WORKING PAPER
2012-02

REPA

Resource Economics
& Policy Analysis
Research Group

Department of Economics
University of Victoria

Climate Impacts on Chinese Corn Yields: A Fractional Polynomial Regression Model

Baojing Sun and G. Cornelis van Kooten

July 2012

Copyright 2012 by B. Sun and G.C. van Kooten. All rights reserved. Readers may make verbatim copies of this document for non-commercial purposes by any means, provided that this copyright notice appears on all such copies.

REPA Working Papers:

2003-01 – Compensation for Wildlife Damage: Habitat Conversion, Species Preservation and Local Welfare (Rondeau and Bulte)
2003-02 – Demand for Wildlife Hunting in British Columbia (Sun, van Kooten and Voss)
2003-03 – Does Inclusion of Landowners' Non-Market Values Lower Costs of Creating Carbon Forest Sinks? (Shaikh, Suchánek, Sun and van Kooten)
2003-04 – Smoke and Mirrors: The Kyoto Protocol and Beyond (van Kooten)
2003-05 – Creating Carbon Offsets in Agriculture through No-Till Cultivation: A Meta-Analysis of Costs and Carbon Benefits (Manley, van Kooten, Moeltne, and Johnson)
2003-06 – Climate Change and Forest Ecosystem Sinks: Economic Analysis (van Kooten and Eagle)
2003-07 – Resolving Range Conflict in Nevada? The Potential for Compensation via Monetary Payouts and Grazing Alternatives (Hobby and van Kooten)
2003-08 – Social Dilemmas and Public Range Management: Results from the Nevada Ranch Survey (van Kooten, Thomsen, Hobby and Eagle)
2004-01 – How Costly are Carbon Offsets? A Meta-Analysis of Forest Carbon Sinks (van Kooten, Eagle, Manley and Smolak)
2004-02 – Managing Forests for Multiple Tradeoffs: Compromising on Timber, Carbon and Biodiversity Objectives (Krcmar, van Kooten and Vertinsky)
2004-03 – Tests of the EKC Hypothesis using CO2 Panel Data (Shi)
2004-04 – Are Log Markets Competitive? Empirical Evidence and Implications for Canada-U.S. Trade in Softwood Lumber (Niquidet and van Kooten)
2004-05 – Conservation Payments under Risk: A Stochastic Dominance Approach (Benítez, Kuosmanen, Olschewski and van Kooten)
2004-06 – Modeling Alternative Zoning Strategies in Forest Management (Krcmar, Vertinsky and van Kooten)
2004-07 – Another Look at the Income Elasticity of Non-Point Source Air Pollutants: A Semiparametric Approach (Roy and van Kooten)
2004-08 – Anthropogenic and Natural Determinants of the Population of a Sensitive Species: Sage Grouse in Nevada (van Kooten, Eagle and Eiswerth)
2004-09 – Demand for Wildlife Hunting in British Columbia (Sun, van Kooten and Voss)
2004-10 – Viability of Carbon Offset Generating Projects in Boreal Ontario (Biggs and Laaksonen-Craig)
2004-11 – Economics of Forest and Agricultural Carbon Sinks (van Kooten)
2004-12 – Economic Dynamics of Tree Planting for Carbon Uptake on Marginal Agricultural Lands (van Kooten) (Copy of paper published in the Canadian Journal of Agricultural Economics 48(March): 51-65.)
2004-13 – Decoupling Farm Payments: Experience in the US, Canada, and Europe (Ogg and van Kooten)
2004–14– Afforestation Generated Kyoto Compliant Carbon Offsets: A Case Study in Northeastern Ontario (Biggs)
2005–01– Utility-scale Wind Power: Impacts of Increased Penetration (Pitt, van Kooten, Love and Djilali)
2005–02 –Integrating Wind Power in Electricity Grids: An Economic Analysis (Liu, van Kooten and Pitt)
2005–03 –Resolving Canada-U.S. Trade Disputes in Agriculture and Forestry: Lessons from Lumber (Biggs, Laaksonen-Craig, Niquidet and van Kooten)

2005–04–Can Forest Management Strategies Sustain the Development Needs of the Little Red River Cree First Nation? (Krcmar, Nelson, van Kooten, Vertinsky and Webb)
2005–05–Economics of Forest and Agricultural Carbon Sinks (van Kooten)
2005–06– Divergence Between WTA & WTP Revisited: Livestock Grazing on Public Range (Sun, van Kooten and Voss)
2005–07 –Dynamic Programming and Learning Models for Management of a Nonnative Species (Eiswerth, van Kooten, Lines and Eagle)
2005–08 –Canada-US Softwood Lumber Trade Revisited: Examining the Role of Substitution Bias in the Context of a Spatial Price Equilibrium Framework (Mogus, Stennes and van Kooten)
2005–09 –Are Agricultural Values a Reliable Guide in Determining Landowners' Decisions to Create Carbon Forest Sinks?* (Shaikh, Sun and van Kooten) *Updated version of Working Paper 2003-03
2005–10 –Carbon Sinks and Reservoirs: The Value of Permanence and Role of Discounting (Benitez and van Kooten)
2005–11 –Fuzzy Logic and Preference Uncertainty in Non-Market Valuation (Sun and van Kooten)
2005–12 –Forest Management Zone Design with a Tabu Search Algorithm (Krcmar, Mitrovic-Minic, van Kooten and Vertinsky)
2005–13 –Resolving Range Conflict in Nevada? Buyouts and Other Compensation Alternatives (van Kooten, Thomsen and Hobby) *Updated version of Working Paper 2003-07
2005–14 –Conservation Payments Under Risk: A Stochastic Dominance Approach (Benítez, Kuosmanen, Olschewski and van Kooten) *Updated version of Working Paper 2004-05
2005–15 –The Effect of Uncertainty on Contingent Valuation Estimates: A Comparison (Shaikh, Sun and van Kooten**)**
2005–16 –Land Degradation in Ethiopia: What do Stoves Have to do with it? (Gebreegziabher, van Kooten and.van Soest)
2005–17 –The Optimal Length of an Agricultural Carbon Contract (Gulati and Vercammen)
2006–01 –Economic Impacts of Yellow Starthistle on California (Eagle, Eiswerth, Johnson, Schoenig and van Kooten)
2006–02 -The Economics of Wind Power with Energy Storage (Benitez, Dragulescu and van Kooten)
2006–03 –A Dynamic Bioeconomic Model of Ivory Trade: Details and Extended Results (van Kooten)
2006–04 –The Potential for Wind Energy Meeting Electricity Needs on Vancouver Island (Prescott, van Kooten and Zhu)
2006–05 –Network Constrained Wind Integration: An Optimal Cost Approach (Maddaloni, Rowe and van Kooten)
2006–06 –Deforestation (Folmer and van Kooten)
2007–01 –Linking Forests and Economic Well-being: A Four-Quadrant Approach (Wang, DesRoches, Sun, Stennes, Wilson and van Kooten)
2007–02 –Economics of Forest Ecosystem Forest Sinks: A Review (van Kooten and Sohngen)
2007–03 –Costs of Creating Carbon Offset Credits via Forestry Activities: A Meta-Regression Analysis (van Kooten, Laaksonen-Craig and Wang)
2007–04 –The Economics of Wind Power: Destabilizing an Electricity Grid with Renewable Power (Prescott and van Kooten)
2007–05 –Wind Integration into Various Generation Mixtures (Maddaloni, Rowe and van Kooten)
2007–06 –Farmland Conservation in The Netherlands and British Columbia, Canada: A Comparative Analysis Using GIS-based Hedonic Pricing Models (Cotteleer, Stobbe and van Kooten)

2007–07 –Bayesian Model Averaging in the Context of Spatial Hedonic Pricing: An Application to Farmland Values (Cotteleer, Stobbe and van Kooten)
2007–08 –Challenges for Less Developed Countries: Agricultural Policies in the EU and the US (Schure, van Kooten and Wang)
2008–01 –Hobby Farms and Protection of Farmland in British Columbia (Stobbe, Eagle and van Kooten)
2008-01A-Hobby Farm's and British Columbia's Agricultural Land Reserve (Stobbe, Eagle, Cotteleer and van Kooten)
2008–02 –An Economic Analysis of Mountain Pine Beetle Impacts in a Global Context (Abbott, Stennes and van Kooten)
2008–03 –Regional Log Market Integration in New Zealand (Niquidet and Manley)
2008–04 –Biological Carbon Sequestration and Carbon Trading Re-Visited (van Kooten)
2008–05 –On Optimal British Columbia Log Export Policy: An Application of Trade theory (Abbott)
2008–06 –Expert Opinion versus Transaction Evidence: Using the Reilly Index to Measure Open Space premiums in the Urban-Rural Fringe (Cotteleer, Stobbe and van Kooten)
2008–07 –Forest-mill Integration: a Transaction Costs Perspective (Niquidet and O'Kelly)
2008–08 –The Economics of Endangered Species Poaching (Abbott)
2008–09 –The Ghost of Extinction: Preservation Values and Minimum Viable Population in Wildlife Models (van Kooten and Eiswerth)
2008–10 –Corruption, Development and the Curse of Natural Resources (Pendergast, Clarke and van Kooten)
2008–11 –Bio-energy from Mountain Pine Beetle Timber and Forest Residuals: The Economics Story (Niquidet, Stennes and van Kooten)
2008-12 –Biological Carbon Sinks: Transaction Costs and Governance (van Kooten)
2008-13 –Wind Power Development: Opportunities and Challenges (van Kooten and Timilsina)
2009-01 –Can Domestication of Wildlife Lead to Conservation? The Economics of Tiger Farming in China (Abbott and van Kooten)
2009-02 – Implications of Expanding Bioenergy Production from Wood in British Columbia: An Application of a Regional Wood Fibre Allocation Model (Stennes, Niquidet and van Kooten)
2009-03 – Linking Matlab and GAMS: A Supplement (Wong)
2009-04 – Wind Power: The Economic Impact of Intermittency (van Kooten)
2009-05 – Economic Aspects of Wind Power Generation in Developing Countries (van Kooten and Wong)
2009-06 – Niche and Direct Marketing in the Rural-Urban Fringe: A Study of the Agricultural Economy in the Shadow of a Large City (Stobbe, Eagle and van Kooten)
2009-07 – The Economics and Policy of Global Warming (van Kooten, Beisner and Geddes)
2010-01 – The Resource Curse: A State and Provincial Analysis (Olayele)
2010-02 – Elephants and the Ivory Trade Ban: Summary of Research Results (van Kooten)
2010-03 – Managing Water Shortages in the Western Electricity Grids (Scorah, Sopinka and van Kooten)
2010-04 - Bioeconomic modeling of wetlands and waterfowl in Western Canada: Accounting for amenity values (van Kooten, Withey and Wong)
2010-05 – Waterfowl Harvest Benefits in Northern Aboriginal Communities and Potential Climate Change Impacts (Krcmar, van Kooten and Chan-McLeod)
2011-01 – The Impact of Agriculture on Waterfowl Abundance: Evidence from Panel Data (Wong, van Kooten and Clarke)

For copies of this or other REPA working papers contact:
REPA Research Group
Department of Economics
University of Victoria PO Box 1700 STN CSC Victoria, BC V8W 2Y2 CANADA
Ph: 250.472.4415
Fax: 250.721.6214
http://web.uvic.ca/~repa/

This working paper is made available by the Resource Economics and Policy Analysis (REPA) Research Group at the University of Victoria. REPA working papers have not been peer reviewed and contain preliminary research findings. They shall not be cited without the expressed written consent of the author(s).

Climate Impacts on Chinese Corn Yields: A Fractional Polynomial Regression Model

Baojing Sun and G. Cornelis van Kooten

Abstract

In this study, we examine the effect of climate on corn yields in northern China using data from ten districts in Inner Mongolia and two in Shaanxi province. A regression model with a flexible functional form is specified, with explanatory variables that include seasonal growing degree days, precipitation, technical change and dummy variables to account for regional fixed effects. Results indicate that a fractional polynomial model in growing degree days explains variability in corn yields better than a linear or quadratic model. Among the tested models, the other factors show steady effects on corn yields. Growing degree days, precipitation in July, August and September, and technical change are important determinants of corn yields.

Keywords: Corn yields; fractional polynomial regression

1. Introduction

Because China accounts for nearly 20 percent of global population, it is important to study how climate affects crop yields in that country. China is the second largest maize producing country in the world after the United States (FAO 2010), but its agriculture is labor-intensive and highly vulnerable to weather and other risks. Yet few studies have examined the impact of climate factors on crop yields in China, partly because data are scarce, difficult to obtain and of varying quality. Nonetheless, as a basis for understanding current and future climate risks, it is important to increase knowledge about the relationship between existing weather records and crop yields.

Many studies have investigated climate effects on crop yields in various regions using different methods, but most have employed simple correlations or relationship that is either linear or quadratic (e.g., Williams 1972; Almaraz et al.2008； Chen et al. 2011). A recent study by Chen et al. (2011) examined the impact of weather on corn yields in three provinces (Heilongjiang, Jilin and Liaoning) located along the northeast coast of China. The authors found that the minimum temperature anomalies for May and September had a strong positive effect on corn yield, where yield in a given year was also measured as an anomaly from average yield over the 44 years (1965-2008) for which data were available. Separate regressions were estimated for each province, and for the three provinces combined, with intercept terms included only in regressions for Heilongjiang and Liaoning provinces and not for Jilin province or the aggregated regression. Not unexpectedly, temperatures in May had a stronger impact on yield than temperatures in September.

Chen et al. (2011) sought only to find the climate factors correlated with corn yields. They did not attempt to provide agronomic insights into their results. Corn is planted in early May and a high minimum May temperature is desirable in their study region; moisture is not a limiting factor in May and temperatures are not yet hot enough to adversely impact the early growth stage. Corn is harvested during September (with dates varying according to latitude); since high minimum temperatures are indicative of frost-free days, higher minima contribute to higher quality harvests and yields.

In this paper, we extend the work of Chen et al. (2011) in several directions. First, we examine a partially adjacent region to the west – the provinces of Inner Mongolia and Shaanxi, which are part of the main corn growing regions of China (Figure 1). These provinces respectively accounted for 8.3% and 3.0% of China's total corn production in 2010 (China Statistical Yearbook 2011). Thus, although our results are not directly comparable to those of Chen et al., they nonetheless provide a point for comparison. Second, we employ district level data for our two provinces – ten districts in Inner Mongolia and two in Shaanxi. Because our data are disaggregated to a greater extent than those used by Chen et al., we only have district level data for the period 1989-1999 (11 years) for Inner Mongolia and 1989-2001 (13 years) for Shaanxi. However, because we also use disaggregated weather data and panel regression, our regression models have 136 observations compared to only 44 observations used by Chen et al. Third, our data only cover the period following the opening up of the economy. While the previous research did not test for potential structural breaks in yields associated with, for example, opening up of the economy, we suspect that such a break might have occurred at or prior to 1989 (see Brown 2009, pp. 444-452) and perhaps again around 2001/2002 when China joined the World Trade Organization (WTO).

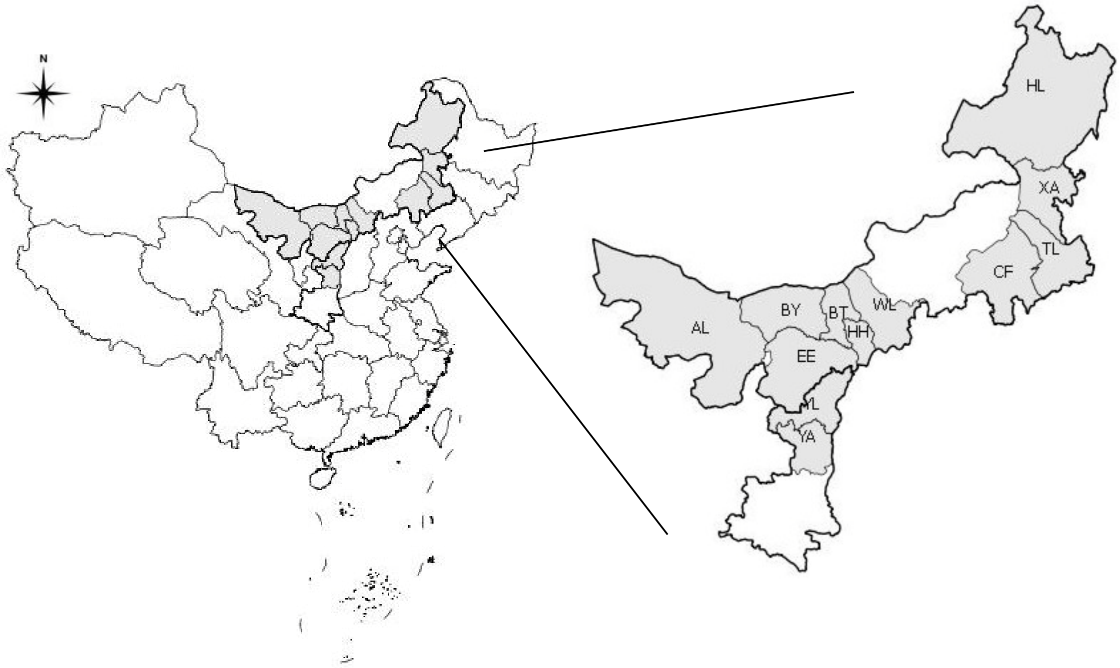

Figure 1: Study area showing 12 districts in two provinces

Finally, there is no reason to think that crop yields are a linear or quadratic function of climate variables. Rather, the relationship is more complicated. For example, Schlenker and Roberts (2006) have shown that estimation using only average temperatures could result in a biased relationship between temperature and yields. Indeed, they specify a highly nonlinear relationship and employ daily temperature information (Schlenker and Roberts 2006, 2008). These aspects are discussed below.

2. Methods

Schlenker and Roberts (2006, 2008), hereafter S&R, make the case that crop yields depend on climate factors in nonlinear fashion and that crop growth throughout the growing season is cumulative. That is, if soil moisture availability is not a constraint, then crop growth is not specifically dependent on one or several periods of warm weather but, rather, a function of the total sum of warm days during the growing season. However, extreme heat (temperatures above 36°C, say) may have an adverse impact on crop yield. Likewise, lack of soil moisture during the growing season may negatively impact crop yields.

S&R begin by postulating that relative plant growth is cumulative over time and that yield is directly dependent on plant growth. Let T represent temperature and y_{jt} represent the log of plant growth in region j in year t. Then, assuming plant growth is given by $g(T)$, the natural logarithm of crop yield is determined by the following relationship:

$$y_{j,t} = \int_{\underline{T}}^{\overline{T}} g(T)\ \Phi_{j,t}(T)\, dT + \sum_i \alpha_i\, z_{i,j,t} + D_j + \varepsilon_{j,t}, \qquad (1)$$

where $\overline{T}$ and $\underline{T}$ refer to the upper and lower bounds that observed temperatures can take; $\Phi_{j,t}(T)$ is the cumulative distribution function of temperatures (heat) over the growing season in region j during year t; $z_{i,j,t}$ are i other factors (precipitation, technology, fertilizers, etc.) that affect crop growth in region j during t; α_i are parameters to be estimated; D_j are time-invariant region-fixed effects; and $\varepsilon_{j,t} \sim N(0,\sigma_{j,t})$ are identical independently distributed error terms.

The issue of concern relates to the growth function. S&R (2006) employ a m^{th}-order Chebychev Polynomial evaluated at the m midpoints of the intervals between $\overline{T}$ and $\underline{T}$. Unfortunately, while S&R have 87,619 observations on corn yields for the period 1950-2004, our data are severely limited as discussed below. When observations are limiting, S&R (2008) recommend that, at the very least, growing degree days during the growing season and the square of growing degree days be used as explanatory variables in the regression model instead of temperatures per se. The quadratic function can capture at least one nonlinear aspect, although they find many more nonlinearities using midpoints of intervals or dummy variables representing number of days that temperatures fall within certain boundaries during a growing season. As noted, they are afforded this luxury by their large data sets.

Temperature is the primary variable of interest because precipitation is assumed not to be constraining. Nonetheless, precipitation may be important at certain times of the year. For example, too much rainfall in September may delay harvests, reduce grain quality, or even reduce overall yield. Likewise, if fields are too wet in fall, harvesting may be delayed and yields may actually decline due to decay; or if fields are too wet in spring the delay in planting leads to reduced exposure to heat and lower yields according to relation (1). In contrast to S&R, we consider precipitation effects for each month, rather than the growing season as a whole, and also consider quadratic effects of monthly precipitation as it is not obvious that the effect of precipitation on crop yield is linear.

Because we employ growing degree days over the season (denoted by G) and not S&R's m^{th}-order Chebychev Polynomial to represent the effect of heat on crop yields, we adopt the method of fractional polynomials (Royston and Altman 1994) to model the nonlinear relation. The use of G and G^2 as explanatory variables is then a special case of the more general m^{th}-order fractional polynomial regression.

Royston and Altman (1994) begin by defining a fractional polynomial of degree m written as:

$$f_m(X; \xi, \mathbf{n}) = \xi_0 + \sum_{j=1}^{m} \xi_j X^{(n_j)}, \qquad (2)$$

where the parentheses on the power term on *X* signify the following transformation:

$$X^{(n_j)} = \begin{cases} X^{n_j} \; if \; n_j \neq 0 \\ lnX \; if \; n_j = 0 \end{cases}$$

For m=2 and $\boldsymbol{n} = \{n_1, n_1\}$, we have the following:

$$F_2(X; \xi, \mathbf{n}) = \xi_0 + (\xi_1 + \xi_2)\, X^{(n_1)} \qquad (3)$$

which is nothing more than a fractional polynomial of degree 1.

Rewrite (3) as:

$$f_2(X; \xi, \mathbf{n}) = \xi_0 + \xi_1\, X^{(n_1)} + \xi_2\, X^{(n_1)}\, X^{(n_2)-(n_1)} - \xi_2\, X^{(n_1)} + \xi_2\, X^{(n_1)} \qquad (4)$$

Rearranging gives

$$f_2(X; \xi, \mathbf{n}) = \xi_0 + \xi_1\, X^{(n_1)} + \xi_2\, X^{(n_1)}\, (X^{(n_2)-(n_1)} - 1) + \xi_2\, X^{(n_1)} \qquad (5)$$

$$\Rightarrow \quad f_2(X; \xi, \mathbf{n}) = \xi_0 + (\xi_1 + \xi_2)\, X^{(n_1)} + \xi_2(n_2 - n_1)\, X^{(n_1)}\, [(X^{(n_2-n_1)} - 1)/(n_2 - n_1)] \qquad (6)$$

As $n_2 \rightarrow n_1$, the last term in parentheses in (6) becomes ln*X*. Then, upon rearranging:

$$f_2(X; \xi, \mathbf{n}) = \xi_0 + (\xi_1 + \xi_2)\, X^{(n_1)} + \xi_2(n_2 - n_1)\, X^{(n_1)}\, lnX. \qquad (7)$$

Letting $\zeta_0 = \xi_0$, $\zeta_1 = \xi_1 + \xi_2$ and $\zeta_2 = \xi_2(n_2 - n_1)$, we can write (7) as

$$f_2(X; \xi, \mathbf{n}) = \zeta_0 + \zeta_1\, X^{(n_1)} + \zeta_2\, X^{(n_1)}\, lnX.$$

This can then be generalized in the same way to m> 2 (Royston and Altman 1994):

$$f_m(X; \xi, \mathbf{n}) = \zeta_0 + \zeta_1\, X^{(n_1)} + \sum_{j=2}^{m} \zeta_j X^{(n_1)} lnX^{j-1}. \qquad (8)$$

Notice that, much like a Taylor series expansion occurs about a particular point, the flexible functional form (7) is obtained by expanding about the power n_1. This is clear from the way that expression (6) is derived. Therefore, we now generalize the fractional polynomial function further:

$$f_m(X; \xi, \mathbf{n_k}) = \zeta_0 + \sum_{k=1}^{K}\left[\zeta_{1,k}X^{(n_k)} + \sum_{j=2}^{m}\zeta_{j,k}X^{(n_k)}lnX^{j-1}\right], \quad (9)$$

where the number of potential powers equals *K*. Essentially there are an infinite number of functions that can be considered, but, in practice, we limit ourselves to n_k values that are integers between -2 and +3, with ± 0.5 ($1/\sqrt{X}$ and $\sqrt{X}$) included, although higher powers are not ruled out.

We illustrate the notation and method for identifying a regression model with several examples. Consider the fractional polynomial for variable X and let K denote powers of X and m the maximum number of terms of a particular power of X; in these examples we exclude an intercept term and terms involving other variables.[1] For example, if K=3 and m=3, we would have three powers of X with one power having three terms, although there might be other powers of X whose terms do not exceed m. Consider {*X*, (-0.5, 0, 1, 1, 1)}. This leads to the following regression equation:

$$f_3(X; \xi, \mathbf{n_3}) = \zeta_0 + \zeta_1\, 1/\sqrt{X} + \zeta_2\, lnX + \zeta_3\, X + \zeta_4\, X\, lnX + \zeta_5\, X\, lnX^2 \quad (10)$$

Likewise, {*X*, (0, 0, 1, 1, 1, 2, 3)} leads to:

$$F_3(X; \xi, \mathbf{n_4}) = \zeta_0 + \zeta_1\, lnX + \zeta_2\, lnX^2 + \zeta_3\, X + \zeta_4\, X\, lnX + \zeta_5\, X\, lnX^2 + \zeta_6\, X^2 + \zeta_7\, X^3 \quad (11)$$

One final note regards the definition of power zero, which is set equal to *lnX* as is clear from expressions (10) and (11).

In the current application, the regression model is specified as

[1] Although treated as a single variable in the text, explanatory variable *X* could just as well be considered a vector of variables.

$$y_{j,t} = \beta_0 + \sum_{k=1}^{K} \left[\beta_{1,k} G_{i,t}^{(n_k)} + \sum_{j=2}^{m} \beta_{j,k} G_{i,t}^{(n_k)} lnG^{j-1} \right] + \alpha_i\, z_{i,j,t} + D_j + \varepsilon_{j,t} \qquad (12)$$

where the dependent variable, $z_{i,j,t}$, α_i and D_j were defined in conjunction with (1), and, importantly, the second term in square brackets in (12) disappears if m=1. In (12), the βs are parameters to be estimated and $G_{i,t}$ refers to total degree days in region i during growing season t. Further, we explore only functional forms with powers $\boldsymbol{n_8} \in \{-2, -1, -0.5, 0, 0.5, 1, 2, 3\}$, although S&R use a sixth power in their function but they do not employ fractions and natural logarithms. We could also employ higher polynomials and fractions, but we limit ourselves to those indicated.

Finally, technological change can be represented by a time variable, while sales of farm chemicals (fertilizer, herbicides, pesticides) in a region might be used to represent other inputs that affect crop yield, assuming such data are available.

3. Study Area and Data

3.1 Study Area

As already noted, the study area consists of 12 districts in two provinces in northwestern China (Figure 1). The northernmost district (Yulin/YL) in Shaanxi Province is adjacent to Inner Mongolia and constitutes 3.703 million hectare (ha), while the one (Yanan/YA) to the south is 4.307 million ha (Government website of Shaanxi Province, 2012); these are part of China's Loess Plateau. Inner Mongolia spans three distinct Chinese administrative units – Northeast China, North China and Norwest China. Inner Mongolia is the third largest province, with a land base of 118.3 million ha which accounts for 12.3% of China's total land area. Of this, 53.4% of the land is plateau area, 20.9% is mountainous, and 0.8% is covered by water

(Government website of Inner Mongolia, 2012). Inner Mongolia consists of 12 districts, although the current study only considers ten of these: From east to west, these are Hulunbeier (HL, 25.0 million ha), Xinganmeng (XA, 6.0 mil ha), Tongliao (TL, 6.0 mil ha), Chifeng (CF, 9.0 mil ha), Wulanchabu (WL, 5.5 mil ha), Huhehaote (HH, 1.7 mil ha), Baotou (BT, 2.8 mil ha), Eerduosi (EE, 8.7 mil ha), Bayannaoer (BY, 6.4 mil ha) and Alashanmeng (AL, 27.0 mil ha).[2]

3.2 Data

Corn yield data are obtained from Inner Mongolia Statistic Yearbook and Shaanxi Statistic Yearbook for 1989 to 1999, 1989 to 2001, respectively. Weather data are from Ecological Environment Database of the Loess Plateau. There are 50 weather stations in Inner Mongolia for which data are available, but only 38 have weather data for a period comparable to the period for which yield data are available. In addition, records from seven weather stations in northern Shaanxi are available from the database. The corn growing season in the study area is from late April or early May to September. Therefore, we use weather records from May to September.

While yield data are available beyond 2001, we lack the same richness of weather station data for the period after 2001. Further, China joined the WTO in

---

[2] Data government websites of each respective districts (June 20th, 2012): http://www.hulunbeier.gov.cn/hlbewh/index.asp; http://www.xam.gov.cn/zwgk/zjxam/136359.htm; http://www.tongliao.gov.cn/gaik_text.asp?bid=194; http://www.chifeng.gov.cn/html/2010-05/259c01bd-a891-4b30-b6dc-40a01acefd31.shtml; http://www.wulanchabu.gov.cn/channel/wlcb/col6722f.html; http://www.huhhot.gov.cn/hhht/index.asp; http://www.baotou.gov.cn/html/btgl/dlqh.html; http://www.ordos.gov.cn/zjeedx/index.html; http://www.bynr.gov.cn/sqgk/; http://www.bynr.gov.cn/sqgk/; http://www.als.gov.cn/main/tour/survey/11012fc7-f4af-4589-817d-cadbde6f886a.shtml.

2001, so world prices could be expected to have a greater influence on crop decisions. Likewise, we lack disaggregated corn yield data prior to 1989, which can also be considered a watershed year. Market liberalization in China began in the earnest after 1985 (Huang et al., 2009), which, along with other developments, suggests that 1989 marks a particular turning point for agriculture. Thus, it makes sense to consider only the period from the turning point that appears to have occurred in 1989 to China's full entry into the world trading system.

For each district, we use the weather stations in the district to determine the temperature and precipitation associated with the crop yields. If there is no weather station in a given district, then the temperature and precipitation data for the nearest weather station to the central point (centroid) of the district is used. If there is only one weather station in the district, the data from that station is used. Finally, if there are two or more weather stations in a district, a weighted average of the precipitation and temperature data are calculated for the centroid of the district. To make the required calculations, a Geographic Information System model was built using the Quantum GIS (QGIS) tool.

After plotting district centroids, the distances between centroids and weather stations are measured using the GIS tool, and then the inverse of these distances are used as weighted coefficient to calculate weather readings. The following formula is used:

$$T_j = \sum_{k=0}^{n_j} \left[T_{k,j} + 0.006(e_{k,j} - e_j)\right](1/d_{k,j}), \quad (13)$$

where T_j is the region j centroid temperature; $T_{k,j}$ is the temperature reading and $e_{k,j}$

is the elevation at weather station k in district j; there are n_k weather stations in region j; e_j is the mean elevation at the centroid of district j, which is used to eliminate elevation differences among stations in the district; 0.006 (measured in ^oC) corrects for elevation; and $d_{k,j}$ is the distance between the centroid in district j and weather station k, with $\sum_{k=0}^{n_j} 1/d_{k,j}$=1.

We calculate growing degree days (G) over the growing season by taking the daily average temperature and subtracting from it 10^{o}C. That is, we use the following relation to calculate growing degree days:

$$G_{j,t} = \Sigma_d \, (T_{d,j,t} - 10),\ d = 1, 2, \ldots, 153, \tag{14}$$

where $G_{j,t}$ is growing degree days in region j in year t, and $T_{d,j,t}$ is the average temperature on day d in region j during the growing season of 153 days (May through September) in year t.

For the response variables, we find that the distribution of unadjusted yields is closer to a normal distribution than either the distributions of the logarithm of yields or the square root of yields. Summary statistics for the variables used in the model are provided in Table 1. Correlations among monthly GDD are strong, as shown in Table 2; thus, to avoid multicollinearity problems, we employ only growing degree days accumulated over the entire growing season rather than separate monthly values. This is not the case for precipitation (Table 3), so monthly precipitation values are employed.

Table 1: Summary statistics for variables

Variable[a]	obs	Mean	Std. Dev.	Min	Max
Yield	136	5600.47	2027.4	1020.0	11525.0
G	136	1464.7	376.9	772.3	2589.5
P_5	136	24.6	20.1	0.05	121.5
P_6	136	47.9	31.8	1.1	135.0
P_7	136	94.1	51.6	4.2	216.1
P_8	136	77.0	45.4	6.5	202.8
P_9	136	32.8	23.2	1.1	147.4

[a] G refers to growing degree days and P to precipitation, with subscripts indicating the month in the growing season (e.g., 5 = May)

Table 2: Correlations of monthly GDD

	GDD_5	GDD_6	GDD_7	GDD_8	GDD_9
GDD_5	1				
GDD_6	0.8666	1			
GDD_7	0.8505	0.928	1		
GDD_8	0.7434	0.8126	0.8531	1	
GDD_9	0.5799	0.6514	0.6621	0.7076	1

Table 3: Correlations of monthly precipitation

	p_5	p_6	p_7	p_8	p_9
p_5	1				
p_6	0.3736	1			
p_7	0.303	0.5463	1		
p_8	0.0959	0.2338	0.4556	1	
p_9	0.0929	0.307	0.3258	0.1707	1

4. Results

Estimation results are provided in Table 4 for linear, quadratic, and degree1 to degree 4 terms of seasonal growing degree days (G); these are represented by models #1 through #6, respectively. In each of the models, dummy variables are used to capture district fixed effects. With some exceptions, unexplained differences among districts turn out to be important in explaining differences in corn yields across our study region.

Table 4 Fractional Polynomial Regression in Different Degrees

Explanatory Variable	#1 Linear	#2 Quadratic	#3 Degree 1	#4 Degree 2	#5 Degree 3	#6 Degree 4
$G^{-0.5}$				60408.94***		
$G^{-0.5}$×(lnG)				33287.70***		
lnG						
$(\ln G)^2$						
$(\ln G)^3$						
$(\ln G)^4$						
G	-2.19	3288.29				
G^2		-1493.91				
G^3			-193.06		6039.99***	-1974.18
G^3×lnG					-12611.59***	10197.98
G^3×$(\ln G)^2$					6789.67***	-18602.66
G^3×$(\ln G)^3$						10234.37
P_5	342.37	637.18	571.42	388.99	-234.23	-246.98
P_5^2	106.39	-89.98	-53.71	108.45	705.36	775.54
P_6	480.73	401.72	484.61	2.02	-233.89	-132.72
P_6^2	-64.52	26.673	-33.58	333.04	446.42	338.35
P_7	1419.79	2007.41**	1883.55**	1665.63*	1423.57	1856.59**
P_7^2	-423.41	-631.72	-591.04	-493.59	-406.24	-581.43
P_8	1464.12*	1516.56*	1524.75**	1731.73**	1735.88**	1594.18**
P_8^2	-669.36*	-701.94*	-696.52*	-837.69**	-850.59**	-766.56**
P_9	-2315.02**	-2284.59**	-2314.33**	-2144.39**	-1986.00**	-2053.41**
P_9^2	934.79	954.23	968.52	817.36	728.83	807.56
time	2600.48***	2441.46***	2437.68***	2543.33***	2681.29***	2545.65***
D_2	-1781.37***	-1645.52***	-1655.57***	-1763.17***	-1894.96***	-1835.49***
D_3	2154.00**	2866.55**	2594.52**	2737.50***	3641.69***	4910.06***
D_4	3103.54**	4326.82***	4066.10***	4166.33***	4545.68***	5083.30***
D_5	692.72	1015.91	989.87*	710.71	347.38	540.65
D_6	-834.02	-648.63	-639.69	-903.86	-1116.74*	-927.03
D_7	398.08	574.87	581.33	394.49	137.583	243.88
D_8	-4073.59***	-2762.26**	-3102.68***	-2217.60*	-1440.22	-1307.12
D_9	-12.39	50.34	70.48	-68.95	-253.14	-234.28
D_{10}	-2172.75**	-1044.17	-1291.66**	-1277.07	-945.61	-456.07
D_{11}	-2517.58***	-2154.62***	-2167.15***	-2517.66***	-2914.38***	-2703.06***
D_{12}	-32.61	324.81	293.71	77.37	-149.64	67.01
cons	6392.88***	6213.29***	6203.66***	6438.78***	6756.99***	6662.62***
$\overline{R}^2$	0.776	0.795	0.788	0.803	0.824	0.832
Deviance	2223.65	2208.53	2215.38	2201.18	2183.19	2176.56
Res.SD	950.38	---	918.66	875.86	823.50	807.36
Dev.dif.	51.00	---	42.73	28.53	10.54	3.91
Prob	0.00	---	0.00	0.00	0.08	0.22

[b] ***, ** and * indicate coefficients are significant at the 0.01, 0.05 and 0.1 levels, respectively; G (season growing degree days)， precipitation and time are standardized in the regressions.

For models # 1, #2 and #3 – models with linear, quadratic and degree-1 terms of G – the estimated parameters are not statistically significant. As the degree on the growing degree days variable increases from 1 to 4 for models #3 through #6, the model fit improves as indicated by $\bar{R}^2$ while the level of statistical significance of the deviance difference statistic equals 0.01 for model #1, and 0.01 and 0.10 for models #3 and #4, respectively; however, for model #1, the estimated parameters on G are not statistically significant. For model #6, the deviance difference is not statistically significant, which indicates that the function form leads to an 'over fitting' of the model; this is also indicated by the lack of statistical significance on any of the estimated parameters on G.

The best models from a statistical standpoint are, therefore, #4 and # 5, which respectively explain 80.2% and 82.3% of the variation in corn yields. For model #4, fractional polynomial terms for growing degree days are combinations of $1/\sqrt{G}$ and $(1/\sqrt{G}) \times ln\, G$, with relationship shown in Figure 2. For model #5, fractional polynomial terms of growing degree days consist of G^3, $G^3 \times ln\ G$, and $G^3 \times (ln\ G)^2$, with this estimated function plotted in Figure 3.

In models #4 and #5, higher levels of precipitation in July (#4) and August (#5) have a positive effect on corn yields, but too much precipitation in any given month will reduce yields as indicated by the negative coefficient on the rainfall squared term (although it is insignificant for July precipitation). Precipitation in September negatively affects crop yields, probably because this is the harvest period; additional rainfall is no longer needed for crop growth and, indeed, rainfall could disrupt harvesting operations causing some crop loss or cause corn yields to fall as precipitation might damage the crop. Both of these findings are similar to those of Chen et al. (2011).

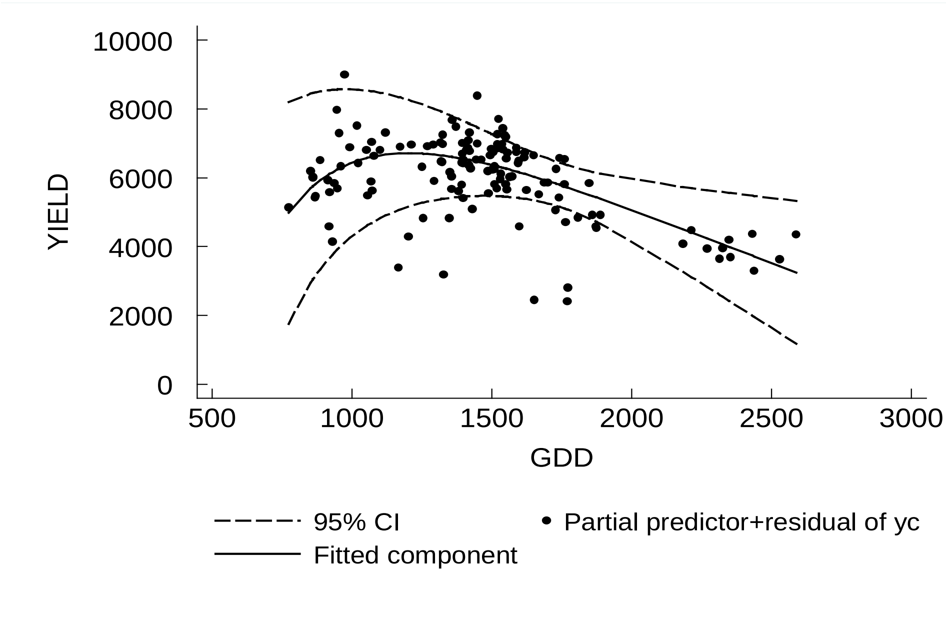

Figure 2: Fractional Polynomial of GDD, of degree 2 with powers (-.5,-.5)

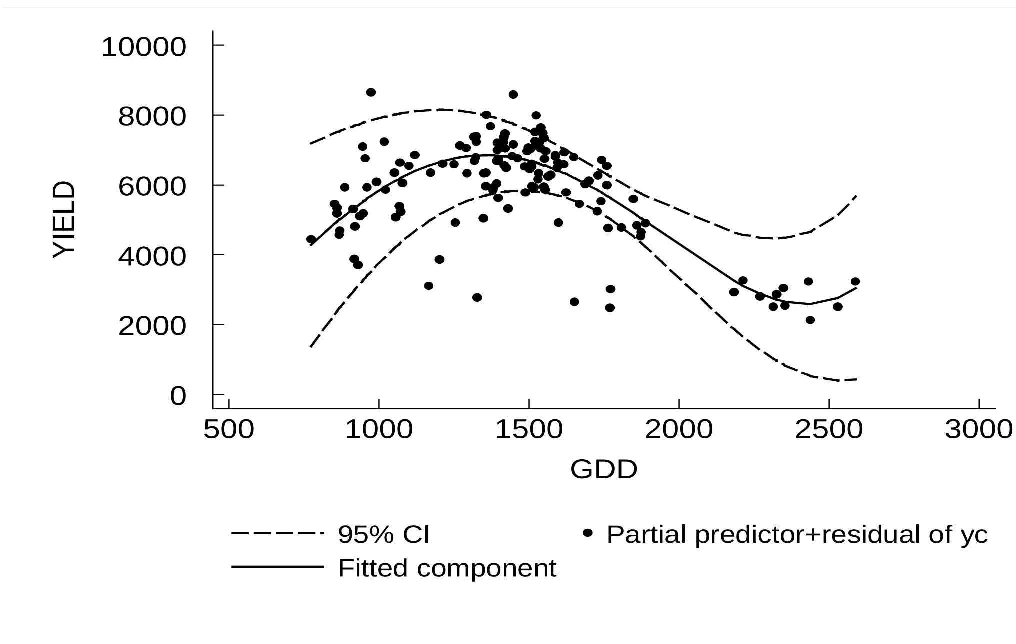

Figure 3: Fractional Polynomial of GDD, of degree 3 with powers (3, 3, 3)

As expected, time variable has a strongly positive impact on corn yields. This indicates that farmers were adopting new technologies, whether improved varieties of corn, more fertilizer, better or newer equipment, or some other improvement.

As a final check on our model, we employ the estimated parameters for models #4 and #5 in a Monte Carlo simulation (with sampling from distributions about the estimated parameters using the estimated standard errors as well as the overall standard error of the estimated model) to determine average corn yield for each of the 12 districts. These are provided in Figure 4. In the figure, estimated yields (Y_4) derived from model #4 are close to actual yields (Y), with the exception of district 3, and estimated yields (Y_5) from model #5 are close to actuarial yields (Y), except for districts 3 and 11. Overall, model #4 appears to better predict corn yields than the other models.

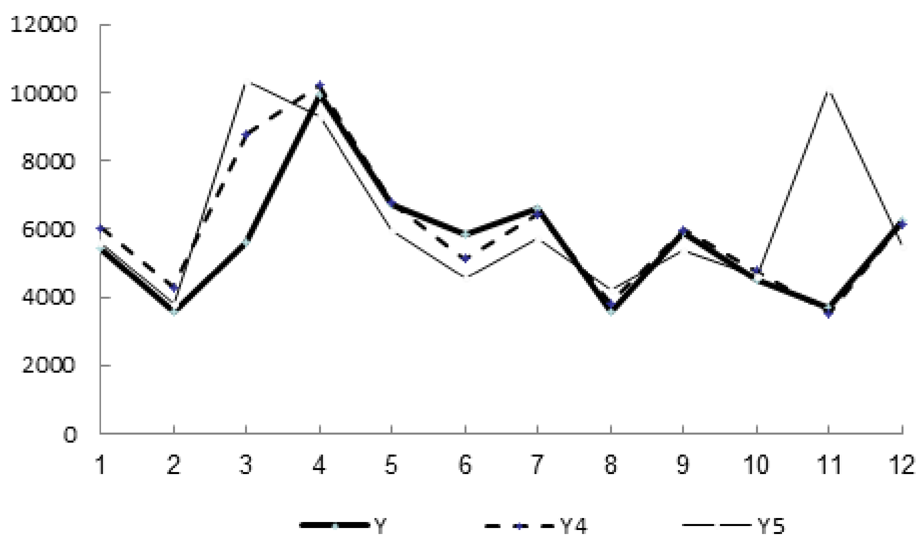

Figure 4: Actual average (Y) and Monte Carlo simulated yields for models #4 (Y4) and #5 (Y5) for the 12 districts

5. Discussion

In this study, we investigated the impact of climate variables on corn yields in northwestern China. The most important result is that accumulated heat throughout the growing season (as measured by seasonal growing degree days) is likely the most important variable influencing corn yields. However, the relationship between GDD and yields is subtle and cannot be captured adequately by a linear or quadratic functional relation. Rather, the relationship is much more complicated and is best determined using a highly nonlinear regression model.

Not surprisingly, precipitation is also important but adds to crop yields primarily during the peak of the growing season, indicating that it is of less import than heat units. That is, for our study region, moisture is important, but there is likely enough soil moisture that rainfall in mid-summer simply provides a boost to yields that is declines rapidly with higher levels of rainfall. Given the size of the estimated parameters, district fixed effects and adopted technical advances are also important factors explaining crop yields.

Finally, we find that the two best fitted models capture 80% or more of the variation in corn yields. In that case, the estimated regression models could potentially be used as a basis for developing weather-indexed insurance products in this study area. Given that farmers in western and central China have expressed interest in weather-indexed insurance to mitigate weather risks (Turvey et al. 2009; Liu et al. 2010), an extension to the current work is to examine how one might use estimated crop-yield regression models to construct such financial instruments.

References

Almaraz, J.J., Mabood, F., Zhou, X., 2008. Climate change, weather variability and corn yield at a higher latitude locale: Southwestern Quebec, *Climatic Change* 88: 187-197.

Brown, A., 2009. *The Rise and Fall of Communism*. Doubleday Canada, Toronto.

Chen, C., Lei, C., Deng, A., Qian,C. Hoogmoed, W., Zhang, W., 2011. Will higher minimum temperature increase corn production in northeast China? An analysis of historical data over 1965-2008, *Agricultural and Forest Meteorology* 151: 1580-1588.

FAO, 2010. http://faostat.fao.org/site/339/default.aspx (June 20th, 2012).

Government website of Inner Mongolia, 2012. http://intonmg.nmg.gov.cn/channel/zjnmg/col6722f.html(June 20th , 2012).

Huang, I., Rozelle, S., Martin, W., Liu, Y., 2009. China. Chapter 3 in *Distortions to Agricultural Incentives in Asia* (pp.117-161) edited by K. Anderson and W. Martin. The International Bank of Reconstruction and Development/The World Bank, Washington DC.

Liu, B.C., Li, M.S., Guo, Y., Shan, L., 2010. Analysis of the demand for weather index agricultural insurance on household level in Anhui, China, *Agriculture and Agricultural Science Procedia* 1: 179-186.

National Bureau of Statistics of China, 2011.China Statistical Yearbook. China Statistical Press, Beijing.

Royston, P., Altman, D.G., 1994. Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling, *Applied Statistics* 43(3): 429-467.

Schlenker, W., Roberts, M., 2006. Nonlinear effects of weather on corn yields, *Review of Agricultural Economics* 28(3): 391-398.

Schlenker, W., Roberts, M., 2008. Estimating the impact of climate change on crop yields: The importance of non-linear temperature effects. NBER Working Paper 13799. 37pp. February. Cambridge, MA: National Bureau of Economic Research. http://www.nber.org/papers/w13799 (accessed May 2, 2012)

Turvey, C.G., Kong, R., Belltawn, B.C., 2009. Weather risk and the viability of weather insurance in western China. Selected paper presented at the annual conference of the American Agricultural Economics Association. 33pp.

Williams, G.D.V., 1972. Geographical variations in yield-weather relationships over a large wheat growing region, *Agricultural and Forest Meteorology* 9, 65-283.