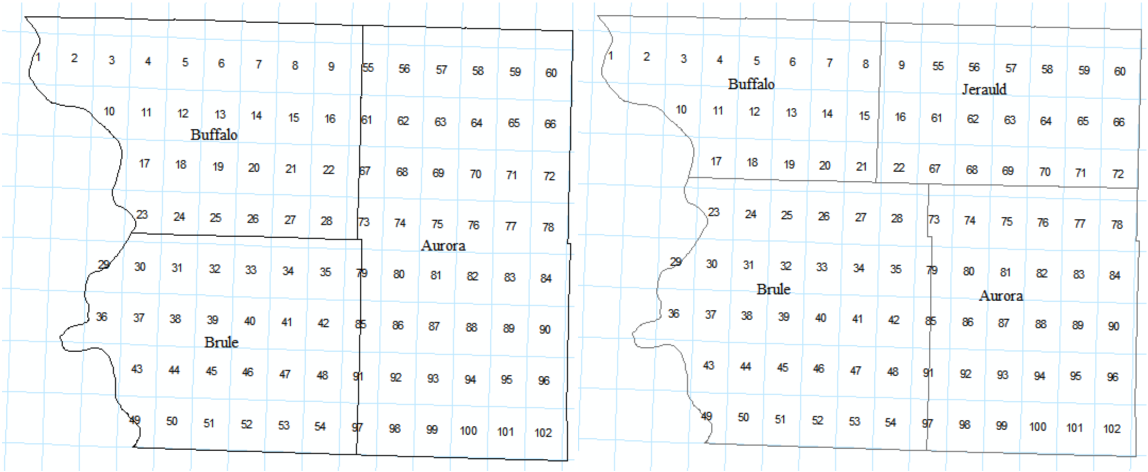
Source: NASS (2011).

---

[75] Shapefiles for each decade from 1880 to 2000 were provided by the National Historical GIS project.

To enable a richer analysis, a procedure was developed to derive a spatially consistent panel from the county data. First, the county data for the coterminous United States were mapped to county boundary files provided by the National Historical GIS project (NHGIS). The boundary files were available for each decade beginning with 1880, and the boundary file that best matched the data year was used for each census dataset. A five arcminute grid was then overlaid on each boundary file, and each grid cell was assigned to the county that contained the largest portion of the cell. As a concrete example, consider Figure 28, which shows the Dakota Territory counties of Buffalo, Brule and Aurora with their 1880 boundaries (Panel a) and their year 2000 boundaries (Panel b). The five arcminute grid is overlaid on these counties and cells 1-28 are assigned to Buffalo county, cells 29-54 to Brule county and the remaining cells to Aurora county. By 2000, Brule County had incorporated portions of Buffalo County, and a new county, Jerauld, was formed out of some area from both Aurora and Buffalo Counties (all now in the state of South Dakota). The same five arcminute grid was overlaid on the year 2000 boundaries, and cells were assigned to each county. The result of this allocation is shown in Table 7.

Figure 28: Area of Buffalo, Brule, Jerauld and Aurora Counties, 1880 and 2000



Panel A: 1880 Panel b: 2000

Table 7: Example Allocation of 5' Pixels

County	1880	2000
Aurora	55-102	73-102
Brule	29-54	23-54
Buffalo	Jan-28	1-8, 10-15 and 17-21
Jerauld	N/A	9, 16, 22 and 55-72

Now consider a simple method for mapping 1880 data on the year 2000 boundaries. Suppose some variable, x_t^i, is observed in each county, *i*, in year *t* and that each x_t^i is the a sum of (possibly) numerous observations within the county (for example, acreage or output as reported by the census). If the component observations that form x_t^i are distributed equally within the county, it would be reasonable to disaggregate data by distributing x_t^i equally among the grid cells that make-up the county. So, for example, if ac_{1880}^{Aurora} represents the 1880 corn acreage in Aurora county, each of the county-year's 47 grid cells could be assumed to have contributed $\frac{ac_{1880}^{Aurora}}{47}$

acres to the total. This value could then be assigned to each cell. If a similar procedure were followed for Brule and Buffalo counties, the 1880 pixilated values could be re-aggregated using a table similar to Table 7 to represent the year 2000 boundaries.

In the example above, each county-year is broken into a set of pixels, $P_t^i \equiv \{p_{t,1}^i, p_{t,2}^i, \ldots, p_{t,m}^i, \ldots, p_{t,m_t^i}^i\}$ so that $\bigcup_{i=1}^{N_t} P_t^i = P_t$, where P_t represents the pixels that comprise the conterminous United States and m_t^i is the number of pixels in county *i* in year *t*. Since the boundaries of the conterminous U.S. and its included territories did not change over the period, $P_t = P_{2000} \forall t$. Further, a pixel is assigned to at most one county in any given year[76] so that $\bigcap_{i=1}^{N_t} P_i^t = \varnothing \quad \forall t$, which greatly simplifies calculations. It is convenient to define a vector W_t^i of numeric weights for the corresponding pixels so that W_t^i is a row-vector of length m_t^i. In the example above, each weight was set to $1/m_t^i$ so that $1_m' W_t^i = 1 \; \forall i,t$ and $W_t^i \geq 0$, where 1_m is a column vector of ones.

However, many counties contain areas that do not produce corn, and it is therefore preferable to not evenly smooth acreage and output across the county. The HarvestChoice Spatial Allocation Model (SpAM) provides a spatially disaggregated estimate of global production of various crops, including corn. The model uses data on prices, population density, cropping systems, locations of cropland and other

---

[76] For each year, pixels were assigned to the county that contained the largest portion of the pixel.

variables to disaggregate sub-national production and area data onto a 5-arcminute grid. The data are allocated to pixels based on a revenue-maximization objective subject to adding-up constraints (see You and Wood 2006 for details).

The SpAM data represent a spatially disaggregated estimate of year-2000 production. The SpAM estimates of harvested corn acreage were used to disaggregate each county-year's area and production data onto its constituent pixels. Let $S(\cdot)$ be a function which, given a vector of pixels, returns a vector of corresponding SpAM values. Then, for example, $S(P_t^i)$ returns a vector of estimated harvested acreage for each pixel in P_t^i. The pixel weights for a county-year are then determined by normalizing the SpAM values so that their sum equals one: $W_t^i = S(P_t^i)\cdot\left[1_m' S(P_t^i)\right]^{-1}$ (note that the bracketed term is a scalar). Then, a county-level scalar variable x_t^i can be disaggregated by multiplication: $x_t^i W_t^i$.

A variable for any given year, s, can be mapped onto the county geometry of another year, τ, by first disaggregating, then re-aggregating using the pixellated landscape. For example, let X_s represent an ordered vector of all disaggregated pixels for year s:

$$X_s = \begin{pmatrix} x_s^1 W_s^1 \\ x_s^2 W_s^2 \\ \vdots \\ x_s^{N_s} W_s^{N_s} \end{pmatrix}$$

X_s can then be aggregated to the year-τ geography by summing:

$$x_{s,\tau}^j = \sum_{k \in P_\tau^j} X_s^k \ \forall j,$$

where $x_{s,\tau}^{j}$ is the year-*s* value for county *j* with year-τ boundaries, and *k* represent the elements of X_s.

A.3 Equiproportional vs. Relative Changes in a Laspeyres Area Index

A Laspeyres area index is defined as $I_A^L = \frac{\mathbf{y}_b'\mathbf{a}_t}{\mathbf{y}_b'\mathbf{a}_b} = \frac{\sum_{i=1}^{n} y_{bi}a_{ti}}{\sum_{i=1}^{n} y_{bi}a_{bi}}$. First consider an increase in area such that each county increases its area by a scalar $p > 0$ percent. This equi-proportional increase in area will generate an index value of $\frac{\mathbf{y}_b'\mathbf{a}_b(1+p)}{\mathbf{y}_b'\mathbf{a}_b} = (1+p)$. Suppose the increase was not equi-proportional, and that county *N* increased its area by some other amount, *α* percent, but that the overall increase in area remained the same. For this condition to hold, one must derive the amount by which the other counties increased their area. Let that parameter be called γ. Then, the total area increase will be maintained if:

$$(1+\gamma)\sum_{i=1}^{n-1} a_{bi} + (1+\alpha)a_{bN} = (1+p)\sum_{i=1}^{n} a_{bi}$$

Solving for γ reveals:

$$(1+\gamma)\sum_{i=1}^{n-1} a_{bi} = \sum_{i=1}^{n} a_{bi} + p\sum_{i=1}^{n} a_{bi} - a_{bN} - \alpha a_{bN}$$

$$(1+\gamma)\sum_{i=1}^{n-1} a_{bi} = \sum_{i=1}^{n-1} a_{bi} + p\sum_{i=1}^{n} a_{bi} - \alpha a_{bN}$$

$$\gamma = p + \frac{a_{bN}(p-\alpha)}{\sum_{i=1}^{N-1} a_{bi}}$$

The goal is to determine how much the index value differs when a change in national area is distributed evenly across counties from its value when the total area changes by the same amount, but is not equally distributed across counties. The numerator of the index will be:

$$\sum_{i=1}^{N-1}(1+\gamma)a_{bi}y_{bi}+(1+\alpha)a_{bN}y_{bN}$$

Subtracting the numerator for a proportional change in area and manipulating the expression yields:

$$(1+\gamma)\sum_{i=1}^{N-1}a_{bi}y_{bi}+(1+\alpha)a_{bN}y_{bN}-(1+p)\sum_{i=1}^{N}a_{bi}y_{b}$$

$$\sum_{i=1}^{N-1}a_{bi}y_{bi}+\gamma\sum_{i=1}^{N-1}a_{bi}y_{bi}+(1+\alpha)a_{bN}y_{bN}-\sum_{i=1}^{N}a_{bi}y_{bi}-p\sum_{i=1}^{N}a_{bi}y_{bi}$$

$$\gamma\sum_{i=1}^{N-1}a_{bi}y_{bi}+\alpha a_{bN}y_{bN}-p\sum_{i=1}^{N}a_{bi}y_{bi}$$

And, substituting for γ:

$$\gamma\sum_{i=1}^{N-1}a_{bi}y_{bi}+\alpha a_{bN}y_{bN}-p\sum_{i=1}^{N}a_{bi}y_{bi}$$

$$p\sum_{i=1}^{N-1}a_{bi}y_{bi}+\frac{a_{bN}(p-\alpha)}{\sum_{i=1}^{N-1}a_{bi}}\sum_{i=1}^{N-1}a_{bi}y_{bi}+\alpha a_{bN}y_{bN}-p\sum_{i=1}^{N}a_{bi}y_{bi}$$

$$\frac{a_{bN}(p-\alpha)}{\sum_{i=1}^{n-1}a_{bi}}\sum_{i=1}^{N-1}a_{bi}y_{bi}+\alpha a_{bN}y_{bN}-p(a_{bN}y_{bN})$$

$$\frac{a_{bN}(p-\alpha)}{\sum_{i=1}^{N-1}a_{bi}}\sum_{i=1}^{N-1}a_{bi}y_{bi}+a_{bN}y_{bN}(\alpha-p)$$

$$a_{bN}(p-\alpha)\sum_{i=1}^{N-1}s_{bi}y_{bi}-a_{bN}y_{bN}(p-\alpha),$$

where s_{bi} is county i's share of the total acreage in counties 1…(N-1). Finally, the numerator of the index for the unequal distribution will be equal to the numerator in the equal distribution plus the following term:

$$a_{bN}(p-\alpha)\left(\sum_{i=1}^{N-1} s_{bi} y_{bi} - y_{bN}\right)$$

Thus, the non-proportional increase in acreage alters the calculated index value by

$$\frac{a_{bN}(p-\alpha)\left(\sum_{i=1}^{N-1} s_{bi} y_{bi} - y_{bN}\right)}{\sum_{i=1}^{N} a_{bi} y_{bi}}.$$

Notice that the bracketed term is equal to the area weighted average yield of counties whose acreage increased by γ percent (hereafter denoted μ_{-N}.) less the yield for the county whose area increased by α percent. Thus, there are several possible scenarios:

1. If $\alpha > p$, county N took more than its area weighted share of the new acres and the first bracketed term will be negative. In this case,
    a. if $y_{bN} < \mu_{-N}$, the yield in county N is less than the area-weighted average of yields in other counties. The second bracketed term will therefore be positive. The index will decrease since a low-yielding county increased its share of area.
    b. If $y_{bN} > \mu_{-N}$, the yield in county N is higher than μ_{-N}. The second bracketed term is also negative and the index will increase since a high-yielding county increased its area share.
    c. If $y_{bN} = \mu_{-N}$, the second bracketed term is equal to zero, and there is no effect on the index.
2. If $\alpha < p$, county N took less than its share of the new acres, and the first term is positive.

    a. If in addition, $y_{bN} < \mu_{-N}$, the second term will be positive and the index will show an increase since acres were allocated away from a low-yielding county.
    b. If $y_{bN} > \mu_{-N}$, the overall effect will be negative.
3. If $\alpha = p$, county *N* increased its area in the same relative proportion as the other counties. The first term will equal zero and there will be no effect on the index.

A.4 Additional Tables and Figures

Table 8: Corn Production by Region, 1879-2007

Year	Appalachian	Corn Belt	Delta	Lake	Mountain	Northeast	North Plains	Pacific	Southeast	South Plains	National
Million bushels											
1879	207	1,031	55	82	1	111	173	2	64	29	1,755
1889	209	1,023	73	88	2	90	489	3	77	69	2,122
1899	229	1,321	105	145	2	115	474	2	92	178	2,665
1909	240	1,276	92	170	7	101	395	2	98	170	2,552
1919	242	1,111	94	174	17	115	293	5	131	162	2,346
1924	191	811	66	107	16	66	364	2	98	102	1,824
1929	202	994	81	146	25	64	404	2	101	111	2,131
1934	206	543	74	102	4	72	13	3	106	47	1,169
1939	213	1,320	92	255	10	75	151	3	97	95	2,311
1944	200	1,420	79	308	16	69	507	2	109	78	2,789
1949	227	1,478	73	356	14	83	374	2	106	63	2,776
1954	164	1,481	42	394	8	91	326	8	66	31	2,612
1959	222	2,152	55	520	19	101	449	18	116	43	3,695
1964	189	2,115	31	437	14	100	325	11	112	24	3,359
1969	207	2,699	13	568	36	150	617	19	105	28	4,442
1974	240	2,437	6	587	59	177	602	23	152	69	4,353
1978	322	3,734	8	1,042	97	244	1,050	42	144	118	6,801
1982	377	4,199	9	1,199	116	280	1,015	64	123	112	7,492
1987	248	3,635	34	1,069	119	217	1,139	41	87	130	6,719
1992	355	4,868	66	1,180	151	242	1,472	43	118	195	8,690
1997	280	4,328	115	1,417	165	227	1,784	60	95	243	8,714
2002	254	4,399	144	1,610	127	144	1,604	45	55	219	8,601
2007	383	6,467	341	1,864	184	266	2,721	65	112	321	12,723

Source: Calculated using described in Section A.2.

Table 9: Corn Harvested Area by Region, 1879-2007

Year	Appala-cian	Corn Belt	Delta	Lake	Mount-ain	North-east	North Plains	Pacific	South-east	South Plains	National
Million acres											
1879	10.6	28.2	3.6	2.4	0.1	3.6	5.1	0.1	6.3	2.5	62.4
1889	10.3	28.3	4.2	3.0	0.2	2.9	13.6	0.1	6.4	3.1	72.1
1899	12.0	35.8	5.9	4.4	0.1	3.5	16.8	0.1	8.6	7.5	94.8
1909	11.6	35.2	6.0	5.1	0.5	3.2	17.6	0.1	8.1	11.0	98.4
1919	11.2	30.5	6.5	4.8	1.1	2.8	13.3	0.2	10.1	7.2	87.8
1924	9.3	28.3	5.1	4.0	1.5	2.0	17.2	0.1	8.7	6.3	82.3
1929	9.0	28.9	4.9	4.2	1.6	1.8	17.9	0.1	7.9	7.0	83.2
1934	9.9	21.4	6.6	3.8	0.4	2.0	1.7	0.1	10.4	5.8	62.2
1939	9.1	27.6	6.7	5.8	0.8	2.0	9.5	0.1	9.8	6.0	77.4
1944	8.4	31.9	5.0	7.3	0.9	2.0	15.8	0.1	7.8	5.2	84.3
1949	7.2	31.1	3.7	7.6	0.6	1.8	13.2	0.0	6.4	3.3	75.1
1954	6.3	29.1	2.6	7.6	0.3	2.0	11.4	0.1	5.4	1.9	66.8
1959	5.3	34.0	1.8	9.5	0.4	1.8	10.9	0.2	4.6	1.5	70.0
1964	3.5	28.8	0.9	7.3	0.2	1.6	7.4	0.1	3.1	0.8	53.7
1969	3.0	28.7	0.4	6.7	0.4	1.7	7.9	0.2	2.8	0.6	52.5
1974	3.1	32.0	0.1	9.2	0.7	2.1	9.6	0.2	2.8	0.8	60.7
1978	4.1	35.4	0.2	11.1	0.9	2.6	11.1	0.4	2.8	1.3	70.0
1982	4.0	36.0	0.1	11.7	0.9	2.9	10.8	0.5	1.7	1.1	69.7
1987	3.0	29.4	0.4	9.5	0.8	2.3	10.4	0.3	1.2	1.3	58.6
1992	3.2	35.0	0.6	11.2	1.1	2.2	12.7	0.3	1.3	1.7	69.3
1997	2.9	34.2	1.0	11.5	1.2	2.3	14.7	0.4	1.0	1.8	70.9
2002	2.7	33.2	1.2	11.4	0.9	1.9	14.0	0.3	0.7	2.0	68.1
2007	3.5	40.2	2.2	13.4	1.3	2.3	19.7	0.3	1.1	2.2	86.1

Source: Calculated using described in Section A.2.

Table 10: Corn Yield by Region, 1879-2007

Year	Appalachian	Corn Belt	Delta	Lake	Mountain	Northeast	North Plains	Pacific	Southeast	South Plains	National
Bushels per acre											
1879	20	37	15	34	17	31	34	27	10	12	28
1889	20	36	17	29	14	31	36	30	12	22	29
1899	19	37	18	33	17	33	28	25	11	24	28
1909	21	36	15	34	16	32	22	24	12	15	26
1919	22	36	15	36	15	41	22	28	13	22	27
1924	21	29	13	27	11	33	21	30	11	16	22
1929	22	34	16	35	15	35	23	33	13	16	26
1934	21	25	11	27	10	35	8	33	10	8	19
1939	23	48	14	44	13	37	16	32	10	16	30
1944	24	45	16	42	18	35	32	34	14	15	33
1949	31	48	19	47	23	46	28	43	17	19	37
1954	26	51	16	52	28	47	29	58	12	17	39
1959	42	63	30	55	50	57	41	71	25	28	53
1964	55	73	34	60	61	64	44	82	36	31	63
1969	68	94	33	85	88	86	78	95	37	45	85
1974	77	76	45	64	90	83	62	104	55	83	72
1978	78	105	50	94	104	93	94	115	51	89	97
1982	95	117	68	102	122	96	94	138	74	101	108
1987	81	124	95	112	140	93	109	159	74	101	115
1992	111	139	106	106	139	108	116	167	90	118	125
1997	97	127	114	123	143	99	121	170	93	133	123
2002	95	133	122	141	148	76	115	175	80	111	126
2007	111	161	157	139	139	117	138	193	100	146	148

Source: Calculated using described in Section A.2.

Table 11: Index Values

	Simple		Paasche			Laspeyres		
Year	I_O^S	I_A^S	I_Y^P	I_A^P	I_R^P	I_Y^L	I_A^L	I_R^L
				(index value, 1909=100)				
1879	69	63	101	67	106	102	68	107
1889	83	73	110	77	105	108	75	103
1899	104	96	107	97	101	107	98	102
1909	100	100	100	100	100	100	100	100
1919	92	89	104	89	100	103	89	99
1924	71	84	86	83	99	87	83	99
1929	83	85	99	86	102	97	84	100
1934	46	63	74	76	120	60	62	98
1939	91	79	113	83	106	109	80	102
1944	109	86	120	93	109	117	91	106
1949	109	76	128	83	109	131	85	111
1954	102	68	132	85	125	121	78	114
1959	145	71	170	82	116	176	85	120
1964	132	55	194	67	122	198	68	124
1969	174	53	260	66	124	263	67	125
1974	171	62	221	72	117	237	77	125
1978	266	71	302	88	123	303	88	124
1982	294	71	328	83	117	355	90	126
1987	263	60	355	70	117	376	74	125
1992	340	70	389	78	111	438	88	124
1997	341	72	387	82	114	416	88	122
2002	337	69	395	84	121	401	85	123
2007	499	88	472	101	116	491	106	121

Source: Calculated using the data described in Section A.2.

Table 12: Listing of States by Region

Appalachian	Corn Belt	Delta States	Lake States
Kentucky	Illinois	Arkansas	Michigan
North Carolina	Indiana	Louisiana	Minnesota
Tennessee	Iowa	Mississippi	Wisconsin
Virginia	Missouri		
West Virginia	Ohio		

Mountain	Northeast	Northern Plains	Pacific
Arizona	Connecticuit	Kansas	California
Colorado	Delaware	Nebraska	Oregon
Idaho	Maine	North Dakota	Washington
Montana	Maryland	South Dakota	
Nevada	Massachuttsets		
New Mexico	New Hampshire		
Utah	New Jersey		
Wyoming	New York		
	Pennsylvania		
	Rhode Island		
	Vermont		

Southeast	Southern Plains
Alabama	Oklahoma
Florida	Texas
Georgia	
South Carolina	

Source: Regions are based on USDA (1998, p. 18).

A.5 Estimated Soil Moisture Capacity

An attempt was made to estimate the average soil water holding capacity for the corn growing areas in each county. Data on soil layers, their depth, and each layer's fractional water holding capacity were obtained from the National Resources Conservation Service's SSURGO database (USDA-NRCS 2011). These data provide variables related to local soils on a relatively fine 1 km grid for the United States. The water holding capacity (SWC_{max}) was calculated for each of the 1 km^2 grid cells then aggregated to the county level using procedures similar to those applied to the production and area data (Section A.2). Figure 29 presents a stylized map of SWC_{max} on the 1 km^2 grid, where darker values indicate a higher overall soil moisture holding capacity. Ultimately, however, the author was uncertain about the methodology applied and decided to use the methods of the CLIMEX model (Sutherst and Maywald 1985) and Trabucco and Zomer (2010), which use a spatially-invariant soil moisture capacity. Incorporating county-specific soil moisture capacities into the corn model is left for future work. It is expected that the SWC_{max} values that were estimated correlate with the correct SWC_{max} values. The correspondence between Figure 29 and the high-yielding corn areas implies that taking account of differences in soil might add to the accuracy of the model.

Figure 29: Estimated Soil Moisture Capacity

Source: Derived from USDA-NRCS (2011).

Notes: Darker areas indicate higher estimated soil moisture holding capacity.