The Evolution of Heat Tolerance of Corn: Implications for Climate Change

Michael J. Roberts♣ and Wolfram Schlenker♠

May 2009

Preliminary Draft - Please Do not Cite

Abstract

We examine how precipitation, moderate temperatures, and extreme temperatures influenced corn yields in Indiana between 1901-2005. Using a fine-scale weather data set of daily weather records we find that the effects of precipitation and extreme heat evolved over time. While the detrimental effect of either too much or too little water seems to have steadily diminished over time, the evolution of tolerance to extreme heat is highly nonlinear, growing with the adoption of hybrid corn in the 1940's, peaking around 1960, and then declining. Corn in Indiana is most sensitive to extreme temperatures at the end of our sample. Since climate change models predict an increase in extreme temperatures, the big question is whether the next breeding cycles can increase both average yields and heat tolerance simultaneously as in the period 1940-1960, or whether an continued increase in average yields can only be achieved at the expense of more sensitivity to extreme heat as in the period from 1960 onwards.

♣ Department of Agricultural and Resource Economics, North Carolina State University, Box 8109, Raleigh, NC. Email: michael_roberts@ncsu.edu.

♠ Department of Economics and School of International and Public Affairs, Columbia University, 420 West 118th Street, Room. 1308, MC 3323, New York, NY 10027. Email: wolfram.schlenker@columbia.edu.

1 Introduction

With evidence accumulating that greenhouse gas concentrations are warming the world's climate, there is growing interest in the potential impacts that may occur under different warming scenarios, and on how economies might adapt to changing climatic conditions. Agriculture is of particular interest due to the fact that climate is a direct natural input in the production process. Agriculture in developed nations, and particularly in the United States, has received considerable attention. Wealthier nations produce a disproportionate share of the world's agricultural commodities, at least partly due to their relatively more temperate climates. Accordingly, climate change impacts on agriculture in developed nations, and particularly the United States, the world's largest producer, have broad implications for food supply and prices worldwide.

In recent research, we have conducted detailed statistical analyses of the relationship between weather and crop yields of corn, soybeans, and cotton, three of the four largest U.S. crops, all of which are important for world commodity prices (Schlenker and Roberts 2008). Corn and soybeans are two are the world's four staple commodities that form the basis of most our calories.[1] The U.S. produces about 40 percent of world production in these two crops, making it, by far, the world's largest producer and exporter of these crops. Cotton is grown in the warmer Southern areas of the United States and might be better suited to warmer temperatures. We find that yields of all three crops grow roughly linearly in temperature up to a threshold above which yield growth declines sharply. The threshold varies by crop and is slightly higher for cotton, the warm weather crop. For all three crops, the slope of the decline above the optimum temperature for yield growth is significantly steeper than the incline below the optimum temperature. If we hold growing areas fixed and extrapolate from this relationship what yields might be as the climate warms, projected impacts are quite severe: production-weighted average yields decrease by 30-46% before the end of the century under the slowest (B1) warming scenario and decrease by 63-82% under the most rapid warming scenario (A1FI). These projected declines are driven by strongly negative yield growth when temperatures exceed 29-32 degrees Celsius combined with the sharp increase in the projected frequency of these extreme temperatures under projections by Hadley III climate model.

There are several reasons to believe these projected damages might overstate actual potential damages. As the climate warms, agricultural production will work to adapt to

[1] Rice and wheat are the other two.

those changes in a ways that mitigate losses and exploit potential gains as much as possible. The most difficult economic questions pertain to how large these adaptation possibilities may be. One obvious and inexpensive form of adaptation would be to simply change the locations where crops are grown. As climates change, so will geographical comparative advantages, so we should not expect crops to be grown the same locations as they are grown today. Ascertaining the potential impact of climate changes therefore calls for analysis of yield potentials of major crops across the globe, even in places where many crops are not grown today. Such analysis can be quite complex and requires some strong assumptions about the suitability of different soil types. For example, there is uncertainty about soil dynamics in the Tundra, a region that is currently too cold to farm but might become farmable under warming. Chapin et al. (1995) conduct experiments of soil changes in Alaska and find that the 3-year response in experimental plots are a bad predictor of 9-year changes in experimental plots. The authors emphasize the difficulty of predicting long-term changes using short-term heat waves.

In neither our earlier work nor in this paper do we attempt such a comprehensive analysis. Rather, by focusing on major crops in the U.S., a climatically diverse country that generates the world's largest agricultural output, we examine forms of potential adaptation observable in historical data. These historical adaptations (or lack thereof) may provide some insight into the scope and nature of potential adaptations that may be available as the climate changes.

Our earlier research found the same nonlinear relationship between yield growth and temperatures when the analysis is narrowed to consider only cooler northern U.S. states or only warmer southern U.S. states. We also found the same relationship if we examined only the early half of the sample (1950-1977), or only the latter half of the sample (1978-2006). These comparisons suggest that innovations since 1950, while increasing average yields approximately three-fold, have not increased relative heat tolerance. And since most regions of the U.S. currently have temperature distributions that are warmer than optimal, there has existed at least some incentive to breed or engineer more heat tolerance into plants. It would appear other kinds of innovations were less costly than development of improved heat tolerance.

Perhaps most indicative of limited adaptation possibilities (albeit still holding growing locations fixed), we found comparable and distinctively non-linear temperature-yield relationships when considering only geographic variations in *climate* paired with average yields in a county and when considering only time-series variations in *weather* paired with the average U.S. annual yield. The time-series relationship identifies a response function in which farmers have little scope to adjust their managerial decisions in the face of arguably random (from the perspective of farmers) weather shocks. In contrast, the cross-sectional relationship compares areas with different expected weather outcomes, i.e., different climates, and thus accounts for a fully adaptive response by farmers, much as might be expected with climate change. The fact that these relationships are similar suggests that, at least historically from 1950 to 2006, there has been little scope for adaptation conditional on the locations where these key crops were grown. This finding is consistent with some earlier work using the hedonic approach, which considers cross-sectional variations in climate to land values (Schlenker et al. 2006).[2]

Where this earlier work found little evidence of adaptation to warmer temperatures between 1950 and 2006, in this paper we look to the earlier and potentially more interesting period between 1901 and 1950. Our focus on this period is motivated in large part by Sutch's (2008) research. Sutch argues that the adoption of hybrid corn, one of history's most remarkable and well documented technological revolutions, was precipitated in part by the extreme weather events of the 1930s. In particular, he argues that hybrid corn demonstrated particularly high yields relative to open-pollinated (non-hybrid) corn during 1934 and 1936, which (by our own crop-related measures) remain the most extreme on record. Thus, it could be that our earlier analysis did not look back far enough to the timing of the key innovation leading to the green revolution.

Specifically, in this paper we examine a panel of corn yields from 1901-2005, a time period

[2]Mendelsohn et al. (1994) first introduced the Ricardian method to measure the effects of climate change on agriculture by estimating a cross-sectional relationship between county-level farmland values and climatic variables in the United States. The predicted impact of changing climatic variables depend largely on the set of weights. Under the cropland weights (fraction of a county that is cropland) the predicted impacts are severely negative, and under the crop-revenue weights (the value of agricultural production sold) the effects are beneficial. The reason why the results diverged under various weights is access to highly subsidized irrigation water rights in the Western United States. These subsidized water rights capitalize into farmland values (Schlenker et al. 2007). Since access to subsidized water rights is correlated with temperature, an increase in temperature implicitly assumes an increase in subsidies, which should not be counted as a societal benefit. The crop-revenue weights aggravate the problem because highly irrigated counties in the Western United States account for a large share of overall revenues, yet the fraction of the county that is cropland (cropland weights) is small. Schlenker et al. (2005) show that if the analysis is limited to rainfed agriculture, the results converge and become unambiguously negative under both sets of weights. (Deschênes and Greenstone 2007) recently proposed to use year-to-year variation in weather to estimate the relationship between profits or yields and weather. They find that agricultural profits and yields are independent of weather. However, their weather data set contains many irregularities and their profit measure, which is the difference between sales in a given year minus expenditures, does not account for storage behavior that smooth profits between periods. Once the data errors are corrected, projected climate change effects on yields are again unambiguously negative (Fisher et al. 2009).

that includes a full 35 years before the beginning of the green revolution as well as some 70 years after the first adoption of hybrid corn on a significant scale. Our analysis focuses mainly on the state of Indiana, which sits in the middle of the so-called "Corn Belt" and is the nation's third largest corn growing state. Our focus on Indiana is mainly due to data availability: it turns out that Indiana has the most comprehensive record of detailed daily weather data that are electronically available. Such detailed weather data is necessary in order to apply our methods of analysis, which require suitably accurate estimates of the entire distribution of temperature outcomes in order to account for variations in temperatures, both within and across all days of each growing season. This detail facilitates correct identification of nonlinear temperature effects, which can be diluted from measurement error, or if temperatures are averaged over time or space. The key focus of our analysis is to examine how heat tolerance and drought tolerance has changed over time, with some particular focus on the time period following the great heat waves of 1934 and 1936 and subsequent widespread adoption of hybrid corn.

2 Corn Yields and Weather in Indiana

Figure 1 presents the evolution of average corn yields in Indiana over the 20th century. These yield data are publicly available from the U.S. Department of Agriculture's National Agricultural Statistical Service (USDA-NASS). All of our data sources are described in further detail in the data appendix. The graph shows the average yield in the state for all years between 1901 and 2005 as black diamonds. For years after 1928, when county-level data becomes available, a box plot shows the range and inter-quartile range of yields across counties in Indiana.

Before 1940 there was no discernible trend in yields. This is true even if one were to extend the time series back many decades before 1901, the earliest year shown on in the figure. Around 1940 yields started a sharp upward linear trend that appears ongoing even today. Typical yields in Indiana were between 30 and 40 bushels per acre before 1940, yet today, a typical Indiana farmer can expect 150 to 160 bushels per acre. Yield variance increased along with typical yields, so we will model the natural log of yield per acre.

As discussed in greater length by Sutch (2008), the beginning of the upward trend in yields began around the time when many key events occurred simultaneously. The Great Depression in the 1930s was followed by the onset of World War II in 1938, which drove commodity prices up sharply from their lows in the early Depression years. At least equally important though was the early adoption of hybrid corn, starting in Iowa and quickly expanding to Illinois, Indiana, and beyond. The superior yields of hybrid corn was discovered in 1918 but it was not until later, perhaps after 1936, that seed production became commercially viable and high-yielding enough for farmers to adopt.

Also, in the decade before 1940, the Midwest, including Indiana, experienced both the hottest and driest temperatures on record for the growing-season months between March through August, shown in the panels of Figure 2. The left graphs shows yearly weather shocks in extreme heat (degree days 29°C, further described below) over the growing season. The right graph shows precipitation deviations from average climatic conditions. The decade of poor weather in the 1930s was most accentuated in the two drought years of 1934 and 1936, which brought about the great Dust Bowl, an event of massive wind erosion in states west and south of Indiana. In those years average yields in Indiana were just 27.6 and 25.6 bushels per acre, two of the three worst yields on record for the state during the 20th century. Indiana still fared much better than states west and south of Indiana. Iowa harvested just 60 percent of its planted acreage in 1934–an all time low–and Dust Bowl states of Nebraska and Kansas were nearly total losses in these years.

Construction of the weather variables presented in Figure 2 is further detailed in the appendix. We construct these data from daily individual weather stations in Indiana. Geographical interpolation is achieved by linking it with the PRISM weather data sets, which gives monthly observations on a 2.5x2.5 mile grid for the entire united States. Indiana is the only state in the U.S. for which the National Climatic Data Center of the National Oceanic and Atmospheric Administration reports having more than three weather stations in the early part of the century. The availability of good, fine-sale weather data is essential for identifying non-linear weather effects since these effects can be diluted with measurement error or if values are averaged over time and space. The geographical locations of weather stations in Indiana that we use to construct our data set for each 25-year period are shown in Figure 3.

The challenge for a regression model that relates yields to weather outcomes is in mapping an entire season of temperature and precipitation outcomes to a single yield response. We achieve this by assuming temperature effects on yields are cumulative over time and that yield is proportional to total exposure. This implies temperature effects are additively substitutable over time, i.e., we sum the daily outcomes over all days of the growing season. Earlier work has shown that there are three weather variables that give the best out-of-sample predictions of corn yields: (i) total precipitation p_{it} in county i in year t; (ii) degree days above 29°C (dd_{it}^H), which captures the harmful effects of high temperatures; and (iii) degree days between 10°C and 29°C degrees (dd_{it}^M), which measures the beneficial effects of moderate temperatures (Schlenker and Roberts 2008). Each measure is simply a truncated integral over the temperature distribution within a day, as outlined below.

Degree days above 29°C (high temperature measure) are defined as:

$$dd_{it}^H = \sum_{j=\text{March 1st}}^{\text{August 31st}} \int_{T=29}^{\infty} (T-29) h_{itj}(T) dT$$

where T is temperature (in degrees Celsius) and $h_{itj}(T)$ is the estimated density of time at each degree during day j in year t in county i. Since the measure is sensitive to geographic variation in temperatures, as wells as variations within and across all days in the growing season, we spend considerable care in estimating $h_{itj}(T)$. Further details are given in the data appendix.

The second temperature measure is degree days between 10°C and 29°C (moderate temperature measure) are defined as:

$$dd_{it}^M = \sum_{j=\text{March 1st}}^{\text{August 31st}} \int_{T=10}^{29} (T-29) h_{itj}(T) dT.$$

3 Regression Model

Where earlier research took care in determining the precise nature of the nonlinear relationship between temperature and yields, in this paper we take as given the two temperature measures described above that are informed by this earlier work. Here our focus is to explore how the relationship between yields and weather has changed over the 105 years from 1901 to 2005. We use a flexible restricted cubic spline model that allows temperature and non-linear precipitation associations to change smoothly over time. Specifically, the regression model is:

$$y_{it} = \beta_0 dd_{it}^M + \beta_1 dd_{it}^H + f_p(p_{it}) + f_t(t) + f_M(t) * dd_{it}^M + f_H(t) dd_{it}^H + f_{t2}(t) f_{p2}(p_{it}) + c_i + \epsilon_{it}$$

where y_{it} denotes the natural log of yield in county i and year t, dd_{it}^M and dd_{it}^H are the degree day measures described above, p_{it} is precipitation, and the function $f_x(\cdot)$ are splines of time or precipitation. Each of the functions is approximated using 5 knots, located at the 0.05, 0.275, 0.5, 0.725 and 0.95 quantiles of the empirical distribution. For time trend knot locations are 1932, 1949, 1967, 1984, and 2001; for precipitation they are 38.9, 51.6, 58, 65.0, and 79.0. The early knot in the time trend is due to the fact that we have only state-level observations prior to 1929, and thus fewer data points per year than after 1929 when we have county-level observations. We also include separate intercepts for each county (i.e., fixed effects, denoted c_i) to account for unobserved time-invariant heterogeneity, like soil quality.

Estimation restricted cubic spline models is easily done using ordinary least squares. Since the errors within each year are likely correlated in space, we adjust our standard errors to account for this (clustering the errors by year) and possible heteroscedasticity using the Huber-White method.

4 Results

The main regression results are shown in figure 4. This figure shows the effects of each of the four variables, time, precipitation, dd^M and dd^H while holding all other three variables fixed at approximately their median value. These results are characteristically similar to what we found in our earlier work that focused on the period from 1950-2005: there is a sharp upward trend in yields over time as shown in the top left panel. Yields have an inverted-U shape with rainfall as shown in the top right panel. Yields increase gradually with temperate degree days between 10°C and 29°C as shown in the bottom left panel. Finally, yields decline sharply with extreme heat, measured as degree days above 29°C, as shown in the bottom right panel.

These median-value predictions, however, do not show how these relationships have changed over time. We explore how these relationships change over time in figure 5 for precipitation, figure 6 for extreme heat dd^H, and figure 8 for moderate temperatures dd^M. Each of these figures plots the relationship of the three weather variables at three points in time. In the appendix we show plots for 15 years covering the entire span of dd^H and precipitation.

The effects of both precipitation and extreme heat have shifted markedly over time. Figure 5 shows that the influence of precipitation has nearly vanished over time. We believe that two explantation are most likely responsible for the fact that yields are no longer directly linked to rainfall during the growing season. First, a lack of precipitation in the growing season might be counter-balanced with irrigation. Continued mechanization of agriculture has lead to the gradual expansion of pivot irrigation systems that can provide supplementary water during especially dry periods. While only a minority of corn fields in Indiana have pivot irrigation systems, the ones that do are probably more prone to dryness or have sandier soils. Second, seed companies may have bred increased drought tolerance into corn plant varieties.

While climate models vary considerably in their predictions for precipitation changes, with some forecasting increases and others decreases, evidence from weather and yields in Indiana suggest this may be of little economic consideration. However, it should be noted that the marginal effect of moderate temperatures (dd^M) was negative during the years of the worst drought years in the 1930s. This suggests that under significant multi-year water shortages the effect of temperatures was more harmful. Since we have not experienced droughts of such severity in the recent past, it is less clear what the effects of such severe droughts would be in more recent years.

Heat tolerance seems to have increased in Figure 6 up until 1960 followed by a sharp decline after 1960. Figure 7 shows the *marginal* effect of extreme heat, i.e., the slope of the regression line in Figure 6 over all years in our sample. The effect of an additional degree day above 29°C is lowest around 1960 and most damaging in recent years when corn varieties were optimized for maximum average yields.[3] The magnitude of the negative coefficient on dd^H is nearly three times as large in 2000 as it is in 1960, and about twice as large in 1901 as compared to 1960.

Estimated slopes in the early years of figure 7 should be interpreted with some caution because there are much fewer data points before to 1929 than afterward, since only state level data are available in the earlier period and county level data afterward. Our spline model places more emphasis where there are more data and linearizes the model in the tails of the data. Closer inspection of the data do suggest that much of the increase in heat tolerance actually took place between 1940 and 1960, rather than being a steady smooth trend up from 1901.[4] This interpretation would be consistent with the relatively stable farming technologies between 1901 and 1936 and rapid technological after 1940. This would also be consistent with Sutch's historical account of the adoption of hybrid corn.

The most interesting and relevant finding that speaks to implications for climate change is the sharp decline in tolerance to extreme heat since 1960. This finding is a powerful counterpoint the apparent increase in drought tolerance. Under the latest climate change models, a sharp rise in maximum temperatures is predicted to significantly increase the occurrence

[3] The appendix lists the evolution of the precipitation and temperature effects at 15 points in Figures A3 and A4 instead of three points in time to make the transition of the relationships more visible.

[4] We intend to present evidence of this in a revised draft.

of temperatures above 29°C. Since degree days above 29°C are a truncated temperature variable, modest shifts in the temperature distribution can have a large relative influence on this temperature measure. For example, a 1°C warming from 29.5°C to 30.5°C triples degree days above 29°C. The historic average number of degree days above 29°C is 25 in Indiana. Under the Hadley II model (IS92a scenario) the number is predicted to increase by 19 at the end of this century. Under the much warmer Hadley III model, degree days above 29°C are projected to increase by 104 under the slow-warming B1 scenario. Thus, even under the slowest-warming scenario, typical weather outcomes in the latter part of this century will be far worse than the worst drought years in the historical record, 1934 and 1936 (refer to figure 1). Under the fastest-warming A1FI scenario, degree days above 29°C are projected to increase by 330, making the measure in a typical year about 3.5 times worse than the worst year on record.

Finally, the relationship between the moderate temperature measure (dd^M, degree days 10-29°C) and log yields has grown more positive with time. This pattern is not statistically significant and thus may be spurious. A summary of significance tests is reported in table 1. For all factors besides dd^M, and all their non-linear interactions with time, have strong statistical significance.

5 Conclusions

This paper has extended earlier research on the the link between weather and yields by examining how various weather variables are associated with corn yields in Indiana for the period 1901-2005. We use restricted cubic spline regressions to let the effect of precipitation, moderate heat, and extreme heat evolve over time in a flexible way. Results for each variable while holding all other variables constant at their median outcomes is comparable to results we obtained for a model using county-level corn yields for all counties east of the 100 degree meridian in the years 1950-2005.

The median association, however, obscures significant evolution of precipitation and temperature effects over time. The effect of precipitation during the growing season is diminished. We hypothesize that attenuation of precipitation effects stems from increased use of supplemental irrigation and development of more drought tolerant crops. A countervailing evolution has made corn in Indiana farm more sensitive to extreme heat. The nonlinear evolution of heat tolerance over time increases until about 1960 and then decreases sharply, with the most damaging marginal effect of temperatures above 29°C occurring most recently.

This might be due to the fact that maximizing corn plants for average yields also makes them more sensitive to suboptimal growing conditions.

Implications for climate change impacts are hence mixed: on the one hand, the sensitivity to extreme heat is highest at the end of the sample and the one feature all climate models agree on is that these extreme heat events are likely to increase, even though the size of the increase varies tremendously between model and emission scenarios. On the other hand, there was a period between 1940-1960 when both heat tolerance and average yields increased at the same time. The question is whether recent increases in yields could only be achieved by making plants less heat resistent, or whether future breeding cycles can increase both heat tolerance and average yields at the same time.

References

Chapin, F. Stuart, Gaius R. Shaver, Anne E. Giblin, Knute J. Nadelhoffer, and James A. Laundre, "Responses of Arctic Tundra to Experimental and Observed Changes in Climate," *Ecology*, April 1995, *76* (3), 694–711.

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

Fisher, Anthony C., W. Michael Hanemann, Michael J. Roberts, and Wolfram Schlenker, "Agriculture and Climate Change Reconsidered," *Working Paper*, May 2009.

Mendelsohn, Robert, William D. Nordhaus, and Daigee Shaw, "The Impact of Global Warming on Agriculture: A Ricardian Analysis," *American Economic Review*, September 1994, *84* (4), 753–771.

Schlenker, Wolfram and Michael J. Roberts, "Estimating the Impact of Climate Change on Crop Yields: The Importance of Nonlinear Temperature Effects," *NBER Working Paper 13799*, February 2008.

——, W. Michael Hanemann, and Anthony C. Fisher, "Will U.S. Agriculture Really Benefit from Global Warming? Accounting for Irrigation in the Hedonic Approach," *American Economic Review*, March 2005, *95* (1), 395–406.

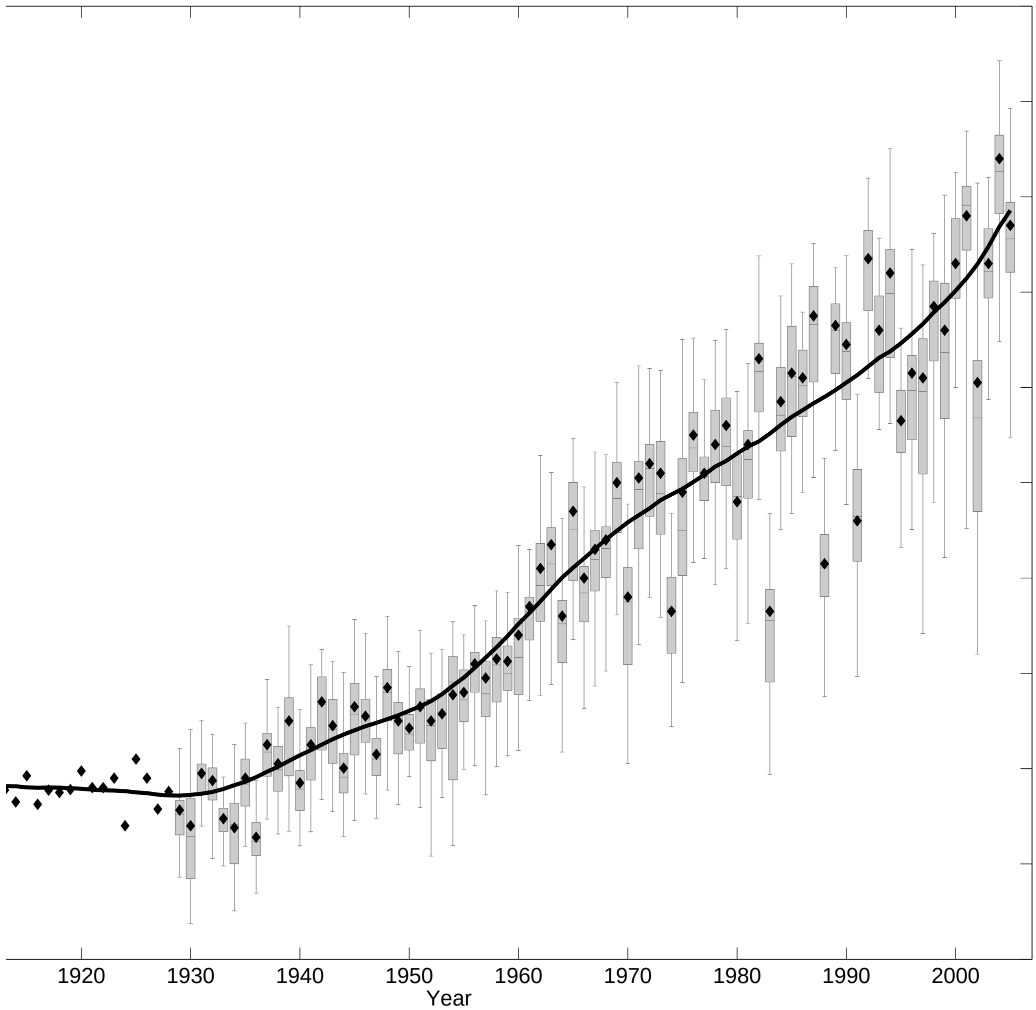
——, ——, and ——, "The Impact of Global Warming on U.S. Agriculture: An Econometric Analysis of Optimal Growing Conditions," *Review of Economics and Statistics*, February 2006, *88* (1), 113–125.

——, ——, and ——, "Water Availability, Degree Days, and the Potential Impact of Climate Change on Irrigated Agriculture in California," *Climatic Change*, March 2007, *81* (1), 19–38.

Snyder, R. L., "Hand Calculating Degree Days," *Agricultural and Forest Meterology*, October 1985, *35* (1-4), 353–358.

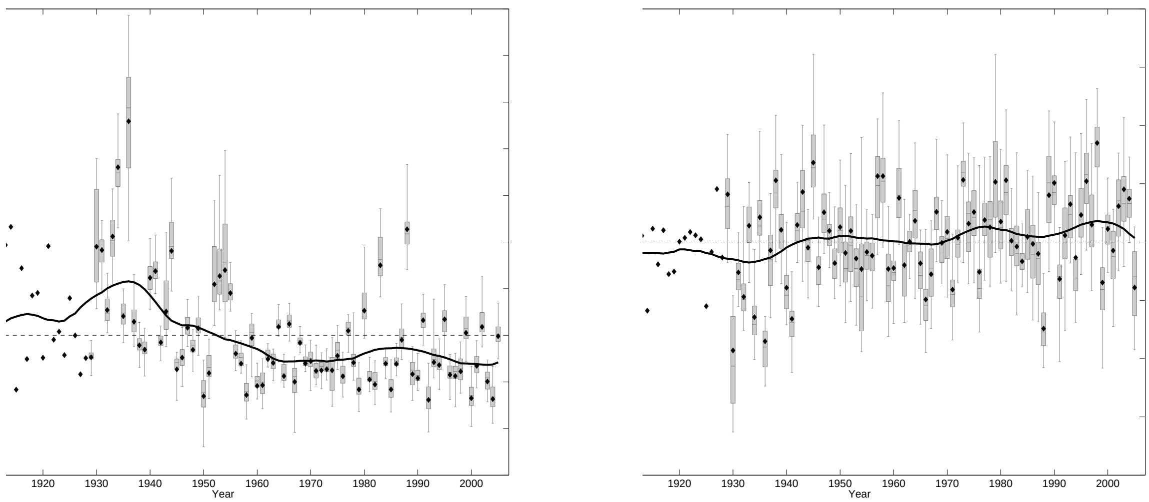
Sutch, Richard C., "Henry Agard Wallace, the Iowa Corn Yield Tests, and the Adoption of Hybrid Corn," *NBER Working Paper 14141*, June 2008.

Figure 1: Average Yields in Indiana 1901-2008



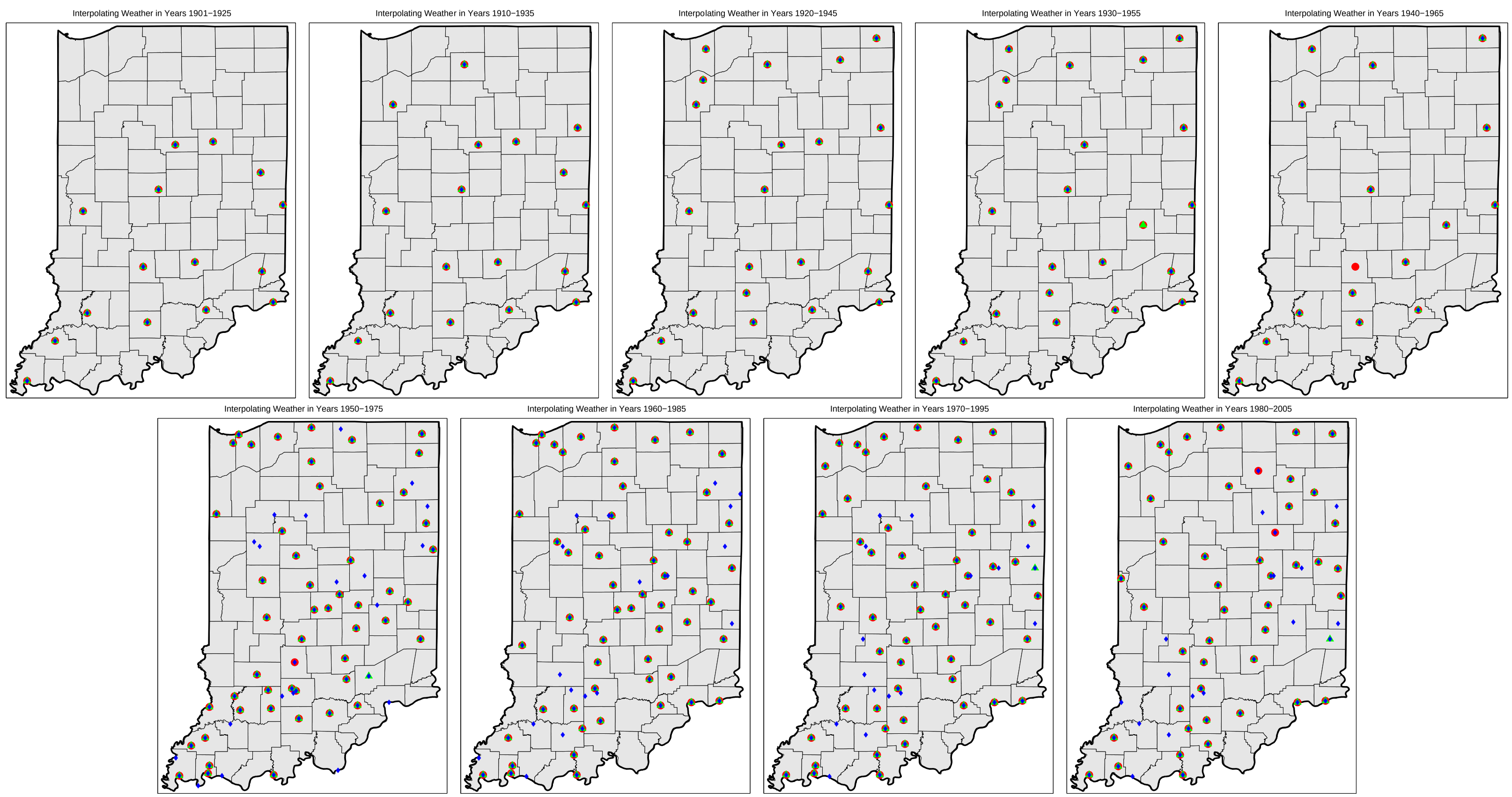
Notes: Graph shows history of corn yields in Indiana. State level averages are shown as diamonds. The range of yields among Indiana's counties is shown as boxplots: The box give the 25%-75% quartile range, the median is shown as a solid line, and whiskers extend to the minimum and maximum. A locally weighted regression (bandwidth of 10 years) is shown as black line.

Figure 2: Weather Shocks in Indiana 1901-2005



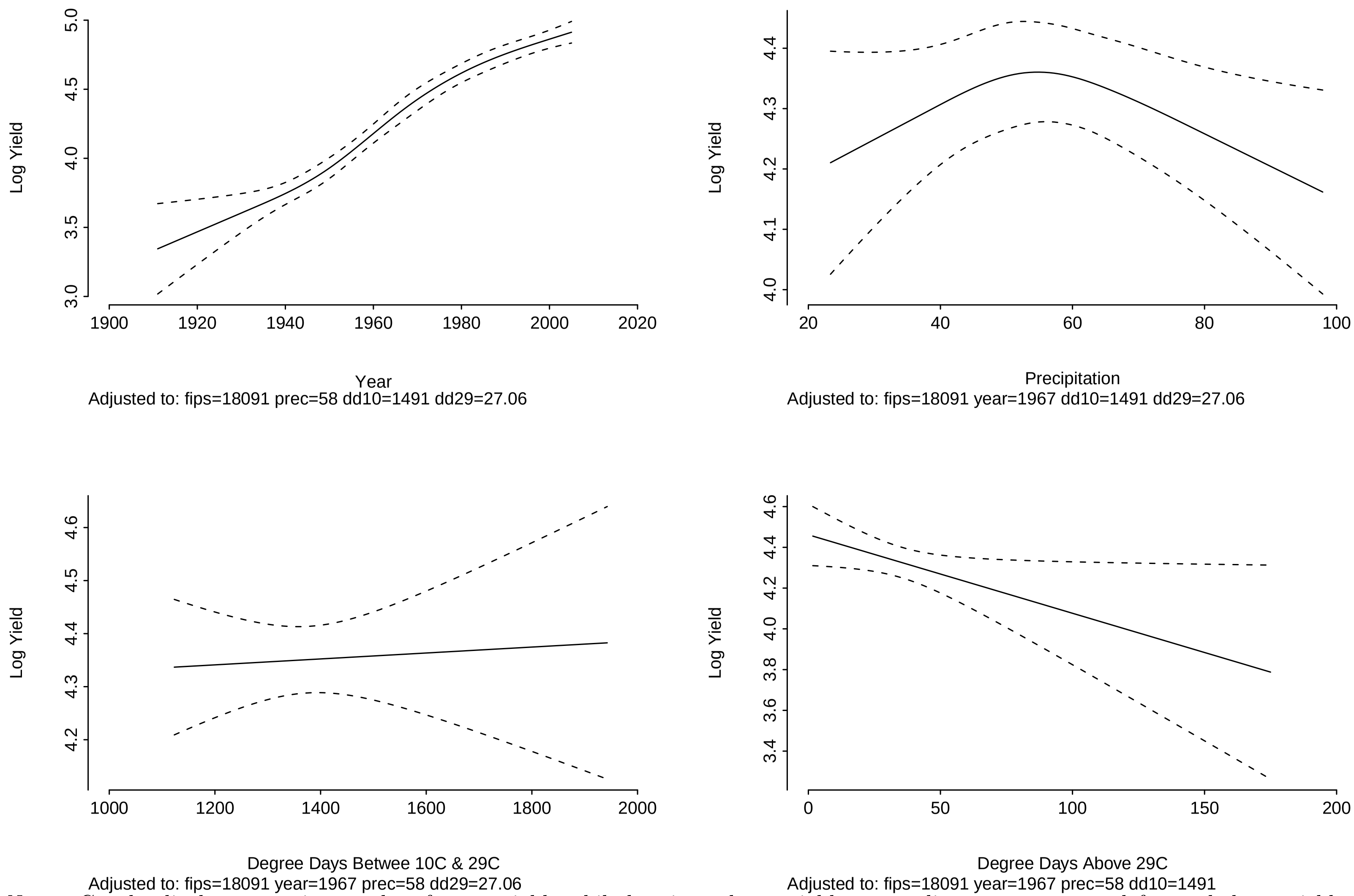
Notes: Graphs shows weather shocks, i.e., the difference to averages in a location. The left graph shows results for degree days above 29°C and the right graph for total precipitation during the growing season March-August. State level averages are shown as diamonds. The range of weather shocks among Indiana's counties is shown as boxplots: The box give the 25%-75% quartile range, the median is shown as a solid line, and whiskers extend to the minimum and maximum. A locally weighted regression (bandwidth of 10 years) is shown as black line.

Figure 3: Weather Station in Indiana By Time Period.



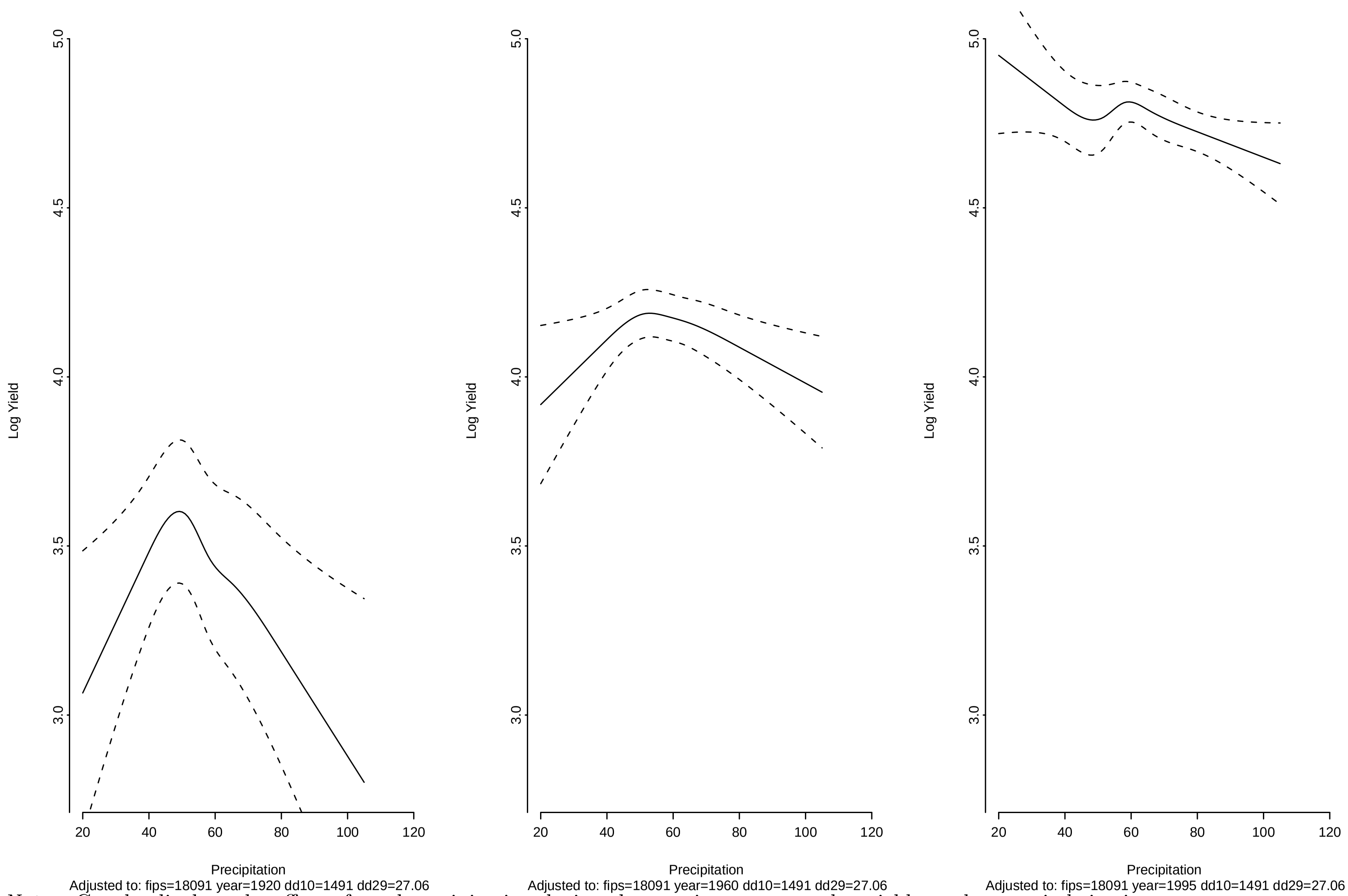
Notes: Weather stations used in interpolation are displayed as red dots for minimum temperature, green triangles for maximum temperature and blue diamonds for precipitation.

Figure 4: Regression Results at Median Values



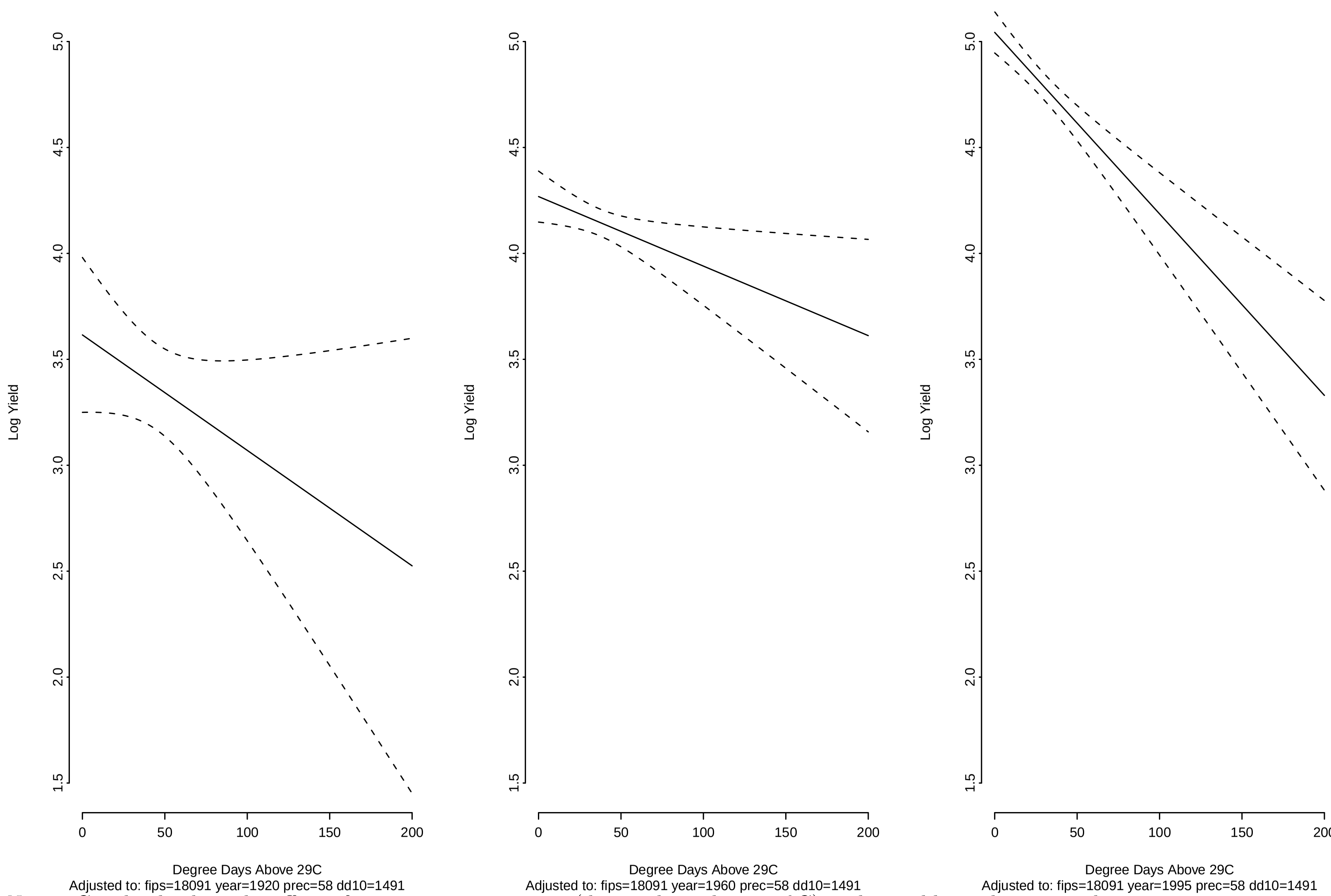
Notes: Graphs display regression results of one variable while keeping other variables at median outcomes: top left panel shows yield trend over time, top right panel shows effect of precipitation, bottom left panel shows effect of degree days 10-29°C and bottom right panel shows effect of degree days above 29°C.

Figure 5: The Evolution of the Impact of Precipitation on Corn Yields



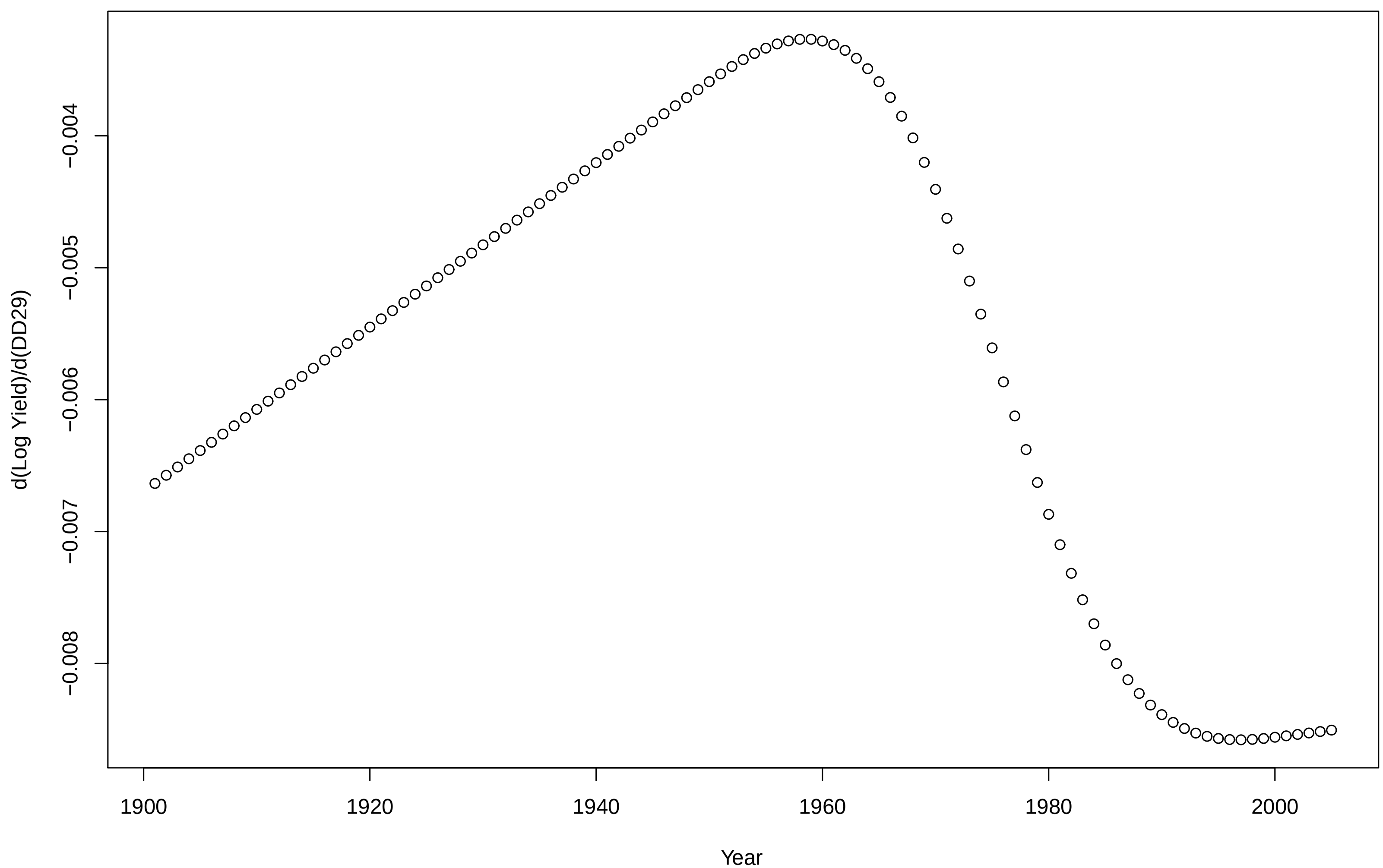
Notes: Graphs displays the effect of total precipitation during the growing season on log yields at three periods in time.

Figure 6: The Evolution of the Impact of Extreme Heat on Corn Yields



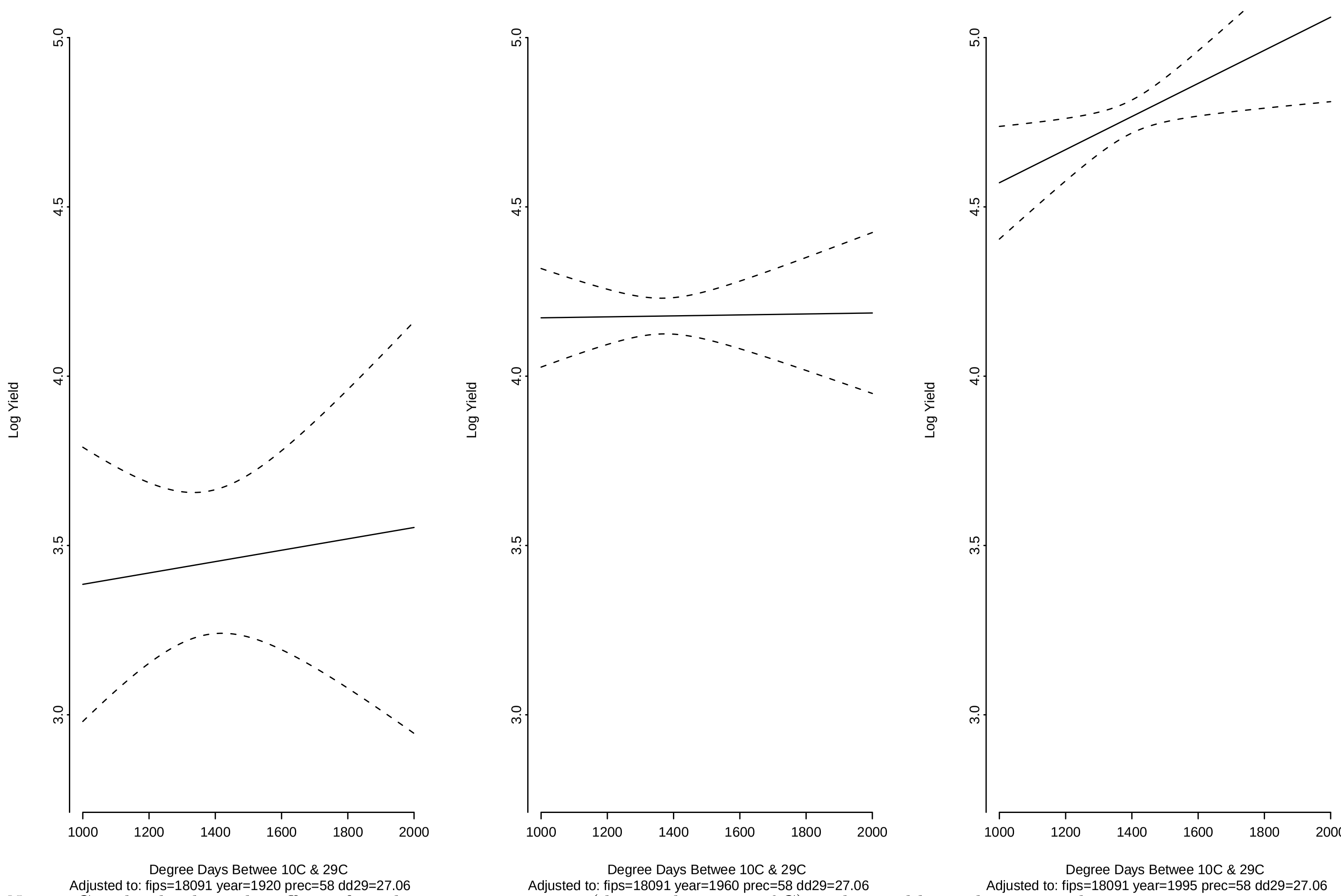
Notes: Graphs displays the effect of extreme temperatures (degree days above 29°C) on log yields at three periods in time.

Figure 7: The Evolution of the Marginal Impact of Extreme Heat on Corn Yields



Notes: Graphs displays the marginal effect of extreme temperatures (degree days above 29°C) on log yields, i.e., the slope of the regression lines in Figure 6.

Figure 8: The Evolution of the Impact of Moderate Heat on Corn Yields



Notes: Graphs displays the effect of moderate temperatures (degree days 10-29°C) on log yields at three periods in time.

Table 1: Analysis of Variance for Log Yield

	d.f.	*PartialSS*	*MS*	*F*	*P*
Degree Days > 29C					
Factor+Interaction with Time	5	1.440	0.288	13.62	<0.0001
Interaction with Time	4	0.22	0.055	2.61	0.0339
Degree Days 10-29C					
Factor+Interaction	5	0.18	0.036	1.71	0.1297
Interaction with Time	4	0.14	0.034	1.62	0.1662
Precipitation					
Factor+Interactions	11	1.86	0.169	7.97	<0.0001
Interaction with Time	7	0.46	0.065	3.09	0.0030
All Nonlinear Components	6	0.57	0.095	4.48	0.0002
Time Trend					
Factor+All interactions	19	52.95	2.787	131.71	<0.0001
All Interactions	15	0.86	0.058	2.72	0.0004
All Nonlinear Components	12	1.17	0.098	4.62	<0.0001
TOTAL NONLINEAR	18	1.99	0.111	5.24	<0.0001
TOTAL	25	132.53	5.301	250.56	<0.0001
ERROR	6950	147.05	0.021		

Notes: The table reports F tests for the joint significance of key explanatory variables and their interactions with other variables. The **Precipitation** factor, for example, is estimated as a restricted cubic spline plus a separate restricted cubic spline of precipitation interacted with a restricted cubic spline of time. This allows precipitation effects to shift over time in a nonlinear way, as illustrated in figure 5. Splines are restricted in the sense that they are forced to be linear in the tails of the data. *All Interactions* refers to model components where the factor is interacted with other variables (usually a spline of time). *All Nonlinear Components* tests all higher-order covariates that capture curvature in either the variable itself or in the interaction of the variable with another variable. These tests shows we fail to reject insignificance of all the factors, nonlinearity, and nonlinear interactions, except for the moderate temperature variable, Degree Days 10-29C. All splines are estimated with five knots, located at the 0.05, 0.275, 0.5, 0.725 and 0.95 quantiles of the explanatory variable's empirical distribution. For time trend the knot locations are 1932, 1949, 1967, 1984, and 2001; for precipitation they are 38.9, 51.6, 58, 65.0, and 79.0. The early knot in the time trend is due there being only state-level data before 1929. The splines are estimated using ordinary least squares using four degrees of freedom for each one, given segments outside the lower and upper knots are forced to be linear. Interactions of two splines (in the case of time and precipitation) use 7 degrees of freedom. All test statistics use the Huber-White method to adjust the variance-covariance matrix for heteroscedasticity and correlated (clustered) errors in each year. The model also includes intercepts for each county (i.e., fixed effects). Estimation was done using the R package *Design* written by Frank Harrell.

Appendix

A1 Data Appendix

This appendix outlines in further detail how we construct our data set.

A1.1 Yield Data

Yield data was obtained from the National Agricultural Statistics Service (accessed March 2009). Yearly state-level yields in Indiana are available from 1866 onwards.[5] County-level yields in Indiana are available starting in 1929.[6] We follow the definition of the Department of Agriculture and calculate yields as the ratio of total production divided by area harvested.

The traditional definition of yields might overstate actual yields if some fields are not harvested. In a sensitivity check we define yields as total production divided by all acres planted. Unfortunately, area planted is only available from 1926 onwards for state totals and from 1972 for individual counties and hence significantly reduces our sample period. The left panel of Figure A1 displays the fraction of the planted area that was harvested in Indiana over time. While there is an upward trend, especially during the 30s, the right panel shows that the year-to-year variation in yields is similar for each definition of yields.

A1.2 Weather Data

Degree days were constructed from daily weather data. We obtained daily observations from the National Climatic Data Center Cooperative station network.[7] The data include *daily* minimum and maximum temperature as well as precipitation. While the NCDC data has great *temporal* coverage, we combine it with the PRISM weather data set that provides better *spatial* coverage.[8] The latter gives monthly minimum and maximum temperatures on a 2.5x2.5 mile grid for the United States from 1895 onward.

To construct a consistent set of weather data, we followed the following procedure for each 25-year period starting in 1901, 1910, 1920, 1930, ..., 1980.

(i) For each of our three weather variables (minimum and maximum temperature as well as precipitation) we determine the set of stations with a consistent record, which we

[5] http://www.nass.usda.gov/QuickStats/Create_Federal_All.jsp
[6] http://www.nass.usda.gov/QuickStats/Create_County_All.jsp
[7] http://ols.nndc.noaa.gov/plolstore/plsql/olstore.prodspecific?prodnum=C00447-CDR-S0001
[8] http://www.prism.oregonstate.edu/

chose to be stations that moved at most by 2.5 miles during the time period and had at most three missing values in at least 90 percent of the months.

(ii) We fill the missing observations at stations with consistent records obtained in step (i) by regressing daily values at each station on daily values at the seven closest stations including half-month fixed effects. We use a linear regression for minimum and maximum temperature and a Tobit regression for precipitation, which has several observations at the truncation value of zero. Intuitively, the regression estimates are used to fill the missing values with a weighted average of surrounding stations with non-missing observations to give us a complete weather record at the stations with consistent weather records.

(iii) We calculated monthly averages for the stations with consistent records in step (i)

(iv) We regress the monthly values at each PRISM grid on the monthly averages at the seven closest weather stations from step (iii) including month fixed effects, again using a linear regression for minimum and maximum temperature and a Tobit model for precipitation. The R-squares are generally in excess of 0.999, suggesting that the PRISM data set is a weighted average of individual stations and we uncovered the weights.

(v) We apply the regression results from step (iv) to the daily weather station data from step (ii) to derive daily weather measures at each 2.5x2.5 mile PRISM grid cell.

(vi) We fit a sinusoidal curve between the minimum and maximum temperature of each day to calculate degree days accounting for the within day distribution of temperatures (Snyder 1985). We evaluate degree days for each bound between -5°C and +50°C using 1° steps at each 2.5x2.5 mile PRISM grid.

Once we have the daily observations on the PRISM grid, we aggregate them spatially

(vii) We obtained the fraction of each PRISM grid cell that is cropland from a one-time LandSat satellite scan in 1992. County-level weather variables are the cropland-weighted average of all PRISM grid cells in a county.

(viii) State-level weather data are the weighted average of all county-level measures in step (vii), were the weights are the amount of harvested corn area reported in the yield data. Since harvested corn area is not reported on a county-level before 1929, we use the average harvested corn area in each county in the years 1929-2005 as weights for years prior to 1929.

Finally, we aggregate the data temporally

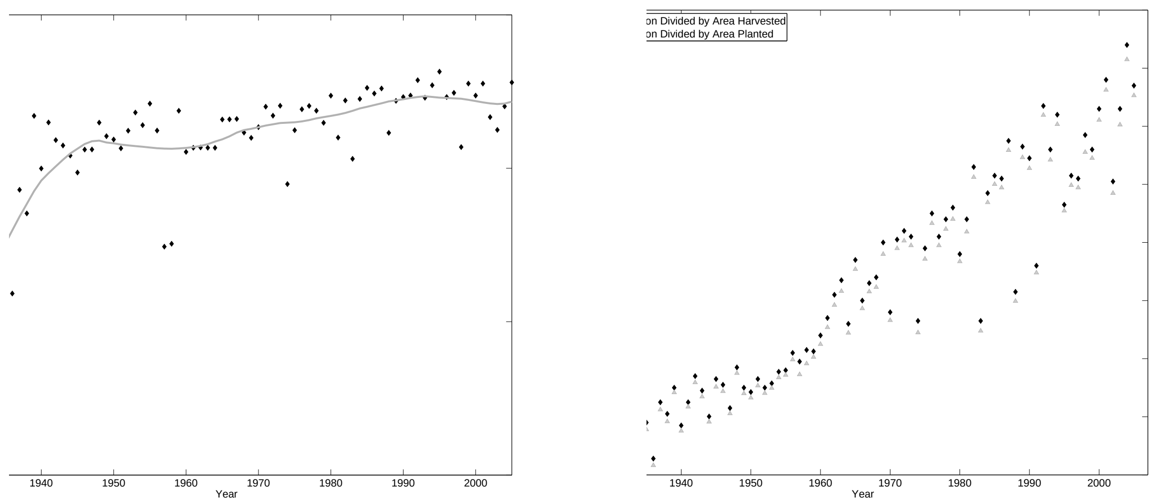
(ix) We define the growing season as the months March through August and add degree days as well as precipitation for all days in these months. Since total precipitation over the growing season is insensitivity to the within-day and between-day distribution, we use the monthly totals in the PRISM data set. For possibly daily interactions between precipitation and temperature we use the interpolated daily precipitation data.

Since it was impossible to get a sufficiently large set of weather stations which had consistent nonmissing records for the entire sample period 1901-2005, we instead derived the measure for 25-year intervals, starting in 1901, 1910, 1920, up to 1980. The results of interpolation series for extreme heat in the state of Indiana (degree days above 29°C) are displayed in colors in Figure A2. They appear to overlap tightly. One might still wonder whether the state results hide the fact that there are substantial errors in the county level data that get averaged out. To examine this further, we take the difference of all overlapping series in the county data. The mean absolute difference is 2.2 degree days above 29°C and the root mean squared prediction error is 3.1 degree days above 29°C, suggesting that the overlapping fit is reasonably close. Our weather data uses the average of all overlapping series.

A2 Sensitivity Checks

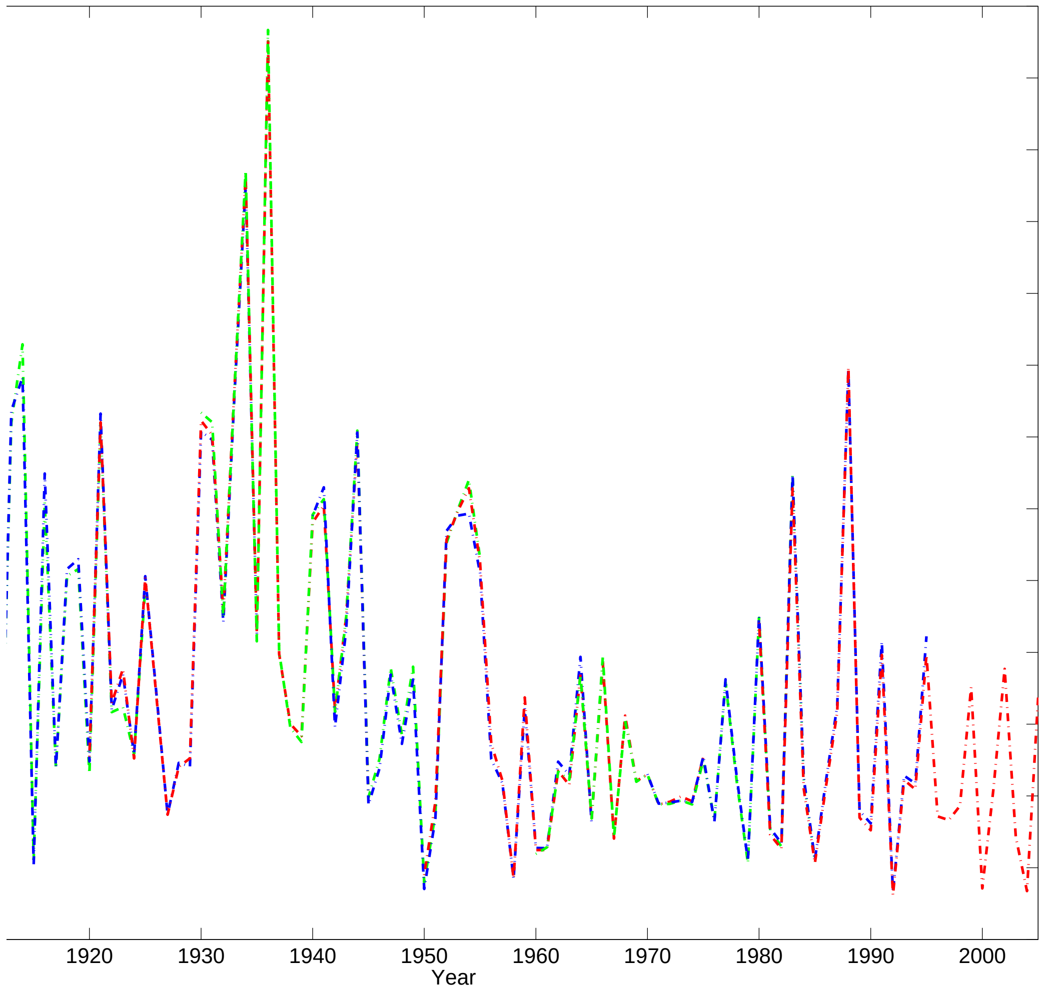
Figure 5 and Figure 6 in the main paper examined how the effect of damaging extreme heat evolved over time at three points in time. To further break down this evolution, we replicate these figures using 15 points in time in Figure A3 and Figure A4, respectively.

Figure A1: Fraction of Corn Area Planted that is Harvested



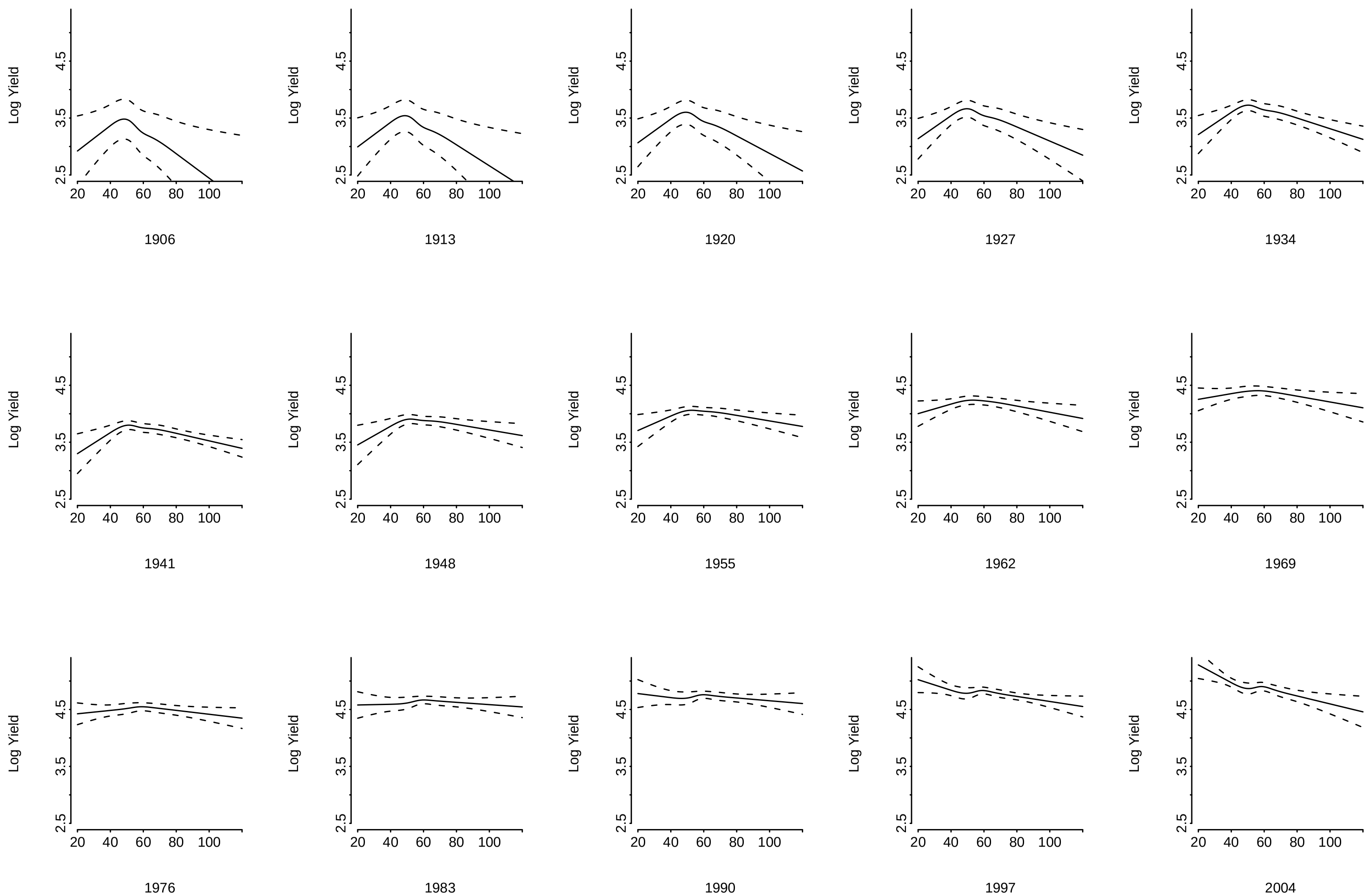
Notes: Left panel shows the ratio of the corn area harvested to the area planted in Indiana 1926-2005 as black diamonds as well as a locally weighted regression with a bandwidth of one decade as grey solid line. The right panel shows yields under the two different definitions. Production divided by area harvested is show as black diamonds, and production divided by area planted as grey triangle.

Figure A2: Interpolation Accuracy (1901-2005)



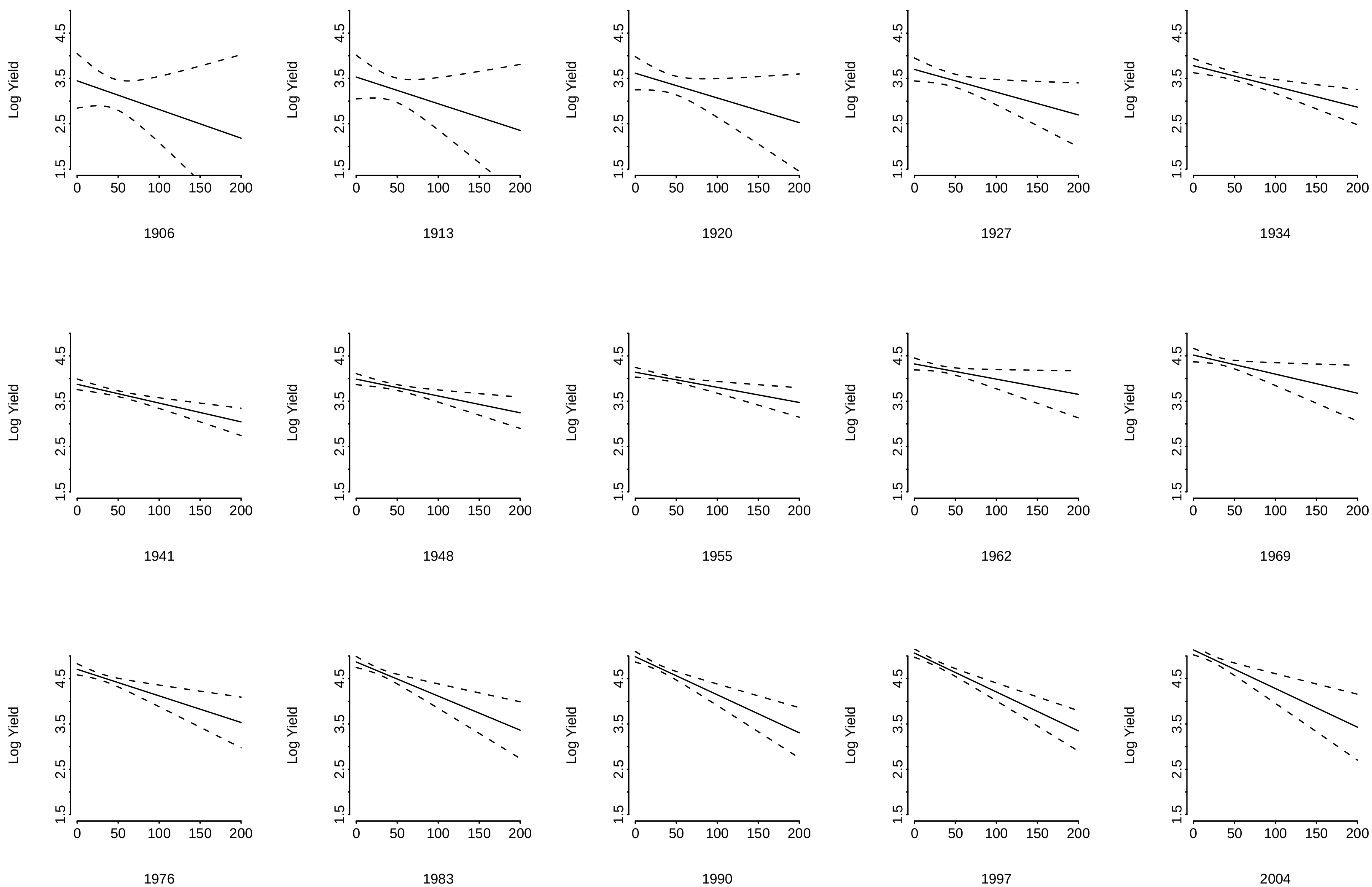
Notes: Graph shows degree days above 29°C in Indiana for each overlapping 25-year interpolation period starting in 1901-1925, 1910-1935, ..., until 1980-2005.

Figure A3: The Evolution of the Impact of Precipitation on Corn Yields



Notes: Graphs displays the effect of total precipitation during the growing season on log yields at 15 periods in time.

Figure A4: The Evolution of the Impact of Extreme Heat on Corn Yields



Notes: Graphs displays the effect of extreme temperatures (degree days above 29°C) on log yields at 15 periods in time.