Crop Yield's Variation and Climatic Conditions in Arizona

by

Haoying Wang

A Thesis Submitted to the Faculty of the

DEPARTMENT OF AGRICULTURAL AND RESOURCE ECONOMICS

In Partial Fulfillment of the Requirements

For the Degree of

MASTER OF SCIENCE

In the Graduate College

The University of Arizona

2010

STATEMENT BY AUTHOR

This thesis has been submitted in partial fulfillment of requirements for an advanced degree at The University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this thesis are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: Haoying Wang

Haoying Wang

APPROVAL BY THESIS DIRECTOR

This thesis has been approved on the date shown below:

Satheesh Aradhyula

Satheesh Aradhyula
Professor of Agricultural and Resource Economics

5/14/10
Date

ACKNOWLEDGEMENTS

I would like to thank my advisor and mentor, Dr. Satheesh Aradhyula for his constructive guidance throughout this research. I am truly grateful for his advice and constant encouragement that helped me complete this thesis.

I would also like to express gratitude to my committee members, Dr. Gary Thompson and Dr. Roger Dahlgran for providing me invaluable insight into different aspects of this research. I would especially like to thank Dr. Michael Crimmins from Department of Soil, Water, and Environmental Science for generously providing me with help in the collection of data used in this Research.

This work was financially and technically supported by the Department of Agricultural and Resource Economics, University of Arizona. Without this support, I am sure I can not go any further in the areas this research has been involved.

Finally, this thesis is dedicated to my parents, who as traditional farmers but taught me to be a good learner in the new world, and my wife, Mrs. Si Luo, whose support and patience made the completion of this research and thesis possible.

Table of Contents

List of Tables

List of Figures

Abstract

This research explores the influence of climatic conditions on the mean and variance of crop yield in Arizona. We develop a stochastic seemingly unrelated regression model for capturing the impact of climate change on crop yield, and the correlation of yield among crops. By design the model can differentiate both long term impact (year-to-year, on the mean of yield) and short term impact (within year, on the variance of yield) of climate change simultaneously. Estimating the model for the 1965-2008 period across seven leading farming counties in Arizona, we find that the climatic conditions-crop yield relationship is crop-specific. Temperature is highly significant in explaining the mean change of both cotton and hay yield; yearly precipitation level is only slightly significant in explaining the mean change of cotton yield, while it is highly significant in explaining the mean change of hay yield. Results also indicate that as the within year variation of precipitation and temperature increase, the yield variation becomes smaller. Finally, the yield variation of cotton and hay tends to be positively correlated, although the strength of correlation varies among different counties.

Chapter 1: Introduction

1.1 Research Motivation

As the key development of human civilization, agriculture has been the basic tool for man's coexistence with physical world. One explanation for this coexistent relationship can be the interrelated processes between climate change and agricultural production, especially for field crop farming. Solar radiation, temperature, and precipitation are the main drivers of crop growth; therefore agriculture has always been highly dependent on climate patterns and variations. Overall, climate change could result in a variety of impacts on agriculture. One of these effects is the change of production patterns due to higher temperature and precipitation variation. While both climate change and agricultural production take place on a global scale, the impact of climate change on agriculture tends to vary from region to region. Climate is generally a concept associated with large geographic area, and the agricultural production driven climate change can only be significant on an equivalent large scale (global scale, etc.). Therefore, for a certain geographic region the climate-to-agriculture impact is dominant among the interaction between climate and agriculture. On the other hand, over the past century, human innovation has led to technological advances in agriculture that have allowed substantial increase in crop yields, in part stimulated to meet population growth and industrialization. These technological advances constitute the essential part of human adaptation in agriculture to both natural and social conditions change.

While agricultural production has experienced stable and substantial growth during the past century (Gardner, 2002), from year to year the production performance has also shown lots of fluctuation. The fluctuation in agricultural production has largely stemmed from a fluctuation in yield of major crops which in turn is the result of the behavior of natural environment (namely climate); and other events like pest attacks on crops, which are indirectly related with climatic conditions. All of these natural fluctuations of agriculture can then be transmitted to market via demand and supply mechanism, but this transmission is mostly featured as short period cycle or seasonality. Often time, beyond the complexity of these relationships we are eager to understand them for different purpose. A most common way to show this aspiration is crop yield forecasting. A traditional way to do crop yield forecasting is meteorological methods, which generally combines simulation model and satellite-derived data or GIS/GPS-based information (Wit and Diepen, 2008; Thomas *et al.*, 2002). Ferris (2006) notes that, agricultural forecasts, whether for the coming year or several years into the future, have been based on assumptions of normal weather and trend of crop yields. That weather is seldom normal and that yields seldom fit trends are well recognized. However, relatively little attention has been given to projecting crop yields stochastically even though computer capacity and software programs are available to do so. Ferris (2006) gives two reasons: one reason is that the task is more challenging than to assign standard deviations to various crop yields and simulate normal distributions using random number generators. For one, deviations of crop yields from trends may be correlated especially if the locations of the crops overlapping. Even to model crop yield individually, those correlations must be taken into account. Secondly, deviations of crop yields from trends may not necessarily be normal. Typically, crop yield deviations are skewed to the low side, with yields lower in poor crop years than higher in favorable crop years. These observations encourage revisiting crop yield forecasting in a more stochastic and more systematic way.

1.2 Research Question

Long term (year-to-year) climate change can have important implications for the adaptation in agriculture production (especially crops); at the same time, influence from irregular change and random shock (short term (within year) variation) of climatic conditions such as temperature, precipitation and humidity are also important. The differences between these two kinds of impacts can be distinct. In Arizona, in 2006, there are about 10,000 farms with an average size of 2,610 acres. At the same time, farm income from crops reached $1,558 million dollars compared to $1,321 million dollars of farm income from livestock. The top agricultural crop commodities in Arizona are lettuce, cotton and hay. On the other hand, Arizona has an unique climate pattern. Due to its large area and variations in elevation the state has a wide variety of localized climate conditions. In the lower elevations, the climate is primarily desert, with mild winters and hot summers. However, the northern third of Arizona is a plateau at significantly higher altitudes than the lower desert, and has an appreciably cooler climate, with cold winters and mild summers. All of these characteristics together give rise to an interest research question on the relationship between Arizona crop yield variation and its unique climatic conditions.

So the general research question comes as what is the impact of climatic conditions' variation on agricultural production, and what is the difference between long term impact and short term impact? Finally, what are the ecological and economic results of these impacts? As an empirical research question, we wonder how to model the climate-crop relationship in a systematic way with appropriate stochastic features, and how would the modeling facilitate the forecasting and inference?

1.3 Literature Review

Aside from the general discussion of crop yield behavior, a thorough understanding of crop yield pattern has to take the consideration of both implicit trends and deviation from these trends. In other words, deviation from crop yield trends is as important as trends themselves since the deviation from past trends indicates the path of the future trends. In literature, considerable attention has been devoted to agricultural effects of climate change (Mendelsohn et al., 1994, 1996; Adams et al., 1998; McCarthy et al., 2001; Seo and Mendelsohn, 2008), but studies focus only on the impact of long term climate change on expected crop yields. Deschênes and Greenstone (2007) measured the economic impact of climate change on US agricultural land by estimating the effect of random year-to-year variation in temperature and precipitation on agricultural profits, which only captures the long term trends into forecasting. In fact, while climatic cycles generate the natural seasonality, this seasonal pattern itself can have significant variation which comes from short term irregular or abrupt changes of climatic conditions such as temperature, precipitation, humidity, etc. An illustration of this irregular change is ENSO (El Niño-Southern Oscillation) events and their impact. Adams *et al*. (1999) found that much of the variation in climate can be traced to the ENSO phenomenon, which can be reflected in agricultural production, prices and profits. Because crop growth generally follows the natural seasonal pattern, so the short term variation of climatic conditions can have significant impact on the planting, growth and harvest of crops. However, little empirical evidence is available on crop yield variations in response to the short term alterations in climatic conditions, while this kind of information is important in simulation studies of climatic conditions-crop yield relationships (Mearns et al., 1997; Isik and Devadoss, 2006) and decision making in production.

Riha *et al*. (1996) for the first time consider the within-year variability of temperature and precipitation (without altering long-term mean values) into research on crop predictions. Their results indicate that average predicted yield decreases with increasing temperature variability where growing-season temperatures are below the optimum specified in the crop model for photosynthesis or biomass accumulation. However, increasing within-year variability of temperature has little impact on year-to-year variability of yield and the influence of changed precipitation variability on yield was mediated by the nature of the soil. Even though these findings stand on pure biological perspectives, it gives the hint that within-year variability of temperature and precipitation could be important in crop yield forecasting.

From a more general statistical perspective, the most important agricultural impact of climate change should be its influence on the distribution of future crop yields rather than solely on average trends. Many studies have been done regarding the effect on the mean of such distributions but few have addressed the effect on variance. Chen *et al.* (2004) examined the potential effects of climate change on crop yield variance in the context of current observed yields and then extrapolates to the effects under projected climate change. In their study, maximum likelihood panel data estimates of the impacts of climate on year-to-year yield variability are constructed for the major U.S. agricultural crops. The estimation results indicate that changes in climate modify crop yield levels and variances in a crop-specific fashion. For sorghum, rainfall and temperature increases are found to increase yield level and variability. On the other hand, precipitation and temperature are individually found to have opposite effects on corn yield levels and variability.

Isik and Devadoss (2006) consider change in climatic conditions from year to year as the major determinants of the crop yield fluctuations, which is a relatively long time period measure of climatic variation. In this research, we construct an alternative within-year measure of climatic variation which can capture the change in climatic conditions more efficiently in a short time span (month-to-month). Since different crops could react differently to the alternations in climatic conditions (Adams et al., 1998), some crops may depend more on certain climatic condition while other crops may prefer different climatic conditions. On the other hand, across different regions, the combination of climatic conditions is also heterogeneous; some regions may be rich in their sunshine (related to temperature), while other regions are more abundant in precipitation. An example is the climatic difference between Arizona and its neighbor California. Mediterranean climate prevails in much of California, while Arizona is known for its dry desert climate, which presents exceptionally hot summers and mild winters. These differences and heterogeneity indicate that the relationship between climatic conditions and crop production is not only crop-specific, but also region-specific. So we would expect that the main crops in a certain region are generally more sensitive to its more significant climatic conditions, which is the consequence of adaptation in crop growth and agricultural production patterns in the long term (Mendlesohn *et al*., 1996).

Mendlesohn *et al*. (1994) found that higher temperatures in all seasons except autumn reduce average farm values, while more precipitation outside of autumn increases farm values. Actually, in terms of impact on agricultural production, different climatic conditions have complicated interaction while crop growth and agricultural production patterns respond to them adaptively. In this research, we consider a quadratic function form to capture the mean change of crop yield, which will account for the interaction among all of the climatic conditions included. As another climatic condition varies, the yield response of a crop to certain climatic condition will keep shifting. This is consistent with Mendlesohn *et al*. (1994)'s finding that the effects of higher temperature range from mildly harmful to unequivocally beneficial. And the range of this shift is always determined by the change of other climatic conditions. As another improvement, we intend to combine the impacts of climatic conditions on different crops into one system, and within this framework we can estimate the correlation of variations among different crops more efficiently. A traditional way to estimate this correlation is a two-step procedure, which estimates each climatic conditions-crop yield relationship being studied first, and then calculates the correlation coefficients of all residuals (Isik and Devadoss, 2006). The unobserved components of error for different equations (climatic conditions-crop yield relationships) are likely to be correlated, and these structural errors may come from some other common factors (such as policy impact) which can not be efficiently estimated by an equation-by-equation estimation procedure.

This research uses a seemingly unrelated regression (SUR) model to systematically explore how certain climatic conditions can affect both mean and variance of crop yields. A candidate benchmark model is based on a stochastic function specification which possesses sufficient flexibility so that the effects of inputs on the deterministic component of function are different than on the stochastic component (Just and Pope, 1978). By design our model specification makes it feasible to differentiate the impacts of long term climate change and short term climatic variation simultaneously. To fit the specification of model, several new measures of short term climatic variation are also developed, which are different from traditional measure of climatic factors (Stallings, 1961). A main technical feature of this research is that, the proposed stochastic SUR model can handle both correlated structural errors and heteroskedasticity simultaneously while resulting in more efficient estimates. Note that, the estimated system can also be used for forecasting of crop yield in coming years.

1.4 Structure of This Research

The focus of this research is to examine the impact of climatic variation on crop yield in Arizona, an area with unique climate patterns. The main part of this research has a structure of five chapters. This first chapter gives a brief introduction to the motivation and basic question of this research. The following chapter will talk about the methodology and model specification employed by this research. The methodology section includes discussion of forecasting methods, the optimality of OLS and efficiency of seemingly unrelated regression (SUR) model, heteroscedasticity and correlation in SUR Model. Different extension of SUR model will also be discussed briefly in this chapter, which includes SUR model with identical regressors, SUR model with unequal numbers of observations, and nonparametric SUR models. The third chapter will discuss data and descriptive statistics, including data coverage, data collection and treatment of missing values, descriptive statistics of variables. The fourth chapter analyzes estimation results, marginal effect, forecasting and correlation among crop yields. This chapter will also discuss heterogeneity cross different counties in Arizona. The last chapter will give implications of this research, and a brief conclusion and hints for future research.

Chapter 2: Methodology and Model Specification

2.1 All about Forecasting

As mentioned in chapter 1, this research uses a seemingly unrelated regression (SUR) model to systematically explore how certain climatic conditions can affect both the mean and the variance of crop yields. In the other words, we try to use a parametric model to capture the interrelated process between crop yield and climatic conditions, and then implement forecasting based on the relationship. Econometric forecasting is based on the application of specific economic models that suggest the relationship(s) between various variables. Generally, econometric forecasting methods can be classified into two very broad categories (Kennedy, 1992). (1) Structured econometric models. Once estimates of the parameters of an economic model are available, the model can be employed to forecast the dependent variable if the associated values of the independent variables are given. The model used can range in sophistication from a single equation with one or two explanatory variables to a large simultaneous-equation model with scores of variables. The SUR model considered in this research can be categorized into this case. (2) Time series models. Time series can be characterized as consisting of a time trend, a seasonal factor, a cyclical element and an error term. A wide variety of techniques is available to break up a time series into these components and thereby to generate a means of forecasting future behavior of the series. These methods are based on the supposition that history provides some guide as to what to expect in the future. For the forecasting and prediction involved in this research, an applicable time series model is multivariate GARCH model (Bauwens *et al*., 2006). The most obvious application of multivariate GARCH models is the study of the relations between the volatilities and co-volatilities of several markets. Is the volatility of a market leading the volatility of other markets? If we consider a crop as a market, then the similar questions rises: Is the fluctuation of a crop yield leading or intervening the yield fluctuation of other corps?

In this research, we choose to focus on the first category: structured econometric models. The essential reason is that we are motivated to examine the interaction between crop yield and climatic conditions, and the crop yield forecasting is the derivative product of this process. In this way, more exogenous information can be used, which can explicitly relate climate variables to crop yield-an objective of the study.

2.2 Basic Model Specification

We assume first that the technological change and capital investment in agricultural production follow a stable increasing pattern which can be efficiently captured by a time trend variable. In fact, technological change and most of capital investment (such as infrastructure, irrigation system) can only affect crop yield in long term, which also suggests the introduction of a time trend variable in the deterministic part of equation. Another basic assumption is that, there is no abrupt land use change which the data seems to support except possibly for Yuma County during 1970s. The land use change in Yuma County in 1970s is discussed in next chapter. Hence, as the first step, the impacts of climatic conditions on crop yield can be isolated from the impacts of other anthropogenic input factors. To represent the relationship between climatic conditions and crop yield, we use the stochastic production function specification proposed by Just and Pope (1978):

$$y = F(x) = f(x) + \varepsilon h(x), \qquad E(\varepsilon) = 0, \quad V(\varepsilon) = \sigma \quad (\sigma \text{ is constant})$$

Where y is crop yield and x is a set of explanatory variables in our case, note that the explanatory variables in $f(x)$ and $h(x)$ are not necessarily identical. After customizing the function into the climatic conditions-crop yield relationship being examined, the basic model (for a specific crop) can be written as:

$$y_{it} = f(x_{it};\alpha) + \omega_{it} h(z_{it};\beta)^{1/2} \qquad \omega_{it} \sim N(0,\sigma_{\omega}^2) \tag{2-1}$$

Where: $f(x_{it};\alpha) = \alpha_0 + \alpha_1 Year_t + \alpha_2 P_{it} + \alpha_3 T_{it} + \alpha_4 P_{it} T_{it} + \alpha_5 P_{it}^{\ 2} + \alpha_6 T_{it}^{\ 2}$

$$\begin{aligned} h(z_{it};\beta) &= e^{\beta_0 + \beta_1 P_{it} + \beta_2 T_{it} + \beta_3 VP_{it} + \beta_4 VT_{it}} \\ &= e^{\beta_0} e^{\beta_1 P_{it} + \beta_2 T_{it}} e^{\beta_3 VP_{it} + \beta_4 VT_{it}} \end{aligned}$$

$y_{it} - crop\ yield,\ \ i - county,\ \ t - year$

$$\begin{cases} Year_t - time\ trend \\ P_{it} \ - yearly\ cumulative\ precipitation \\ T_{it} \ - yearly\ average\ temperature \\ VP_{it} - within\ year\ precipitation\ \text{variance} \\ VT_{it} - within\ year\ temperature\ \text{variance} \end{cases}$$

The specification in equation (2-1) has desirable properties. First, it can separate the deterministic and stochastic parts of the function, and then it gives the mean and variance of yield in a concise way:

$$E(y_{it}) = f(x_{it};\alpha) \tag{2-2}$$

$$V(y_{it}) = \sigma_{\omega}^2 h(z_{it};\beta) = h(z_{it};\beta) \qquad (if\ \sigma_{\omega}^2 = 1) \tag{2-3}$$

In this framework, there is no need to use identical explanatory variables in both deterministic and stochastic parts. We create two new variables for stochastic part of the function (1): within year precipitation variance (VP) and within year temperature variance (VT). These two newly introduced variables are calculated based on monthly average observations of precipitation and temperature, which are simply the variances of 12 monthly average observations respectively. Because within year variation of climatic conditions is also important to crop growth and hence crop yield, we expect that these two variables would explain a large part of crop yield's short-term variation.

On the other hand, for the purpose of comparison, we also include yearly cumulative precipitation and average temperature in the stochastic component of function (1). As measures of long-term climate change, we expect these two variables are significant in accounting for the mean crop yield change (deterministic component of function). However, to provide a robust explanation for the variance change of crop yield, we expect second moments of within-year variance of temperature and precipitation are equivalently significant and convincing in terms of both methodology and economic intuition. The reason is that more information will be used with both first and second moments, and within-year variance can also account for the change of seasonality.

Another implication of this specification is that, we assume there is heteroskedasticity in individual equation (climatic conditions-crop yield relationship). The observed part of heteroskedasticity will be captured by $h(z_{it};\beta)$, and the remaining unobserved random part of heteroskedasticity will be captured by ω_{it} which follows normal distribution by assumption. In a correlated system, these unobserved random parts comprise structural errors which come from some common unobserved factors. In this case, using a SUR model which combines all of the equations will be more efficient than imposing a Least Squares Method or MLE (Maximum Likelihood Estimation) on each of single equations individually. The reason will be discussed in next section.

2.3 The Optimality of OLS

By choosing seemingly unrelated regression methods, we need to be convinced that a SUR model which combines all of the equations will be more efficient than imposing a Least Squares Method. The efficiency of SUR model over OLS model (ordinary least squares) has been discussed by different authors (Kmenta, 1986;

Srivastava and Giles, 1987; Baltagi, 1998). In estimating the coefficients of the seemingly unrelated regression equations, one possible approach is to apply the OLS method to each equation separately. And the OLS estimators of the regression coefficients are unbiased and consistent. Thus the major question is that of efficiency, by estimating each equation separately and independently, we are disregarding the information about the mutual correlation of the disturbances, and the efficiency of the estimators becomes questionable. In the context of the seemingly unrelated regressions, Kmenta (1986) proved that, when taking into account the correlation of the disturbances across equations, OLS estimation of the seemingly unrelated regressions is not efficient. However, there are situations the OLS and FGLS/MLE estimators are identical. One special case is that variance covariance matrix of the equation system is know to be diagonal, which means there is no correlation of the disturbances across equations (Srivastava and Giles, 1987).

2.4 Heteroscedasticity, Correlation and SUR Model

In this research, we expect there are disturbance correlations among different objects (crops and regions) being studied, and which is of importance in studying behavior of all crops involved in agricultural production as a whole. Since a least squares estimation on each separate equation does not necessarily ensure efficient estimators, more generalized estimation methods should be implemented. Based on the basic model specification (2-1), we construct a stochastic SUR model with identical/different regressors in each equation which represents the relationships being studied respectively (2-4). And the system is also featured by heteroskedasticity and spatial (cross section) correlation.

$$y_{jit} = f_j + u_{jit},\ \ u_{jit} = \omega_{jit} h_j \left(z_{jit}; \beta_j\right)^{1/2} \quad j = j\text{th equation} \qquad \text{(2-4)}$$

We can write the N-equations system as following:

$$\begin{cases} y_{1it} = f_1(x_{1it};\alpha_1) \ + \omega_{1it} h_1\left(z_{1it};\beta_1\right)^{1/2} \\ y_{2it} = f_2(x_{2it};\alpha_2) + \omega_{2it} h_2\left(z_{2it};\beta_2\right)^{1/2} \\ \vdots \qquad \vdots \qquad\qquad \vdots \\ y_{nit} = f_n(x_{nit};\alpha_n) \ + \omega_{nit} h_n\left(z_{nit};\beta_n\right)^{1/2} \end{cases} \tag{2-5}$$

More specifically, there are different ways to define this system according to different cross section measurements.

(1) By crops: j = jth crop (cotton, hay, corn, wheat, etc.) (2-5-1)

In this case, we put all of the data together as a pool since every crop is planted over the area being studied. In other words, we have a SUR specification with identical regressors here.

(2) By regions: j = jth region (can be county, district or state level) (2-5-2)

In this case, by definition we treat different regions in the area being studied separately and the data is no longer used as a pool. Therefore, we have a SUR specification with different regressors. As mentioned above, in both cases, our model is featured by heteroskedasticity and spatial (cross section) correlation.

As far as estimation procedure is concerned, most of theoretical derivation assumes that the elements of the variances and covariances matrix of the regression disturbances are known. However, if they are not known, as is the more general case, we need to find consistent estimators of the variances and covariances matrix. One possibility is to estimate these variances and covariances from OLS residuals as suggested by Zellner (1962), which involves application of Aitken's generalized least-squares to the whole system of equations (Aitken, 1935, 1942). It is found that the regression coefficient estimators so obtained are at least asymptotically more efficient than those obtained by an equation-equation application of least squares. The two-stage Aitken estimator of coefficients is asymptotically equivalent to the generalized least squares estimator and, therefore, to the maximum likelihood estimator of coefficients (Kmenta, 1986). An alternative solution to the problem of estimating variances and covariances matrix of the regression disturbances is to use the maximum likelihood method. The maximum likelihood estimator can also be used for the purpose of testing the hypothesis that variances and covariances matrix is a diagonal matrix, which means that the regression equations are actually unrelated (no disturbances correlation). In this research, we use maximum likelihood estimators.

In proceeding specification we have a white noise ω in stochastic part of the function. To move on, we explicitly assume that:

(For the case of $j = 3$)

$$\boldsymbol{\omega} = \begin{bmatrix} \omega_{1it} \\ \omega_{2it} \\ \omega_{3it} \end{bmatrix} \sim N\left(0, \quad \Sigma\right), \text{ where } \Sigma = \begin{bmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{bmatrix} \tag{2-6}$$

Note that two additional conditions are also assumed:

$$|\rho_{jk}| < 1 \quad \forall \quad j, k = 1, 2, 3; \tag{2-7}$$

$$|\Sigma| > 0; \tag{2-8}$$

For assumption (2-7), it is very straightforward. For assumption (2-8), it is a necessary condition in the procedures of maximum likelihood estimation. Actually, if $|\rho_{jk}|$ is not very large, this condition will always hold. And in our case, we expect that the structural errors are correlated but not highly correlated, because identical explanatory variables have been using in variance function $h(z_{it}; \beta)$. Hence, the variance-covariance matrix for yields is given by:

$$\operatorname{var}\begin{bmatrix} y_{1it} \\ y_{2it} \\ y_{3it} \end{bmatrix} = \operatorname{var}\begin{bmatrix} u_{1it} \\ u_{2it} \\ u_{3it} \end{bmatrix} = \Omega = \begin{bmatrix} h_1 & \rho_{12}\sqrt{h_1 h_2} & \rho_{13}\sqrt{h_1 h_3} \\ \rho_{12}\sqrt{h_1 h_2} & h_2 & \rho_{23}\sqrt{h_2 h_3} \\ \rho_{13}\sqrt{h_1 h_3} & \rho_{23}\sqrt{h_2 h_3} & h_3 \end{bmatrix} \tag{2-9}$$

Note that (2-9) can also be written into products among different elements as following:

$$\Omega = \begin{bmatrix} \sqrt{h_{1it}} & 0 & 0 \\ 0 & \sqrt{h_{2it}} & 0 \\ 0 & 0 & \sqrt{h_{3it}} \end{bmatrix} \cdot \Sigma \cdot \begin{bmatrix} \sqrt{h_{1it}} & 0 & 0 \\ 0 & \sqrt{h_{2it}} & 0 \\ 0 & 0 & \sqrt{h_{3it}} \end{bmatrix} = H^{0.5} \cdot \Sigma \cdot H^{0.5}$$

Here $|H| = h_{1it} h_{2it} h_{3it}$

It is well known that the classic way to estimate variance - covariance matrix is to follow the first step of several asymptotically equivalent two-step GLS (generalized least squares) procedures (Schmidt, 1977; Hwang, 1990) which can give consistent estimators of variance-covariance matrix. However, the variance-covariance matrix in (2-6) is not exactly the matrix estimated by classic ways. Note that variance - covariance matrix (Σ) is only part of variance-covariance matrix (Ω) for yields. This decomposition enables us to focus only on unexplained heteroskedasticity and corresponding correlation induced by these unexplained errors. On the other hand, with model specification as (2-5), however, more efficient estimator can be achieved by one-step maximum likelihood estimation which can estimate both Σ and coefficients in deterministic and stochastic parts simultaneously. Therefore, by combining (2-4), (2-5), (2-6), (2-7), (2-8) and (2-9) together, we can get the log-likelihood function of this three-equation system:

$$\underset{\alpha,\beta,\Sigma}{Max} \ln L = \sum_{t=1}^{T} \sum_{i=1}^{N} \ln L_{it} \tag{2-10}$$

Where,

$$\ln L_{it} = -\frac{3}{2}\cdot\ln(2\pi) - \frac{1}{2}\ln|\Omega| - \frac{1}{2}\begin{pmatrix} y_{1it} - f_1(x_{1it};\alpha_1) \\ y_{2it} - f_2(x_{2it};\alpha_2) \\ y_{3it} - f_3(x_{3it};\alpha_3) \end{pmatrix}' \Omega^{-1} \begin{pmatrix} y_{1it} - f_1(x_{1it};\alpha_1) \\ y_{2it} - f_2(x_{2it};\alpha_2) \\ y_{3it} - f_3(x_{3it};\alpha_3) \end{pmatrix}$$

And,

$$\begin{aligned} \ln|\Omega| &= \ln\left|H^{0.5}\right| + \ln|\Sigma| + \ln\left|H^{0.5}\right| \\ &= \ln(h_{1it}) + \ln(h_{2it}) + \ln(h_{3it}) + \ln|\Sigma| \end{aligned}$$

Finally, we have the log-likelihood function as:

$$\begin{aligned} \underset{\alpha,\beta,\Sigma}{Max} \ln L &= \sum_{t=1}^{T}\sum_{i=1}^{N} \ln L_{it} = -\frac{3NT}{2}\cdot\ln(2\pi) \\ &-\sum_{t=1}^{T}\sum_{i=1}^{N}\ln h_{1it} - \sum_{t=1}^{T}\sum_{i=1}^{N}\ln h_{2it} - \sum_{t=1}^{T}\sum_{i=1}^{N}\ln h_{3it} - \frac{NT}{2}\ln|\Sigma| \\ &-\frac{1}{2}\sum_{t=1}^{T}\sum_{i=1}^{N}\begin{bmatrix} (y_{1it} - f_1(x_{1it};\alpha_1))/\sqrt{h_{1it}} \\ (y_{2it} - f_2(x_{2it};\alpha_2))/\sqrt{h_{2it}} \\ (y_{3it} - f_3(x_{3it};\alpha_3))/\sqrt{h_{3it}} \end{bmatrix}' \Sigma^{-1} \begin{bmatrix} (y_{1it} - f_1(x_{1it};\alpha_1))/\sqrt{h_{1it}} \\ (y_{2it} - f_2(x_{2it};\alpha_2))/\sqrt{h_{2it}} \\ (y_{3it} - f_3(x_{3it};\alpha_3))/\sqrt{h_{3it}} \end{bmatrix} \end{aligned} \tag{2-11}$$

To implement the estimation on this log-likelihood function, nonlinear optimization procedure will be used and here we choose Double-Dogleg Method and Newton-Raphson method. For the data, as we discussed early in this section, in the case (2-5-2) of SUR by regions (for individual crop) we deal with the panel nature of data. Otherwise, in the case (2-5-1) of SUR by crops (both for single region and all regions) we pooled the entire dataset by regions and years together. The data characteristics and descriptive statistics will be presented in next chapter.

2.5 SUR Model with Identical Regressors

Before we proceed on data analysis and estimation, it is also necessary to clarify the case of SUR model with identical regressors. This is the most crucial framework in this research (case 2-5-1), and in some situations it tends out that The OLS and FGLS estimators are identical and, therefore, equivalent to the maximum likelihood estimator of coefficients (α in (5-2)). Srivastava and Giles (1987) give a case that the same explanatory variables appear in each of the equations of SUR model. While there is no heteroskedasticity in each equation, they prove that SUR model collapses to the multivariate regression model, and hence OLS and FGLS estimators are identical:

$X_1 = X_2 = \cdots = X_M = Z$ (M=number of equations)
$\mathrm{K}_1 = \mathrm{K}_2 = \cdots = \mathrm{K}_M = K$ (K=number of parameters in each of M equations)
$\mathrm{T}_1 = \mathrm{T}_2 = \cdots = \mathrm{T}_M = T$ (T=number of observations in each of M equations)

$$X = \begin{pmatrix} Z & 0 & \cdots & 0 \\ 0 & Z & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & Z \end{pmatrix} = (I_M \otimes Z) \tag{2-12}$$

$$\begin{aligned} \beta_{OLS} &= (X^{'}X)^{-1}X^{'}Y \\ &= (I_M \otimes Z^{'}Z)^{-1}(I_M \otimes Z^{'})Y \\ &= [I_M \otimes (Z^{'}Z)^{-1}Z^{'}]Y \end{aligned} \tag{2-13}$$

$$\begin{aligned} \beta_{GLS} &= [(I_M \otimes Z^{'})(\hat{\Omega}^{-1} \otimes I_T)(I_M \otimes Z)]^{-1}(I_M \otimes Z^{'})(\hat{\Omega}^{-1} \otimes I_T)Y \\ &= (\hat{\Omega}^{-1} \otimes Z^{'}Z)^{-1}(\hat{\Omega}^{-1} \otimes Z^{'})Y \\ &= [I_M \otimes (Z^{'}Z)^{-1}Z^{'}]Y \end{aligned} \tag{2-14}$$

($\hat{\Omega}$ is the estimate of variance-covariance matrix)

In our specification, however, we have heteroskedasticity in each equation. As it indicates in maximum likelihood function (2-11), we can divide each equation by the corresponding square root of heteroskedastic variance $\sqrt{h_{jit}}$. Then as a result of this transformation, the SUR model with identical regressors becomes a SUR model with different regressors (2-15), in which case OLS estimators is not efficient (Kmenta, 1986).

$$X = \begin{pmatrix} Z & 0 & \cdots & 0 \\ 0 & Z & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & Z \end{pmatrix} = (I_M \otimes Z) \Rightarrow X = \begin{pmatrix} Z\big/\sqrt{h_{1it}} & 0 & \cdots & 0 \\ 0 & Z\big/\sqrt{h_{2it}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & Z\big/\sqrt{h_{Mit}} \end{pmatrix} \quad (2\text{-}15)$$

Therefore, in the situation of a system of equations with identical regressors, if the following two conditions exist then SUR methods are more efficient: (1) there is disturbance correlation among different equations; (2) heteroskedasticity is the appropriate specification for each equation. Our model meets both conditions.

Chapter 3: Data and Descriptive Statistics

Applied research is designed to help solve particular existing problems and advance people’s knowledge and understanding of these issues. While methodology is the framework of research, data serves as the foundation of that research. The characteristics of the data, particularly their type, quantity, and sampling methods, constrain the choice of analytical techniques applied to the data. So in this chapter we will be discussing data collection and coverage, characteristics of data, missing values and descriptive statistics.

3.1. Crop Data

Agriculture is important to Arizona. According to the National Agricultural Statistical Service of USDA, in 2007, there were 10,000 farms operating in Arizona, the average farm size was about 2,600 acres. But not all farmers grow the same crops; farmers in Arizona grow almost 100 different crops. Among these crops, lettuce, hay, cotton, cantaloups, wheat, broccoli, watermelons are leading crops for the state’s farm cash receipts. These leading crops account for more than 80% of the state’s farm cash receipts (table 3.1).

The top three agricultural crop commodities in Arizona are lettuce, cotton and hay. They represent 10.8%, 5.4% and 7.5% of the state’s total farm cash receipts (2008) respectively. Arizona grows enough cotton each year to make more than one pair of jeans for every person in the United States. Arizona hay alfalfa yield led the nation at 8.3 tons per acre compared to 3.4 tons nationally. So in this research we focus on two of the top three crops, hay and cotton. The raw data on annual county-level crop yields comes from the National Agricultural Statistics Service (USDA) online database.

Table 3.1 Arizona: Leading commodities for cash receipts, 2007-2008

Items	Value of receipts (1,000 dollars)		Percent of Total Receipts		Cumulative Percent	
Year	2007	2008	2007	2008	2007	2008
All commodities	3,420,496	3,464,560	100.0	100.0	--	--
Livestock and products	1,577,007	1,505,488	46.1	43.5	--	--
Crops	1,843,487	1,959,072	53.9	56.5	--	--
Dairy products	801,627	763,136	23.4	22.0	23.4	22.0
Cattle and calves	677,393	637,016	19.8	18.4	43.2	40.4
Greenhouse/nursery	79,125	417,748	2.3	12.1	45.5	52.5
Lettuce	560,160	373,290	16.4	10.8	61.9	63.2
Hay	195,177	260,926	5.7	7.5	67.6	70.8
Cotton	169,894	188,797	5.0	5.4	72.6	76.2
Cantaloups	103,880	118,825	3.0	3.4	75.6	79.7
Wheat	40,196	118,808	1.2	3.4	76.8	83.1
Broccoli	67,444	57,988	2.0	1.7	78.8	84.8
Watermelons	41,016	46,656	1.2	1.3	80.0	86.1

Source: Economic Research Service/USDA

Another issue associated with crop data is which county and how many counties should be included in this research. There are 15 counties in the state of Arizona. During the time period 1965-2008, almost all of the counties have been geographically stable except that La Paz County was established in 1983 after voters approved separating the northern portion of Yuma County, making it the first and only new county created since Arizona statehood in 1912. As a result, Arizona laws were changed to make splitting other existing counties much more difficult. On the other hand, it happens that La Paz is one of the leading agricultural counties in Arizona. According to the Arizona Agricultural Statistics, La Paz is among the top five counties in agricultural sales in 2008 (table 3.2). And the observations from Yuma County are also affected by the discontinuity of La Paz County during the time period being studied. To remedy this, we combine Yuma County and La Paz County together as a joint county Yuma-La Paz. Since counties in northern Arizona and other small counties (Greenlee and Santa Cruz County) only count for a very small percent of state receipts, and statistical data is not continuously available for these counties, we exclude them to reduce measurement error and the complexity of analysis. In the end, we include 7 counties in this research, and Yuma and La Paz are combined into one county for the reason discussed above. These counties are Yuma-La Paz, Pinal, Pima, Maricopa, Graham, and Cochise, which in total account for 94.91% of the state cash receipt.

By considering crop choice and county choice jointly, the alternative way is looking at the production value of cotton and hay in those seven leading counties (table 3.3). For cotton upland, seven leading counties sum up to 96.89% and 96.44% of state total production in 2007 and 2008 respectively. For hay all, these percentage are 86.58 (data of Cochise and Pima county is missing in 2007) and 95.46%. Therefore, the data collection in this research covers most of the crop farming activities in Arizona, which ensures the applicability of this research.

Table 3.2 Cash Receipts: All Farm Commodities by County (2008)

	Percent of state total receipts	Thousands $
1. Yuma County	28.88%	1,000,578
2. Pinal County	25.52%	884,175
3. Maricopa County	25.43%	881,115
4. Graham County	5.37%	186,143
5. La Paz County	4.05%	140,174
6. Cochise County	3.60%	124,873
7. Pima County	2.06%	71,209
Sub Total	94.91%	3,288,267
State total	100%	3,464,560

Source: Arizona Agricultural Statistics (2008).

Table 3.3 Production of Cotton and Hay by Leading Counties (2007-2008)

County	Cotton Upland (bales)				Hay All (tons)			
	2007		2008		2007		2008	
	Production	Percent	Production	Percent	Production	Percent	Production	Percent
Cochise	--	--	--	--	--	--	153,500	6.44%
Graham	60,000	11.67%	52,100	12.86%	16,500	0.75%	14,300	0.60%
Maricopa	89,000	17.32%	56,500	13.95%	658,000	30.02%	777,000	32.61%
Pima	22,000	4.28%	19,000	4.69%	--	--	23,000	0.97%
Pinal	238,000	46.30%	202,700	50.05%	455,500	20.78%	512,000	21.49%
Yuma	51,000	9.92%	29,000	7.16%	289,000	13.18%	293,500	12.32%
La Paz	38,000	7.39%	31,300	7.73%	478,800	21.84%	501,500	21.04%
Sub Total	498,000	96.89%	390,600	96.44%	1,897,800	86.58%	2,274,800	95.46%
State Total	514,000	100%	405,000	100%	2,192,000	100%	2,383,000	100%

Source: Economic Research Service/USDA

3.2. Climatic Data

Arizona is on the western end of the Rocky Mountain chain and the northern half of the state is mountainous. So the main agricultural area is the southern third of the state (except for Mohave County, which is near the Colorado River). The southern half of the state is mainly desert and is good for year round crop growth in irrigated areas. The irrigation rate in Arizona agricultural area is generally high, most of years the irrigation rate is higher than 70% (table 3.4). For some special crops, all of the planted and harvested area is irrigated. With Arizona’s diverse topography, the temperature for southern Arizona, including the Phoenix metropolitan area, is lows of 30°F in the winter to over 100°F in the summer. In the northern portion of the state, the temperature fluctuates from 20°F to 95°F. The highest temperature recorded was 127°F. The lowest temperature recorded was -40°F. A major climate feature of the U.S. Southwest is the North American monsoon (a distinct seasonal change in wind direction of at least 120°). Arizona receives a majority of it’s rainfall during this late summer period. There are, on average, 257 clear, sunny days with an average rainfall of 12.7 inch a year giving Arizona very low relative humidity (NOAA).

Table 3.4 Farm Characteristics of Arizona

1997, 2002 and 2007 Census of Agriculture			
	1997	2002	2007
Approximate total land area (acres)	72,731,030	72,726,122	72,696,492
Total farmland (acres)	27,169,627	26,586,577	26,117,899
Percent of total land area	37.4%	36.6%	35.9%
Cropland (acres)	1,354,820	1,261,894	1,205,425
Percent of total farmland	5.0%	4.7%	4.6%
Percent irrigated	**75.3%**	**70.4%**	**68.3%**
Harvested Cropland (acres)	1,026,359	887,966	832,406

Source: USDA, National Agricultural Statistics Service
http://www.agcensus.usda.gov/Publications/

Climatic data comes from Western Regional Climate Center. The climatic dataset includes two measures: temperature and precipitation. Both are monthly average observations on the county level, by using these monthly measures we calculated four main explanatory variables (table 3.5): yearly average temperature (T), yearly cumulative precipitation (P), within year temperature variance (VT) and within year precipitation variance (VP). To sum up, the dataset contains data for cotton yield (cotton upland, 1935-2008) and hay yield (hay all, 1965-2008) as dependent variables; time trend and climatic variables as explanatory variables. The descriptive statistics and observation curves of these variables will be shown in section 3.4. To match the specification and estimation procedure of the SUR model, it is convenient to maintain a balanced dataset. A balanced dataset requires that, however, we have to deal with missing values in our case, which we are going to discuss in details in next section. Finally, we end up with a new balanced dataset which includes two main crops (cotton upland and hay all) from 7 leading agricultural counties (Cochise, Graham, Maricopa, Pima, Pinal, Yuma-Lapaz) in Arizona for the time period of 1965-2008 (table 3.5).

3.3. Missing Values

As mentioned in last section, to maintain a balanced dataset, we have to deal with the missing values in the raw data on crop yield. 17 out of 528 observations for two crops across six counties were missing, which is about 3% of the entire dataset. The reason for these missing values includes: estimates too small to warrant quantitative estimate or not published to avoid disclosure of individual operations. For these missing values, a strategy to fill them is simple imputation, which substitutes a value for each missing value. Standard statistical procedures for complete data analysis can then be used with the filled-in data set (Rubin, 1976, 1987). In our dataset, for simplicity, we employ two procedures to fill these missing values: mean filling and filling by percent increase. Another reason for choosing mean filling and percent increase filling rather than random imputation is that, we believe these values are missing not because of random factors but anthropogenic factors. So these missing values could have been consistent with observed values from other counties.

In the case there is only one missing value, such as hay yield for Cochise County in 1996 and 2007, we simply use the average of the values from the previous and the next years. In the case there are missing values for continuous years, such as cotton yield for Cochise County from 2005-2008, we use filling by percent increase. The principle of this method is that, while we have continuous missing values for a certain county we have observations for other five counties during the same time period. This means we can use information from other counties in the same area (southern Arizona) to infer the pattern of continuous missing values in that particular county. One way to formulate and capture this pattern is to calculate the percent increase of yield by year from other counties, and then we implement these percent increases on the particular county with continuous missing values. By following this procedure, we assume that the yield increase patterns for that particular time period with missing values in all counties are similar. This assumption makes sense when all of these counties are involved in the same agricultural area, which is the case in this research.

3.4. Descriptive Statistics

Before we proceed with estimation and analysis of results, it is necessary to get an overall impression of the data we are using by looking at the descriptive statistics of the data. The table 3.5 shows the descriptive statistics of crop yield and climatic measures which we are going to use as dependent and independent variables in estimation.

Table 3.5 Descriptive Statistics of Crop Yield and Climatic Data (all counties)

Variable (unit)	N	Mean	Std Dev	Minimum	Maximum
Cotton yield (lb/acre)	264	1079.82	255.66	435.00	1686.00
Hay yield (ton/acre)	264	6.37	1.26	2.67	9.16
Year	44	1986.50	12.72	1965.00	2008.00
P (inch)	264	11.64	4.71	1.22	25.73
VP (inch*inch)	264	1.12	0.83	0.04	4.34
T (F)	264	66.36	4.85	57.19	74.16
VT (F*F)	264	190.35	23.05	136.06	242.25
P*T (F*inch)	264	756.80	281.29	89.43	1538.54
P^2 (inch^2)	264	157.64	115.59	1.49	662.03
T^2 (F^2)	264	4427.12	634.77	3270.89	5499.46

Data Source: (1) National Agricultural Statistics Service, USDA.
(2) Western Regional Climate Center, United States.

Another way to look at these variables is to draw the scatter-line chart of them. Since basically the data we are using in this research has both the nature of being cross section and time series, so scatter-line chart can show the time trend and cross section heterogeneity more visually and directly (figure 3.1-3.6). From these charts, following basic patterns can be observation:

(1) During sample period, average temperature has been rising at a slow rate

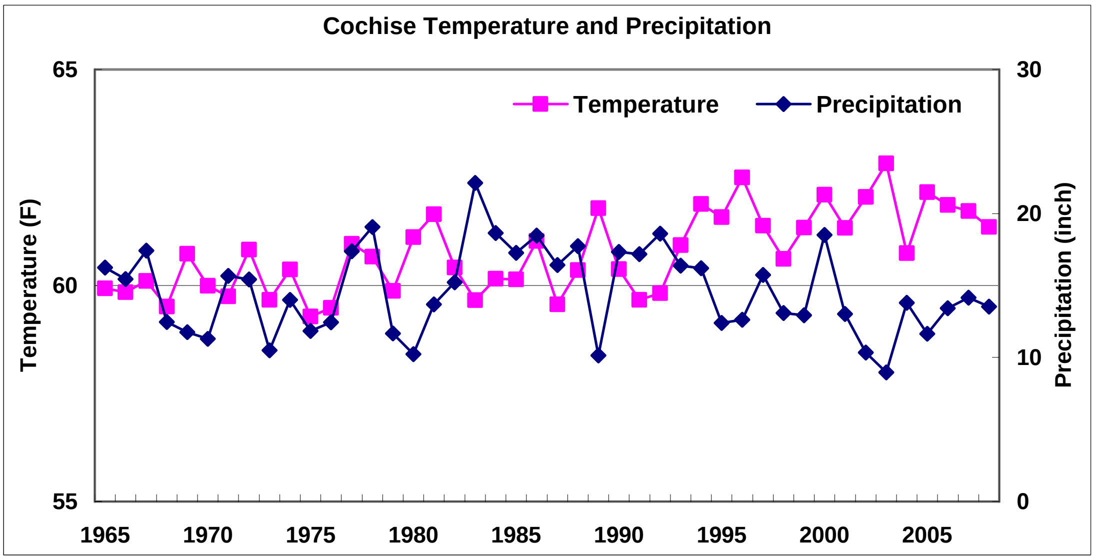
consistently.

(2) During sample period, yearly cumulative precipitation has not changed very much, but since 1980s there has been a decline in precipitation.

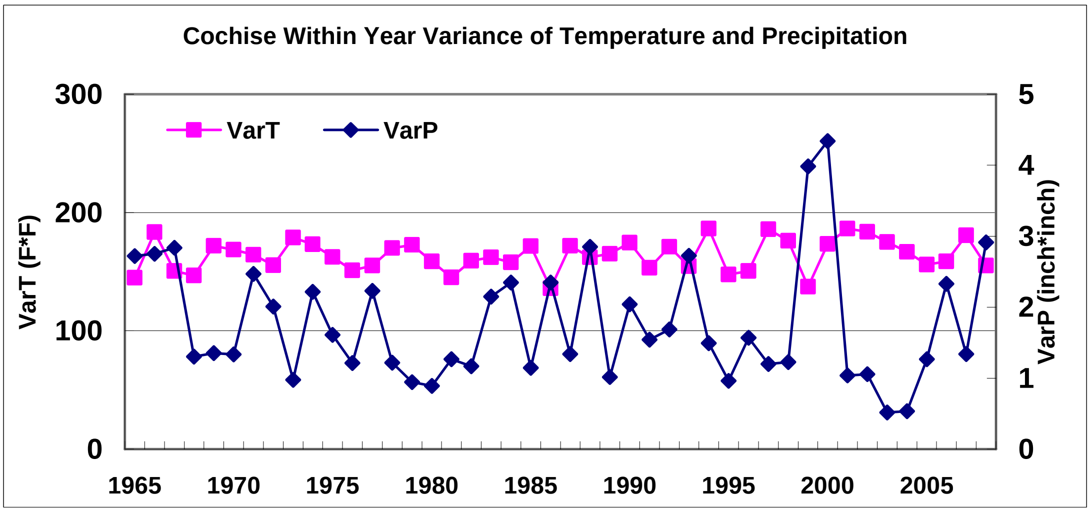
(3) The within-year variance of temperature and precipitation can come from two sources: seasonality and irregular climatic change and shocks, which are both important to crop yield.

(4) The within year variance of temperature and precipitation have very significant fluctuation from year to year. For within year variance of precipitation, it is evident that most of time higher within year variance is associated with a higher observation on precipitation.

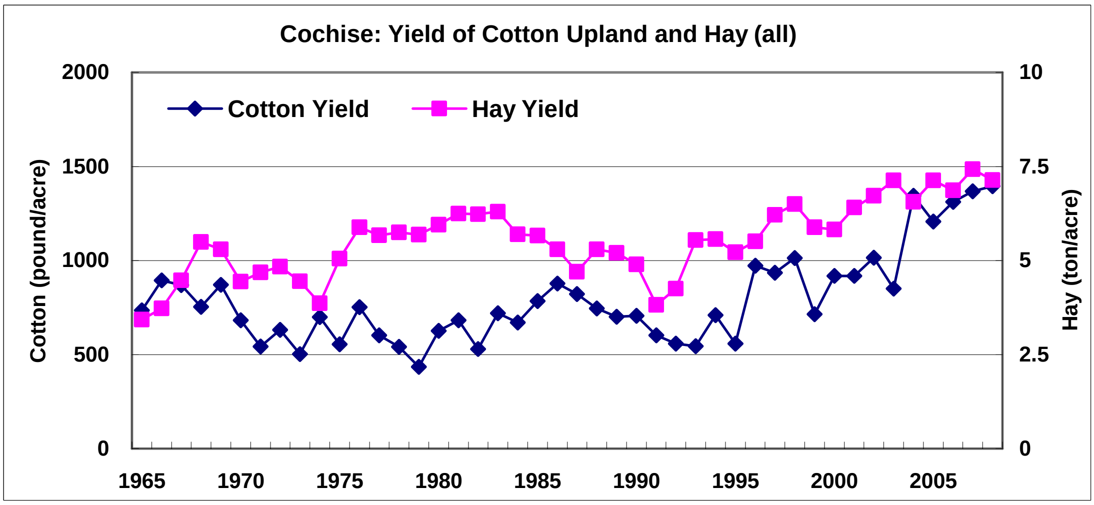
(5) During 1965-2008, both cotton and hay have experienced a yield increase while there also exists significant yield fluctuation.


Cochise Temperature and Precipitation
Temperature
Precipitation
Temperature (F)
Precipitation (inch)
65
60
55
30
20
10
0
1965
1970
1975
1980
1985
1990
1995
2000
2005

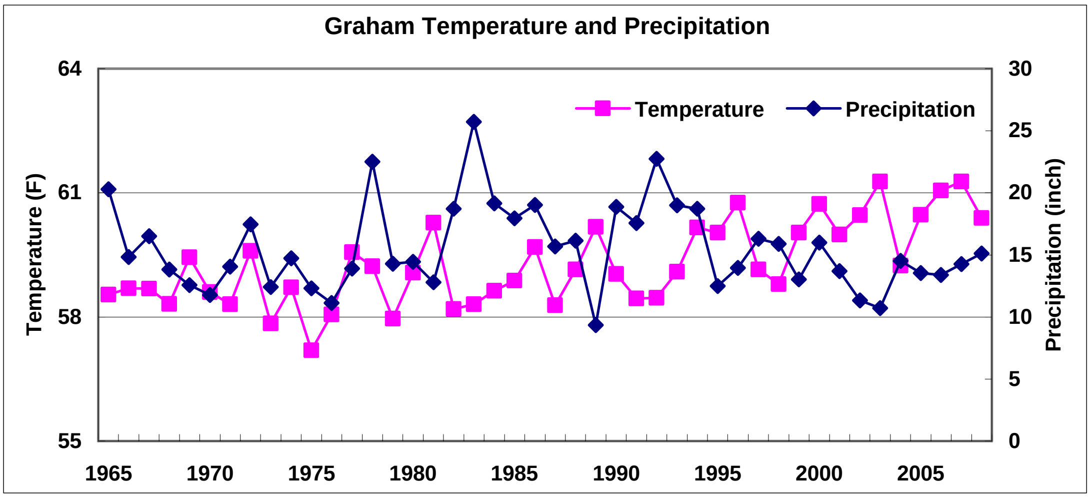
3-1(a)


Cochise Within Year Variance of Temperature and Precipitation
VarT
VarP
VarT (F*F)
VarP (inch*inch)
300
200
100
0
5
4
3
2
1
0
1965
1970
1975
1980
1985
1990
1995
2000
2005

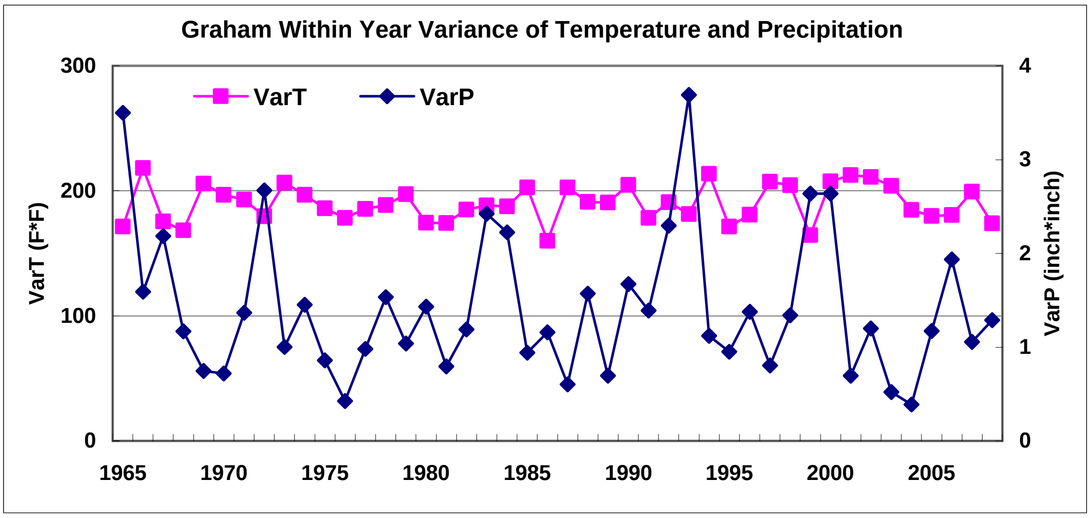
3-1 (b)


Cochise: Yield of Cotton Upland and Hay (all)
Cotton Yield
Hay Yield
Cotton (pound/acre)
Hay (ton/acre)
2000
1500
1000
500
0
10
7.5
5
2.5
0
1965
1970
1975
1980
1985
1990
1995
2000
2005

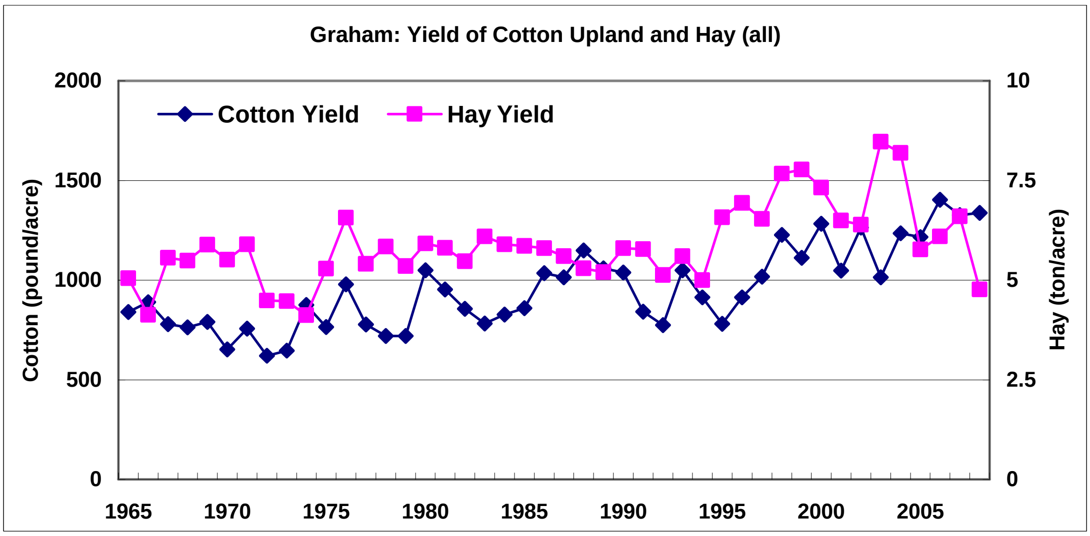
3-1 (c)


Graham Temperature and Precipitation
Temperature
Precipitation
Temperature (F)
Precipitation (inch)
64
61
58
55
30
25
20
15
10
5
0
1965
1970
1975
1980
1985
1990
1995
2000
2005

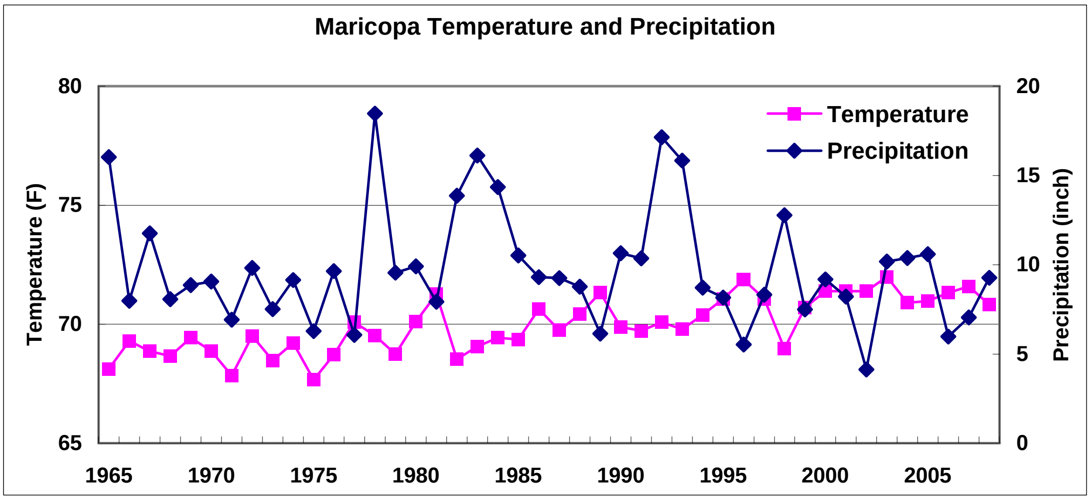
3-2 (a)


Graham Within Year Variance of Temperature and Precipitation
VarT
VarP
VarT (F*F)
VarP (inch*inch)
300
200
100
0
4
3
2
1
0
1965
1970
1975
1980
1985
1990
1995
2000
2005

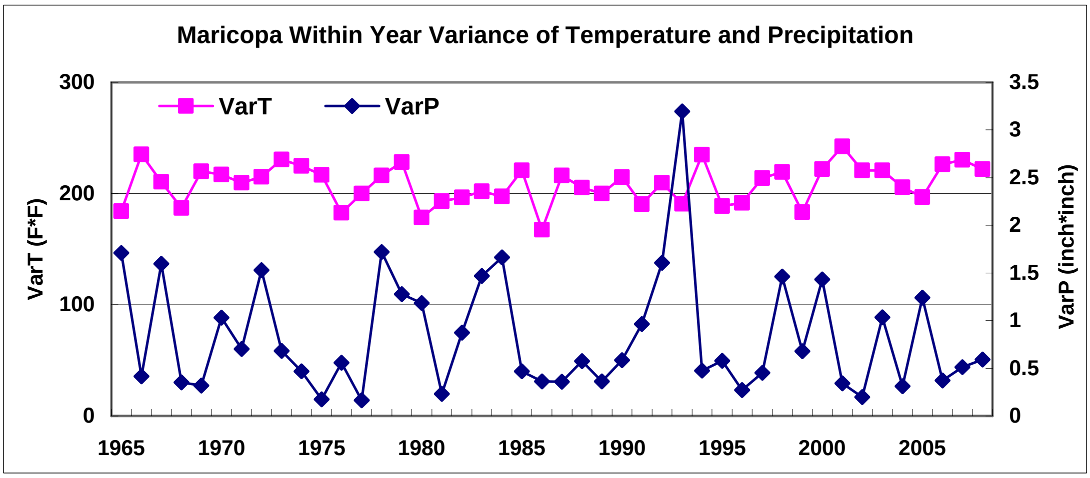
3-2 (b)


Graham: Yield of Cotton Upland and Hay (all)
Cotton Yield
Hay Yield
Cotton (pound/acre)
Hay (ton/acre)
2000
1500
1000
500
0
10
7.5
5
2.5
0
1965
1970
1975
1980
1985
1990
1995
2000
2005

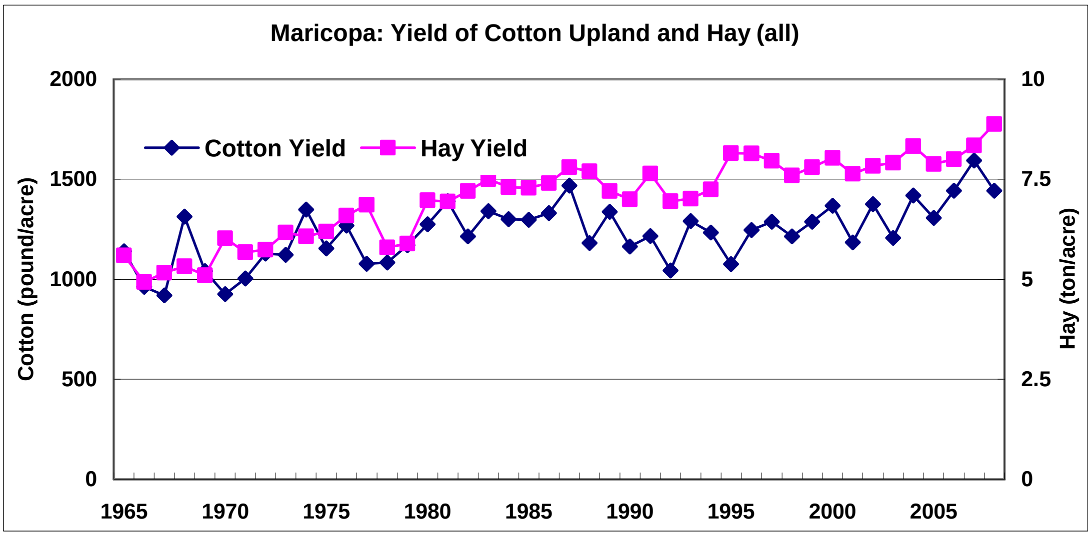
3-2 (c)


Maricopa Temperature and Precipitation
Temperature
Precipitation
Temperature (F)
Precipitation (inch)
80
75
70
65
20
15
10
5
0
1965
1970
1975
1980
1985
1990
1995
2000
2005

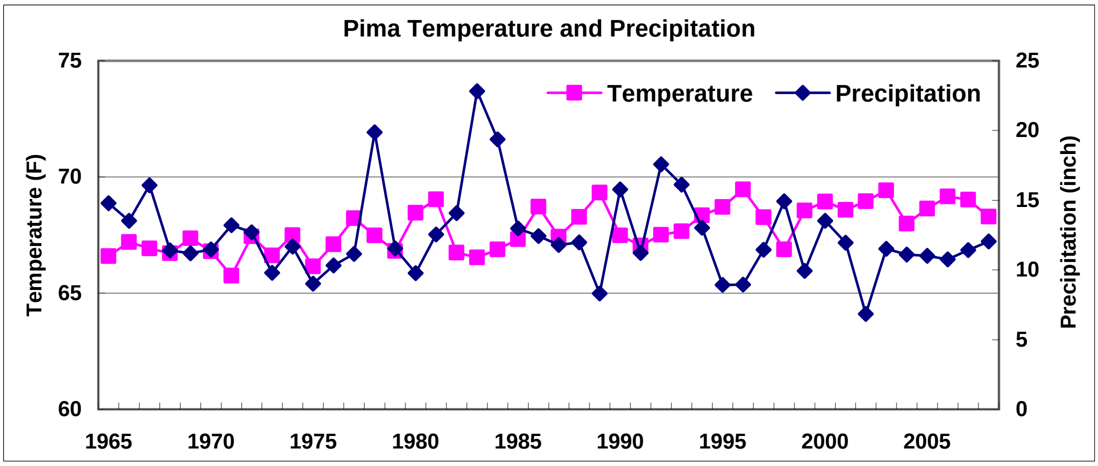
3-3 (a)


Maricopa Within Year Variance of Temperature and Precipitation
VarT
VarP
VarT (F*F)
VarP (inch*inch)
300
200
100
0
3.5
3
2.5
2
1.5
1
0.5
0
1965
1970
1975
1980
1985
1990
1995
2000
2005

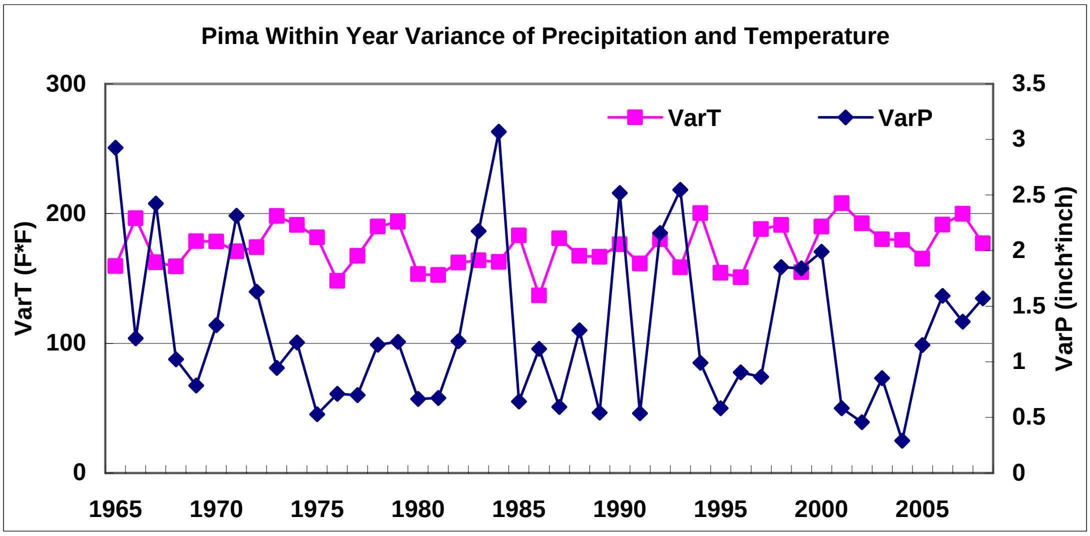
3-3 (b)


Maricopa: Yield of Cotton Upland and Hay (all)
Cotton Yield
Hay Yield
Cotton (pound/acre)
Hay (ton/acre)
2000
1500
1000
500
0
10
7.5
5
2.5
0
1965
1970
1975
1980
1985
1990
1995
2000
2005

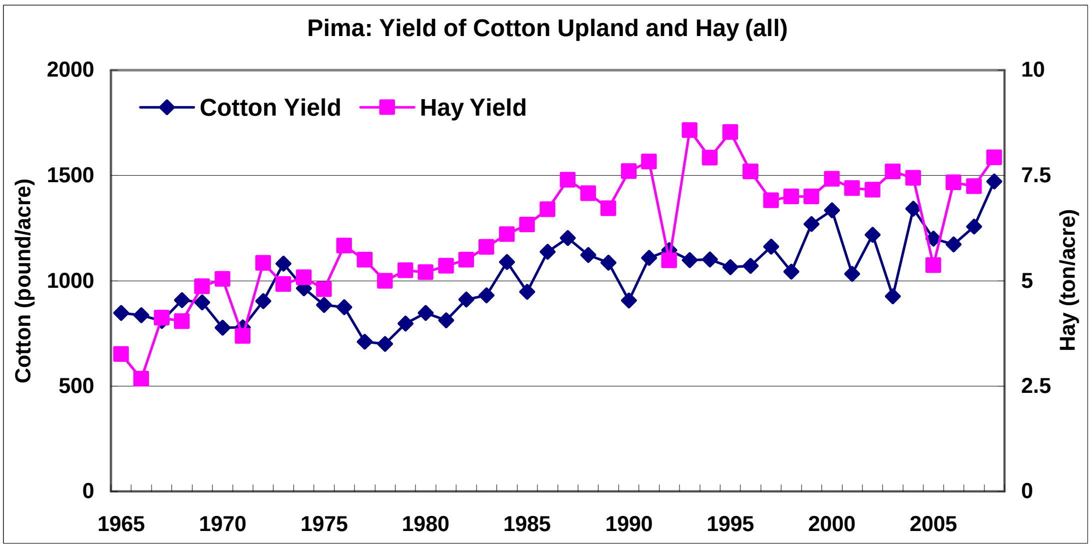
3-3 (c)


Pima Temperature and Precipitation
Temperature
Precipitation
Temperature (F)
Precipitation (inch)
75
70
65
60
25
20
15
10
5
0
1965
1970
1975
1980
1985
1990
1995
2000
2005

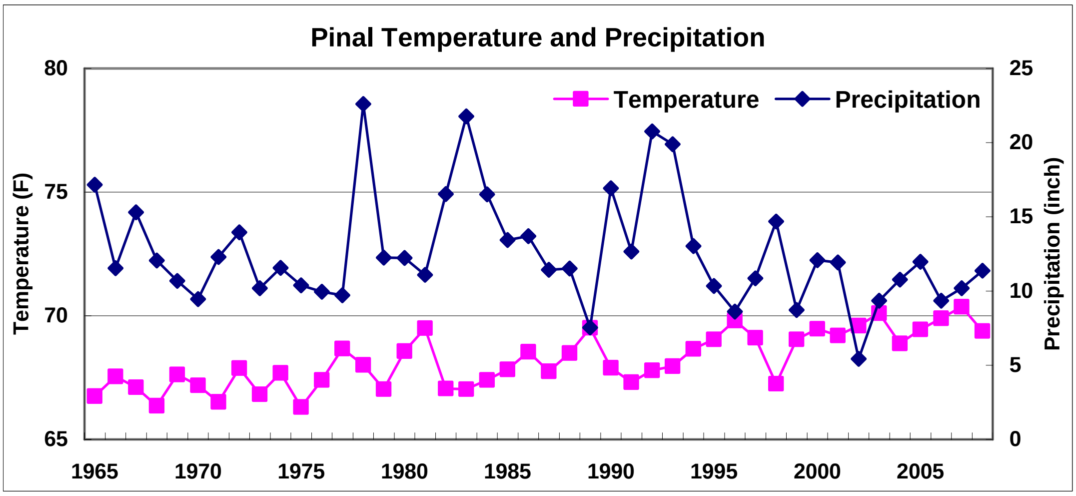
3-4 (a)


Pima Within Year Variance of Precipitation and Temperature
VarT
VarP
VarT (F*F)
VarP (inch*inch)
300
200
100
0
3.5
3
2.5
2
1.5
1
0.5
0
1965
1970
1975
1980
1985
1990
1995
2000
2005

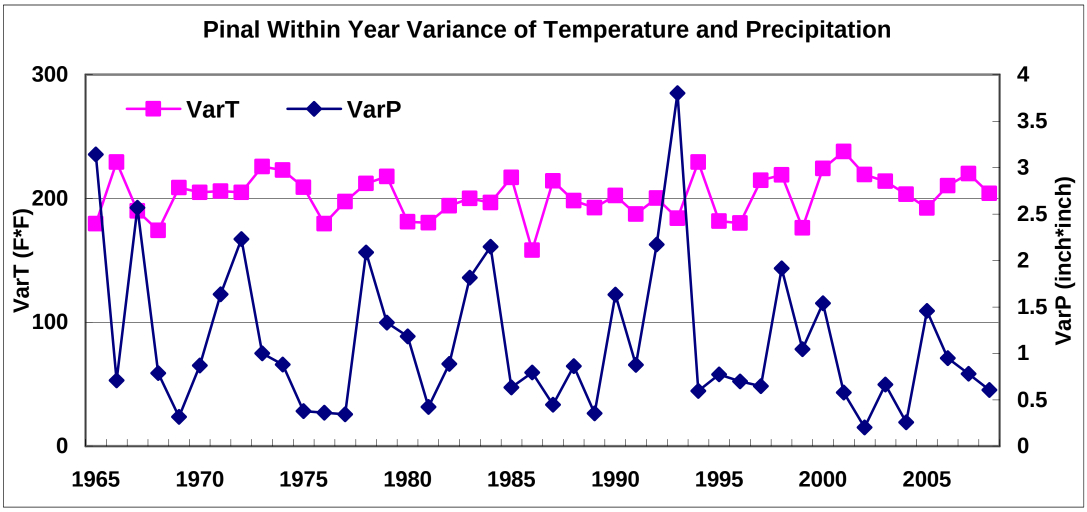
3-4 (b)


Pima: Yield of Cotton Upland and Hay (all)
Cotton Yield
Hay Yield
Cotton (pound/acre)
Hay (ton/acre)
2000
1500
1000
500
0
10
7.5
5
2.5
0
1965
1970
1975
1980
1985
1990
1995
2000
2005

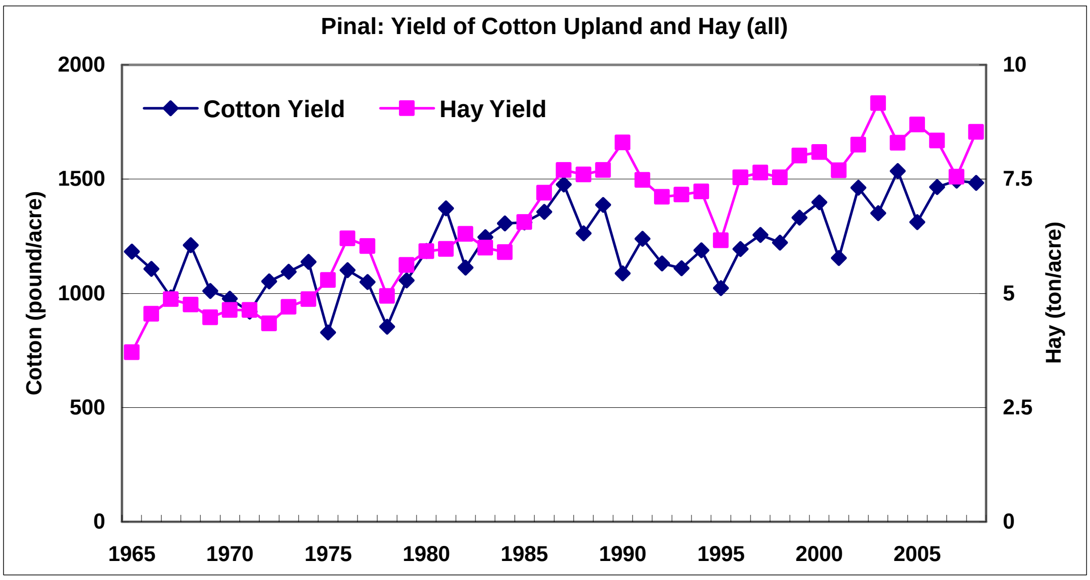
3-4 (c)


Pinal Temperature and Precipitation
Temperature
Precipitation
Temperature (F)
Precipitation (inch)
80
75
70
65
25
20
15
10
5
0
1965
1970
1975
1980
1985
1990
1995
2000
2005

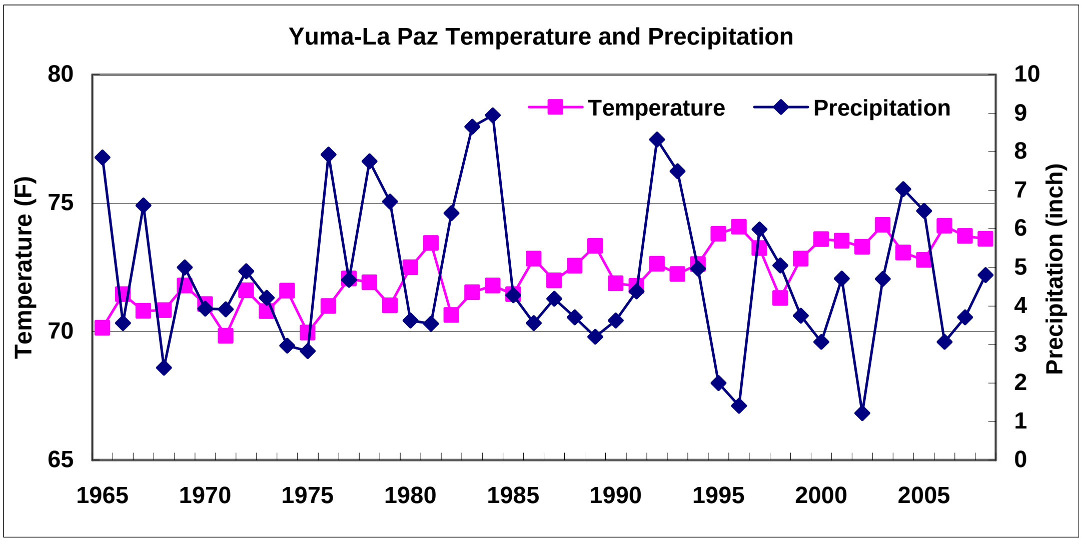
3-5 (a)


Pinal Within Year Variance of Temperature and Precipitation
VarT
VarP
VarT (F*F)
VarP (inch*inch)
300
200
100
0
4
3.5
3
2.5
2
1.5
1
0.5
0
1965
1970
1975
1980
1985
1990
1995
2000
2005

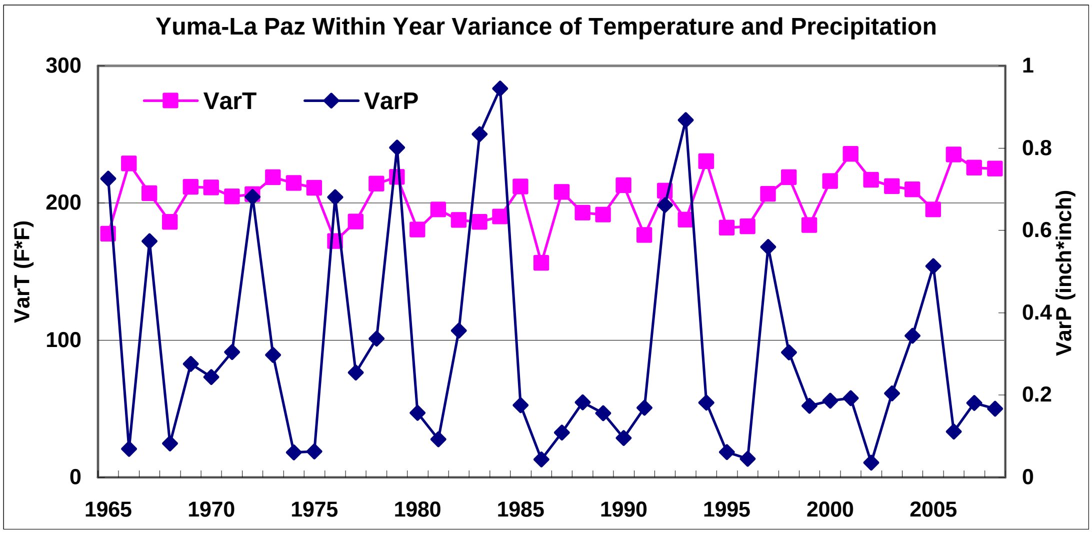
3-5 (b)


Pinal: Yield of Cotton Upland and Hay (all)
Cotton Yield
Hay Yield
Cotton (pound/acre)
Hay (ton/acre)
2000
1500
1000
500
0
10
7.5
5
2.5
0
1965
1970
1975
1980
1985
1990
1995
2000
2005

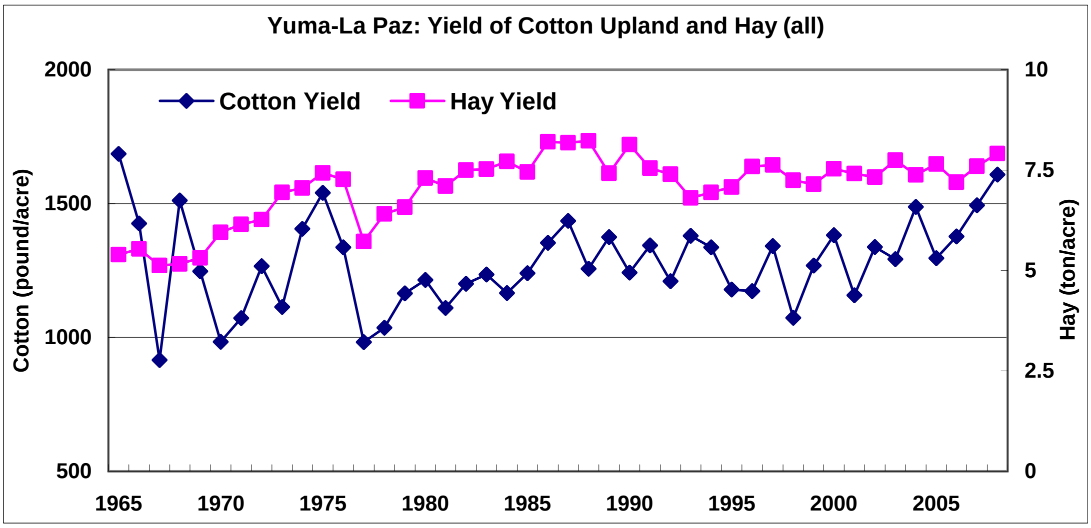
3-5 (c)


Yuma-La Paz Temperature and Precipitation
Temperature
Precipitation
Temperature (F)
Precipitation (inch)
80
75
70
65
10
9
8
7
6
5
4
3
2
1
0
1965
1970
1975
1980
1985
1990
1995
2000
2005

3-6 (a)


Yuma-La Paz Within Year Variance of Temperature and Precipitation
VarT
VarP
VarT (F*F)
VarP (inch*inch)
300
200
100
0
1
0.8
0.6
0.4
0.2
0
1965
1970
1975
1980
1985
1990
1995
2000
2005

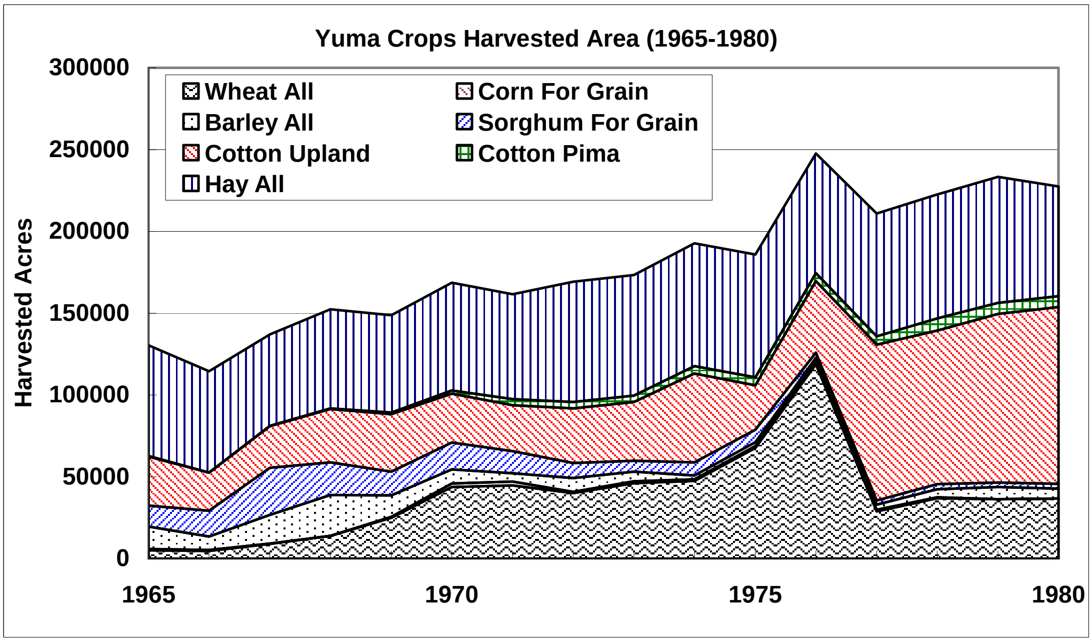
3-6 (b)


Yuma-La Paz: Yield of Cotton Upland and Hay (all)
Cotton Yield
Hay Yield
Cotton (pound/acre)
Hay (ton/acre)
2000
1500
1000
500
10
7.5
5
2.5
0
1965
1970
1975
1980
1985
1990
1995
2000
2005

3-6 (c)

3.5 Land Use Change in Yuma

Aside from the missing values we discussed in section 3.3, another particular concern for this research is land use change. Land use change could be an important source of yield fluctuation. Unlike technology-induced yield change, land use change-induced yield change can be abrupt mutation since the productivity of land has experienced essential change. Therefore, if there is land use changed involved during the period being studied, and then it is difficult to tell the difference of impacts from climatic change and land use change. After a careful examination on the dataset, it is obvious that Yuma County has experienced a big land use change during the mid of 1970s (figure 3-7). As the graph indicates, after 1976, the harvested area of cotton upland had jumped up while the harvested area of wheat had been reduced. The harvested area of hay had not changed much during this period. This kind of observed pattern in the dataset requires us to take special treatment in estimation.



3-7: Yuma Crops Harvested Area (1965-1980)

As a result of land use change, and which affects the land productivity directly, there is a quick yield decline after 1976 (figure 2-6(c)). The high yield fluctuation from 1965 to 1975 can also be a sign of land use change (the incubation for the change). There are two possible explanations for the yield decline after 1976. First, the new land used for cotton is of low productivity, which can level down the average productivity of all cotton cropland. Another explanation is that, the new land used for cotton had been used for some other crops which require less water before the land use change. So the irrigation system of the new included land has not been well developed, while cotton cropping is very demanding on irrigation resources. From figure 3-7, we can see that there is a land use change between wheat and cotton upland. And generally wheat cropping requires less irrigation resources than cotton cropping.

Chapter 4: Empirical Results

4.1 Tests of Specification and Structure Change

Before we proceed to model estimation, two questions are investigated. First, for SUR model by regions is it enough to implement a single overall estimation for all included counties or do we need to estimate separate sets of parameters for each county? On the other words, are the estimates for different counties significantly different, and different impact of climatic conditions on crop yields? Since we identified there is land use change in Yuma County, another question rises as, is this change significant enough to affect crop yield? In this regard, statistic tests are necessary to answer these two questions.

Table 4.1 Likelihood Ratio Test on SUR Model with 6 counties

Model		Likelihood Value	Number of Estimates
Unrestricted Model	Cochise	-502.56	29
	Graham	-490.29	29
	Maricopa	-416.95	29
	Pima	-475.32	29
	Pinal	-448.40	29
	Yuma La Paz	-479.03	29
Restricted Model	All Counties	-3137.34	29
Number of Restrictions	145		
Test Statistics (Z)	19.04		
Decision	Reject Null Hypotheses		

Table 4.2 Likelihood Ratio Test on SUR Model with 5 counties

Model		Likelihood Value	Number of Estimates
Unrestricted Model	Cochise	-502.56	29
	Graham	-490.29	29
	Maricopa	-416.95	29
	Pima	-475.32	29
	Pinal	-448.40	29
Restricted Model	5 Counties	-2577.23	29
Number of Restrictions	116		
Test Statistics (Z)	16.02		
Decision	Reject Null Hypotheses		

For the first question, we run a likelihood ratio test. The restricted model is the overall

SUR model for all counties, and the unrestricted models include six SUR models for each county. We also implement the likelihood ratio test for other five counties without Yuma, since we have already realized that Yuma County has a unique data pattern because of land use change. In these two likelihood ratio tests, the numbers of restrictions are 145 and 116 respectively. And in this case (degree of freedom is larger than 100) chi-square test can be transformed into a standardized normal distribution test:

$Z = \sqrt{2\chi^2} - \sqrt{(2k-1)}$ (k is the number of restrictions)

And the null hypotheses are:

(1) $\begin{cases} \alpha_{all} = \alpha_{cochise} = \alpha_{graham} = \alpha_{maricopa} = \alpha_{pima} = \alpha_{pinal} = \alpha_{yumalapaz} \\ \beta_{all} = \beta_{cochise} = \beta_{graham} = \beta_{maricopa} = \beta_{pima} = \beta_{pinal} = \beta_{yumalapaz} \end{cases}$ (For all counties model)

(2) $\begin{cases} \alpha_{5\,counties} = \alpha_{cochise} = \alpha_{graham} = \alpha_{maricopa} = \alpha_{pima} = \alpha_{pinal} \\ \beta_{5\,counties} = \beta_{cochise} = \beta_{graham} = \beta_{maricopa} = \beta_{pima} = \beta_{pinal} \end{cases}$ (For 5 counties model)

The test results are shown in table 4.1 and table 4.2. For both tests we reject the null hypotheses; which means it is necessary to examine the impacts from climatic conditions on crop yield from county to county.

For the second question, we need to test the effect of land use change on cotton yield in Yuma County. For simplicity, we use dummy variable to test if there is significant yield change due to land use change in the mid of 1970s. Here we define the dummy variable as:

$D\_land,cotton = \begin{cases} 1, & \text{if } year < 1977; \\ 0, & \text{if } year \geq 1977; \end{cases}$ (crop=cotton upland, county=Yumalapaz)

Null Hypothesis: $\alpha_{D\_land,cotton} = 0;$

According to the hypothesis testing results in table 4.3, there does exist a yield structure change between years before 1976 and years after 1976. The cotton yield decline in Yuma County during the mid of 1970s is probably related with the land use change (Table 3-6(c)). As a summary of all the inferences we can reach from these testing results, first, while we can estimate a combined SUR model to see the general relationship between cotton and hay yield in all counties it is also necessary to estimate the model for each county. And then, we can estimate the model for Yuma County separately because of the unique observed data pattern. Besides, it is more efficient to estimate the model for Yuma County from 1977 to 2008. For this time period, there is no evident intervention between impacts from change of climatic conditions and land use change.

Table 4.3 Statistic Test: Effect of Land Use Change on Cotton Yield in Yuma County

	Restricted model	Unrestricted model (with dummy for land use change)
Estimate of $\alpha_{D\_land,cotton}$		436.0883 (t value = 5.35)
Likelihood Value	-482.7654	-470.7843
Number of Estimates	29	30
Number of Restrictions	1	
Statistic Test (Chi-square)	23.96	
Decision	Reject Null Hypothesis	

4.2 Estimates and Predicted Yield Curves

As discussed in section 2.4, the proposed SUR model can be estimated in different level. First and overall, it will be interesting to look at the correlation between cotton and hay and their response to climatic variation in all counties being studied generally. The table 4.4 shows all of the coefficients for both county level SUR model and overall SUR model. From the estimates of overall model we can tell that: (1) precipitation is highly significant in explaining the change of both cotton and hay yield, while the variation of precipitation is only significant in explaining hay yield change; (2) temperature level is only significant in explaining the mean change of hay yield, but variation of temperature is highly significant in explaining the mean change of cotton yield; (3) variation of temperature (VT) is highly significant in explaining the variation of cotton yield, and the cotton yield variation is also sensitive to variation of precipitation (VP); (4) for hay yield variation, it is sensitive to variation of temperature (VT); but it is also sensitive to yearly precipitation level and variation of precipitation (VP) (at 10% confidence level). As a brief summary, the results indicate that cotton and hay yields are sensitive to different combination of climatic factors. Cotton is sensitive to precipitation and variation of temperature, while hay is sensitive to level of precipitation and temperature, and variation of precipitation. And overall it turns out that hay yield relies more on factors of precipitation. (5) The yield of cotton and hay are highly correlated, note that the correlation may come from some common factors (other than climatic factors) which affects both cotton and hay yield, such as policy and some other anthropic factors.

Table 4.4 (a) Coefficient Estimates for both County Level and Overall SUR Model (Cotton Upland)

		Cochise		Graham		Maricopa		Pima		Pinal		Yuma-La Paz*		Overall**	
	Variable	**estimate**	**t value**	**estimate**	**t value**	**estimate**	**t value**	**estimate**	**t value**	**estimate**	**t value**	**estimate**	**t value**	**estimate**	**t value**
Cotton	**constant**	-30529	-0.2827	-49256	-0.9715	-47310	-0.7429	-219356	-3.3740	-92728	-1.4436	-161926	-3.1106	-16469	-4.3283
	year	10.0227	3.0046	12.1278	7.6610	8.4706	4.1779	12.2648	9.1487	6.6405	5.1013	10.4735	5.4458	8.2008	13.9541
	P	707.07	1.1856	-156.28	-0.3994	691.5438	2.1251	-544.38	-1.1118	1792.16	5.3558	479.28	1.3112	68.08	1.2981
	T	192.34	0.0551	910.99	0.5400	847.2580	0.4790	5911.35	3.1342	2036.46	1.0951	3909.15	2.7552	-14.08	-0.1386
	P*T	-9.0451	-0.9886	2.9234	0.4669	-9.6484	-2.0919	6.9161	0.9957	-25.2199	-5.1989	-5.8061	-1.1623	-0.6993	-1.0510
	P^2	-5.7244	-1.8326	-0.9876	-1.0302	-0.9159	-0.7028	1.9939	1.7879	-2.7649	-3.8426	-5.5987	-2.1964	-0.9559	-2.0354
	T^2	-0.4389	-0.0154	-7.9623	-0.5592	-5.5766	-0.4422	-44.4102	-3.2209	-12.5527	-0.9327	-26.83	-2.7473	0.3727	0.4930
	VT	1.4404	0.8153	0.0097	0.0083	-1.4223	-1.8680	-0.4103	-0.5504	-0.8988	-1.0964	-0.9227	-1.0750	2.1818	5.6845
	VP	42.1889	1.3401	63.8320	2.3757	-45.4207	-0.4645	78.9258	3.4256	-35.2197	-1.9393	10.1817	0.2523	10.3806	0.8070
	constant	10.5993	0.6056	9.9371	0.7981	10.0735	0.6205	-7.5981	-0.4911	9.7076	0.7880	9.7822	0.5812	12.4909	8.7050
	VT	-0.0230	-1.4549	-0.0162	-1.1168	-0.0052	-0.4420	0.0127	1.1281	-0.0104	-1.0927	0.0196	1.8500	-0.0134	-3.5744
	VP	-0.0622	-0.2376	0.2950	0.7552	-0.0773	-0.0543	-0.2078	-0.5135	-0.6783	-2.0440	0.1346	0.2231	-0.2139	-1.9365
	P	0.0082	0.0828	-0.1001	-1.3250	-0.0507	-0.2907	0.0429	0.3872	-0.0007	-0.0069	-0.0994	-0.6852	0.0302	1.0781
	T	0.0455	0.1594	0.0507	0.2631	0.0033	0.0151	0.1997	0.9356	0.0253	0.1463	-0.0670	-0.2959	-0.0114	-0.5332

* The model is estimated only for time period 1977-2008 to avoid the intervention from land use change.

**The overall model is estimated for 5 counties (except Yuma-La Paz) during time period 1965-2008.

Table 4.4 (b) Coefficient Estimates for both County Level and Overall SUR Model (Hay all, correlation)

		Cochise		Graham		Maricopa		Pima		Pinal		Yuma-La Paz		Overall	
	Variable	**estimate**	**t value**	**estimate**	**t value**	**estimate**	**t value**	**estimate**	**t value**	**estimate**	**t value**	**estimate**	**t value**	**estimate**	**t value**
Hay	**constant**	-481.49	-1.3949	-162.42	-0.3877	-215.82	-1.4023	-1338.63	-2.7563	-871.70	-3.2621	964.93	2.6726	-23.69	-1.3741
	year	0.0601	8.3835	0.0357	3.4752	0.0729	12.7519	0.1039	9.9513	0.1120	16.5226	0.0167	2.3351	0.0645	25.2364
	P	-2.4009	-2.0353	0.0651	0.0314	1.8042	1.6516	-3.9193	-1.3034	1.7726	1.0839	-13.86	-4.1706	-0.7152	-2.7276
	T	12.6276	1.1007	3.3372	0.2372	2.0762	0.4788	34.4486	2.4676	19.0849	2.4821	-26.30	-2.6841	-2.9890	-6.5867
	P*T	0.0326	1.7470	-0.0072	-0.2184	-0.0235	-1.5502	0.0551	1.2524	-0.0245	-1.0423	0.1886	4.1958	0.0100	3.0015
	P^2	0.0138	1.5475	0.0096	1.7424	-0.0064	-1.8812	0.0065	1.0457	-0.0040	-1.1836	0.0161	1.3723	0.0026	1.1410
	T^2	-0.1067	-1.1320	-0.0272	-0.2276	-0.0134	-0.4347	-0.2596	-2.5331	-0.1385	-2.4810	0.1747	2.6173	0.0229	6.8258
	VT	-0.0112	-2.5448	-0.0084	-1.3435	-0.0104	-4.2002	-0.0143	-2.8732	-0.0096	-2.7609	-0.0108	-2.8912	-0.0017	-1.0742
	VP	-0.2887	-4.2742	-0.1222	-0.9343	-0.3060	-3.8640	-0.0963	-0.5292	-0.2145	-2.0101	-0.5255	-1.1276	-0.2158	-3.6043
	constant	-0.2637	-0.0182	-18.2179	-1.1474	-2.0280	-0.1339	-55.1356	-2.4462	-1.5105	-0.0863	-1.2634	-0.0388	-0.4570	-0.5735
	VT	-0.0411	-2.7092	0.0101	0.6806	0.0009	0.0688	-0.0119	-0.7990	-0.0204	-1.6449	-0.0503	-1.9747	-0.0105	-3.1511
	VP	-1.2700	-4.1664	0.1208	0.3734	-0.5735	-0.9639	-0.0353	-0.0667	-0.0238	-0.0519	-1.2451	-0.2778	-0.2217	-1.6288
	P	0.2111	2.3930	-0.1478	-1.8752	0.0844	0.7164	0.1130	0.7277	-0.0561	-0.4568	0.3420	0.5616	0.0970	3.4545
	T	0.0686	0.3076	0.2869	1.1634	-0.0219	-0.1077	0.8025	2.5471	0.0612	0.2474	0.1041	0.2192	0.0000	0.0093
correlation		0.3112	3.6565	0.1018	1.0938	0.1504	1.5441	-0.0463	-0.5087	0.0368	0.3999	0.3175	3.7797	0.1622	4.2288

* The model is estimated only for time period 1977-2008 to avoid the intervention from land use change.

**The overall model is estimated for 5 counties (except Yuma-La Paz) during time period 1965-2008.

For county level SUR model, the estimates show a lot of heterogeneity. To understand these estimates, we should look at the general differences among six counties first. As table 3.2 indicates, the top 3 counties (Maricopa, Pinal, and Yuma-La Paz) account for more than 80% of the state agricultural production value, while other three counties only account for about 11%. For individual crops, as table 3.3 shows, Pinal, Maricopa and Yuma-La Paz account for more than 80% of cotton production and more than 87% of hay production in 2008. For other three counties (Cochise, Graham and Pima), it is highly possible that the production is operated mainly based on several big farms. This kind of big farm-based production pattern makes the crop's response to climatic variation much more heterogeneous. On the other hand, for three leading counties (Maricopa, Pinal, and Yuma-La Paz), there are far more farms in operation. So on average, the crop's response to climatic variation is more consistent and easy to capture in these big counties.

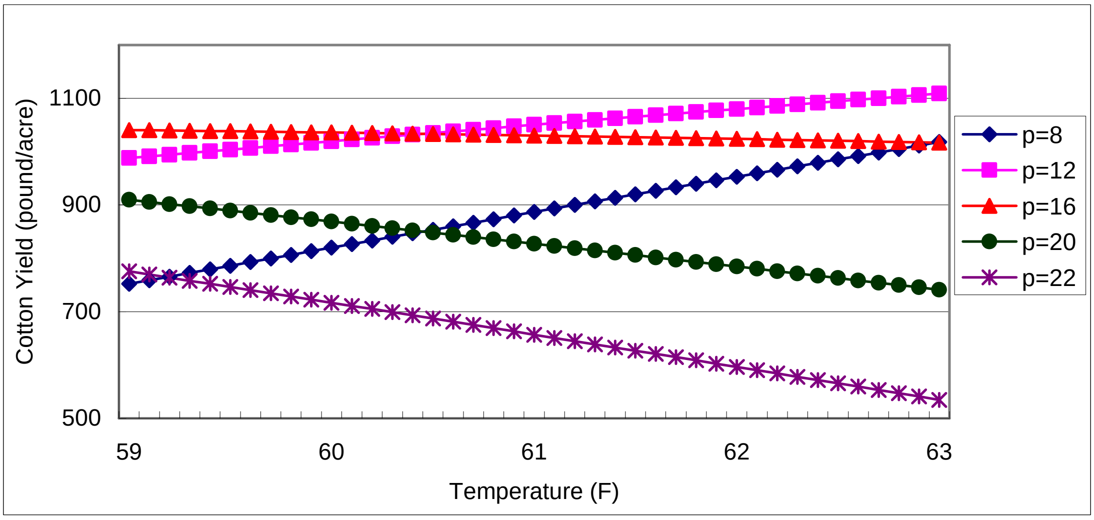
Accompanying with the characteristics of six counties being studied, it is easier to understand the estimates for county level SUR model in table 4.4. First, for cotton: (1) time trend is a highly significant explanatory variable in each of county level SUR model. (2) In Cochise and Graham County, the mean change of yield is mainly explained by time trend. And both temperature and precipitation are unable to explain the mean change of yield significantly. The only exception is the variation of precipitation, which is significant in explaining cotton yield change in Graham. Similarly, both temperature and precipitation factors are unable to explain the variation of yield. As discussed above, the reason may be that in these small counties the crop production is big farm-based. And individual farm can easily adapt to climatic variation and mitigate the impact of climatic impact. (3) In Pima County, the mean change of cotton yield is sensitive to temperature and variation of precipitation. However, it is worthy to note that, Pima County is also growing lots of pima cotton which is not included in this research. In this research, the cotton is referred as upland cotton. (4) In

Maricopa and Pinal County, precipitation is significant in explaining the mean change of yield. This kind of response could be related to the irrigation resources in these areas. Another observation for Pinal County is that, variation of precipitation is significant in explaining both mean and variation change of yield. (5) For the combined county Yuma-La Paz (model is estimated for time period 1977-2008), which is not included in the overall model for the reason to maintain a balanced SUR model with equal number of observations. The explaining power of temperature and precipitation here is highly significant, but the yield variation is only slightly sensitive to variation of temperature. A related fact might be that, Yuma-Lapaz County has a relatively lower precipitation level and higher temperature level than other counties (the historical maximum precipitation is less than 9 inch since 1965).

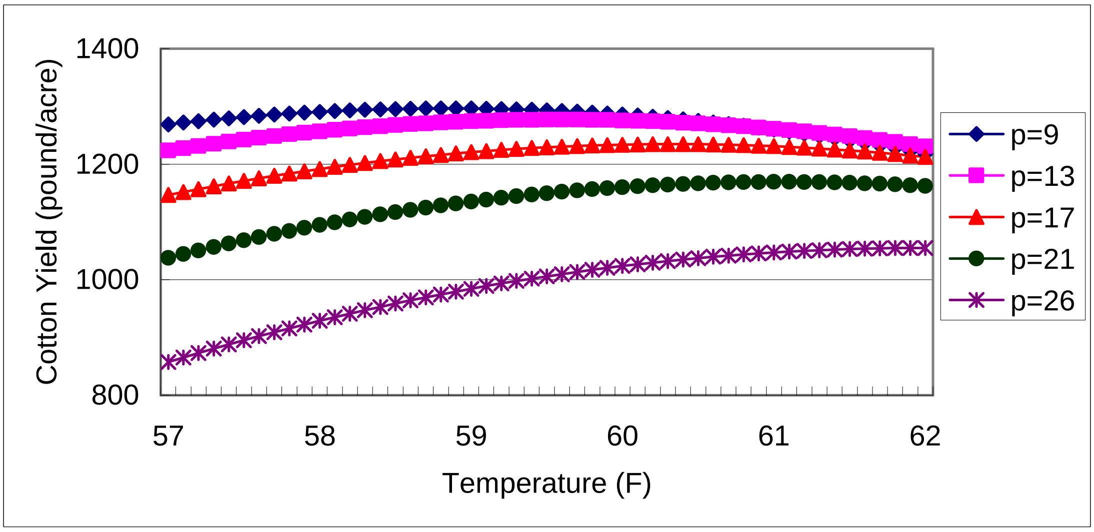
For Hay, the similar patterns are observed. (1) Time trend is a highly significant explanatory variable in each of county level SUR model. (2) In Cochise, precipitation level, variation of precipitation and temperature are all highly significant in explaining both mean and variance of yield. (3) In Graham, both temperature and precipitation are unable to explain the mean change of yield significantly, but precipitation level is slightly significant in explaining yield variation. (4) In Pima County, the mean change of hay yield is sensitive to temperature and variation of temperature. This is different from the observed response in cotton yield of Pima. And the yield variation is also sensitive to temperature level. (5) In Maricopa and Pinal County, variation of temperature and precipitation is highly significant in explaining the mean change of yield. Temperature level is also significant in explaining the yield change in Pinal. For variance of yield, only variation of temperature has slightly significant explaining power. (6) For the combined county Yuma-La Paz, the explaining power of both precipitation and temperature are highly significant. Besides, the variation of temperature is also significant in explaining both the change of yield mean and variation.

Aside from estimates of variable coefficients in mean and variance function, another part of observation is that, in all yield variance function, most of the estimates for constant term are not significant. This result supports our heteroskedastic specification. For estimates of variation correlation coefficients (correlation between yield variation of cotton and hay), as we have already mentioned, overall it is highly correlated. For each of Individual counties, it is highly correlated in Cochise and Yuma-La Paz while it is only slightly significant in Maricopa. And in other three counties, the correlation is not evident. The explanation for this could be that, as it indicates by table 3.3, in Cochise, Maricopa and Yuma-La Paz County cotton and hay production are equally important. For other three counties, at least by looking at the production percentage, cotton is more important than hay production, especially for Graham.

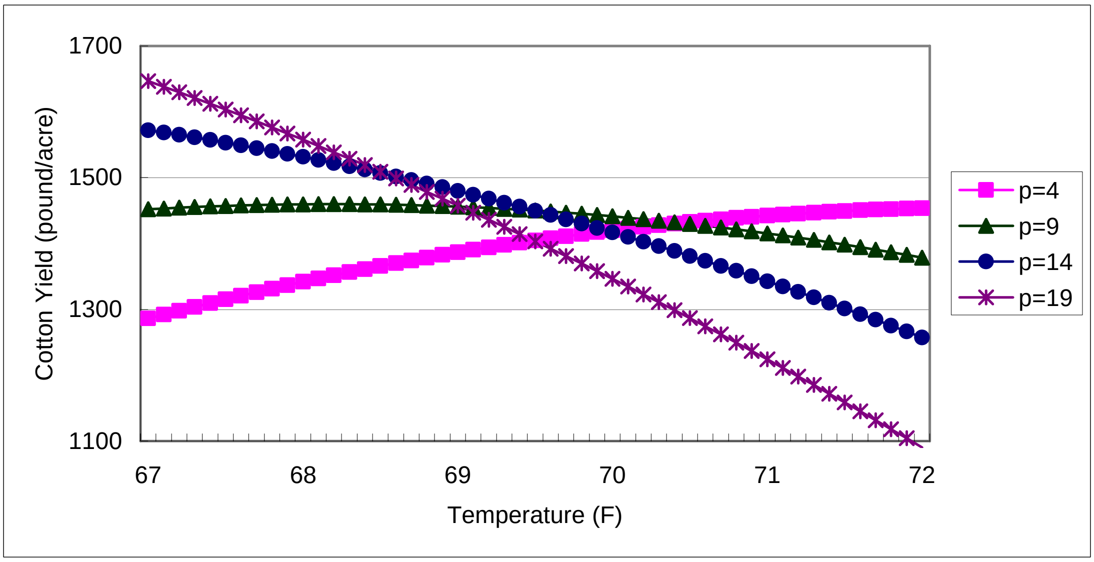
Another way to understand and interpret these estimates is by drawing the predicted curves, which is also convenient in understanding the marginal effect of explanatory variables with interaction and squared terms. In this research, since we use quadratic functional form it is difficult to tell the effect of individual variables only by looking at the coefficient estimates. Figures 4.1 to 4.4 show all of the yield prediction curves for both cotton and hay in county level. The vertical axis shows the predicted yield of certain crop, and the horizontal axis is the corresponding climatic measurement used for prediction. In each figure the predicted curves are given in different level which is measured by another climatic measurement (temperature or precipitation). It is also worth to note that, the range of climatic measurements is chosen based on the range of historical observation of these measurements in time period 1965-2008 for each county. The lower bound of range is a little lower than the minimum of historical observations, and the upper bound of range is a little higher than the maximum of historical observations.



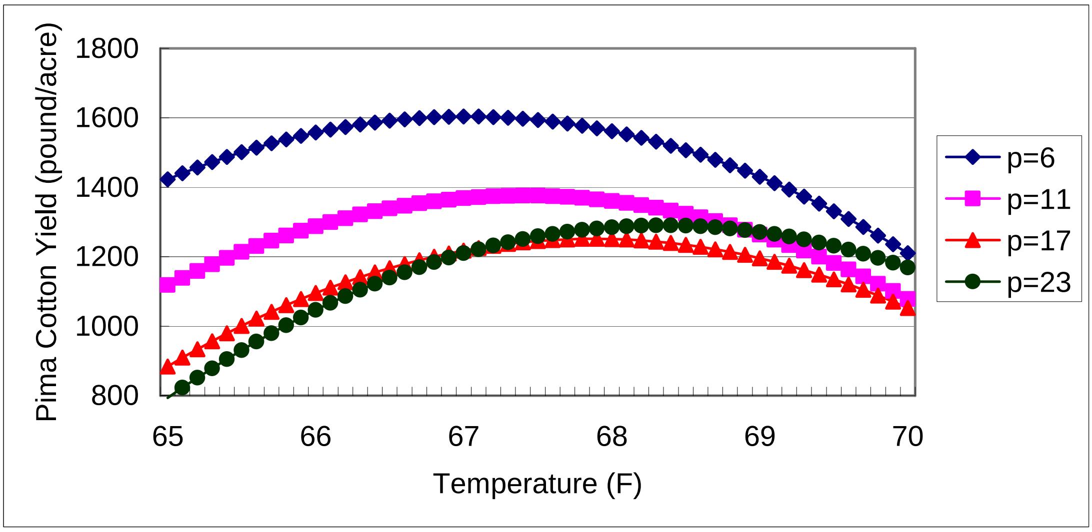
4-1(a): Cochise Predicted Cotton Yield on Temperature (Year 2010)



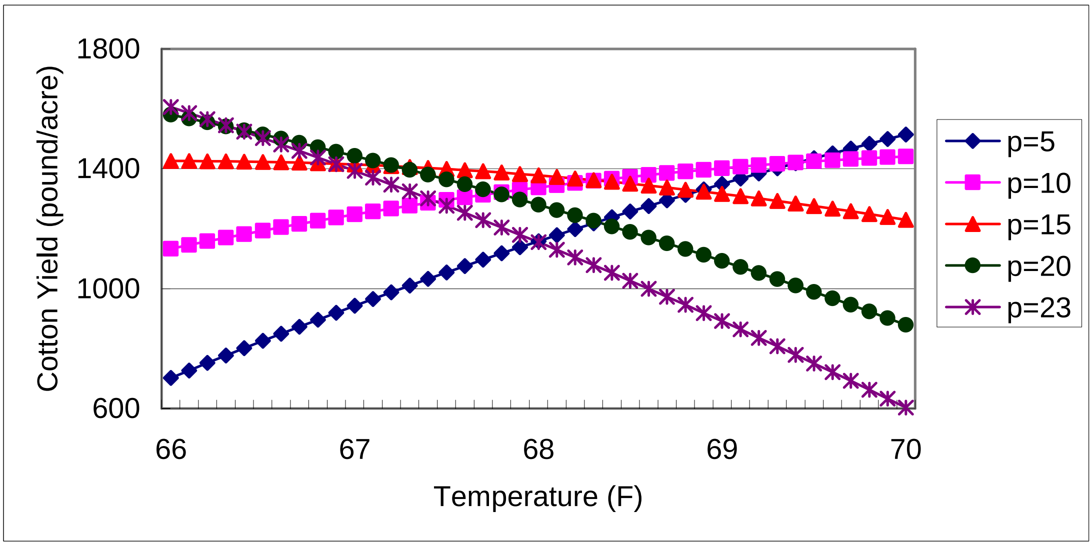
4-1(b): Graham Predicted Cotton Yield on Temperature (Year 2010)



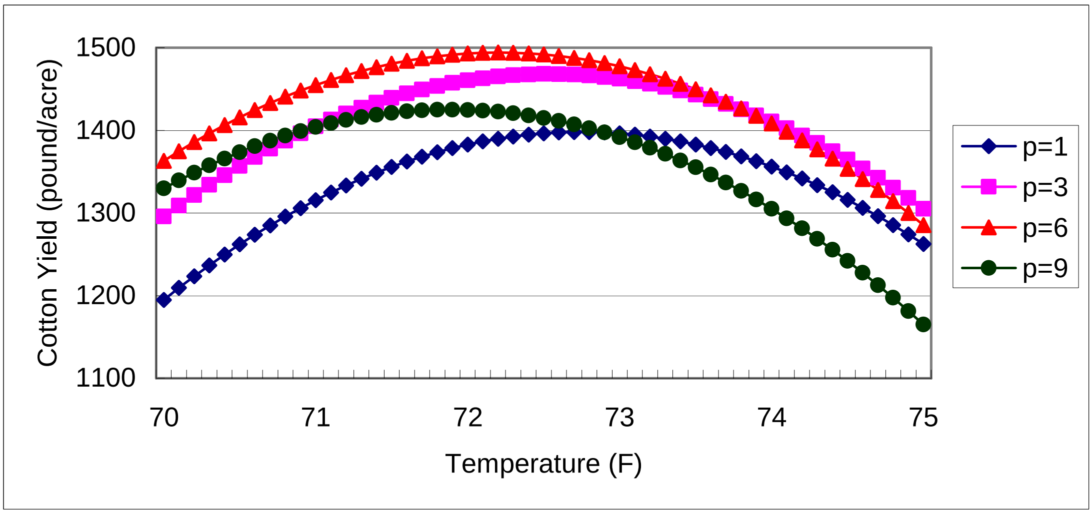
4-1(c): Maricopa Predicted Cotton Yield on Temperature (Year 2010)



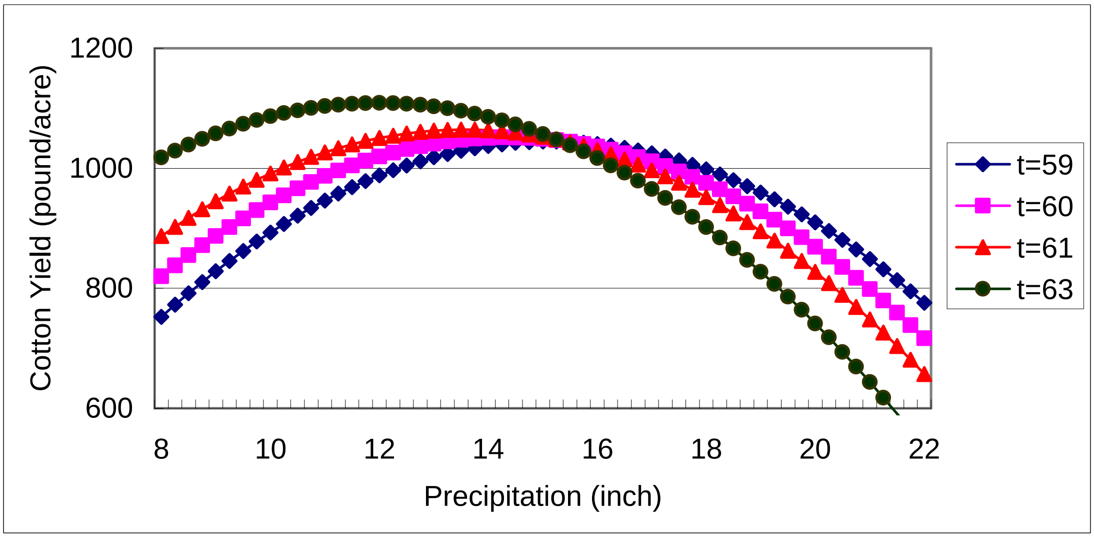
4-1(d): Pima Predicted Cotton Yield on Temperature (Year 2010)



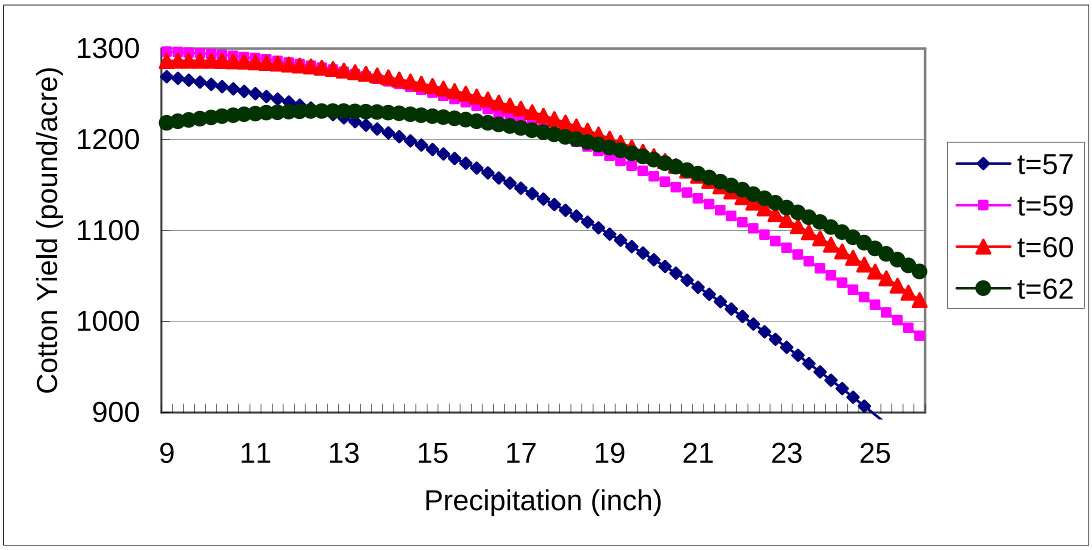
4-1(e): Pinal Predicted Cotton Yield on Temperature (Year 2010)



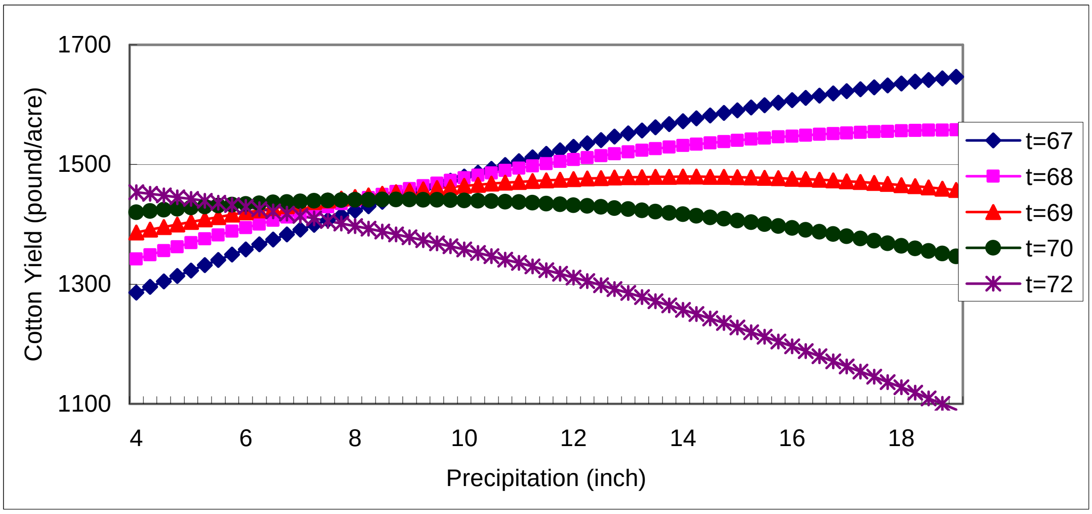
4-1 (f): Yuma-La Paz Predicted Cotton Yield on Temperature (Year 2010)



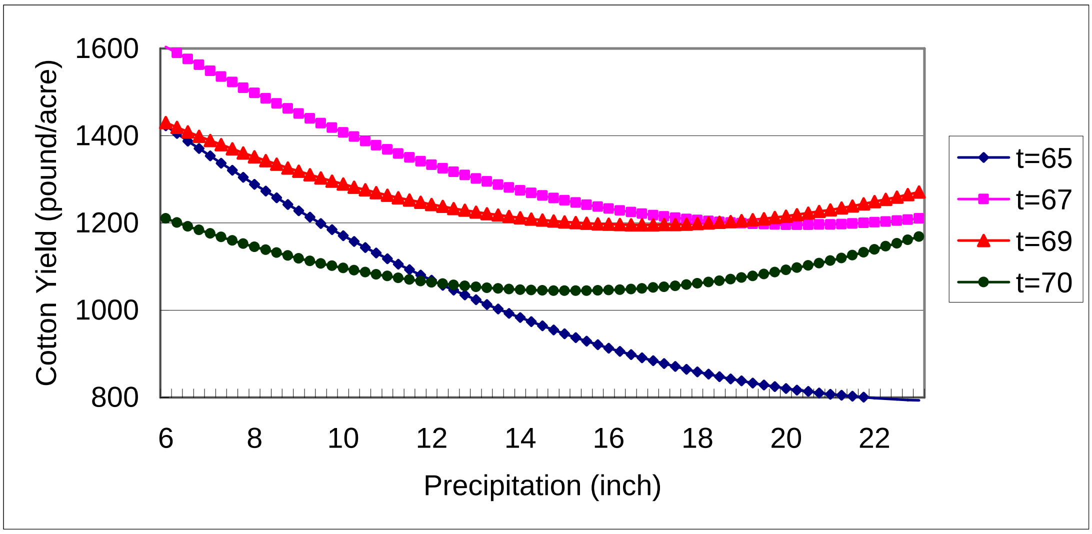
4-2 (a): Cochise Predicted Cotton Yield on Precipitation (Year 2010)



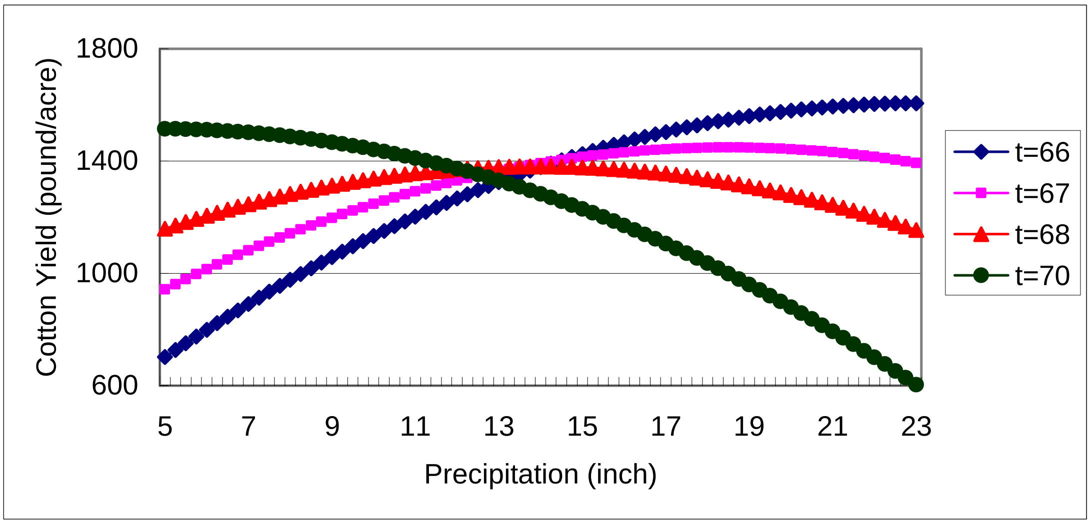
4-2 (b): Graham Predicted Cotton Yield on Precipitation (Year 2010)



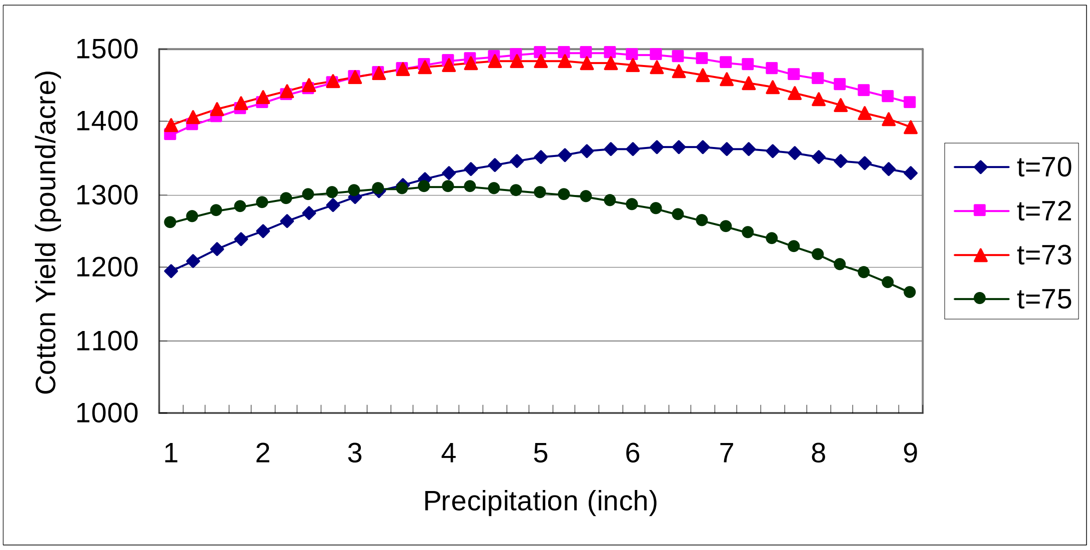
4-2 (c): Maricopa Predicted Cotton Yield on Precipitation (Year 2010)



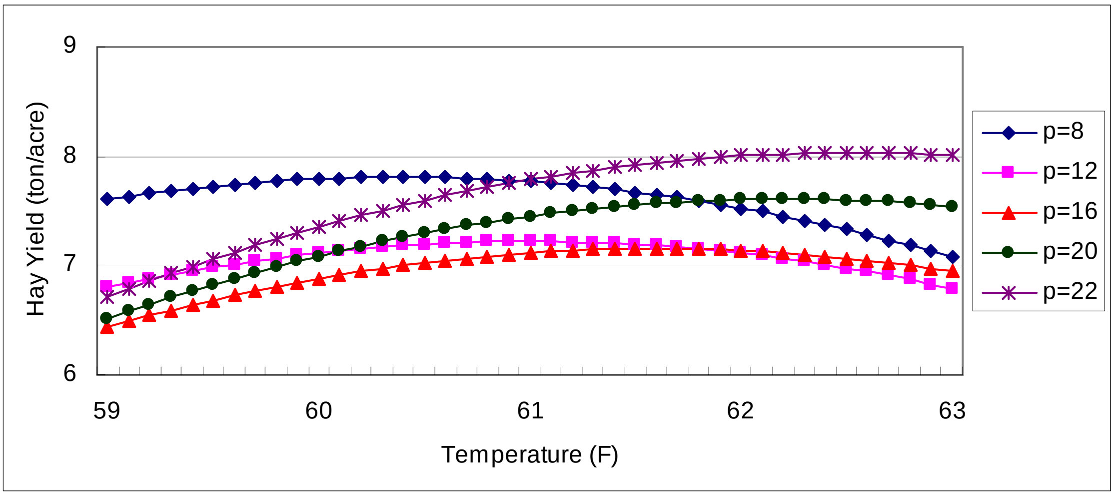
4-2 (d): Pima Predicted Cotton Yield on Precipitation (Year 2010)



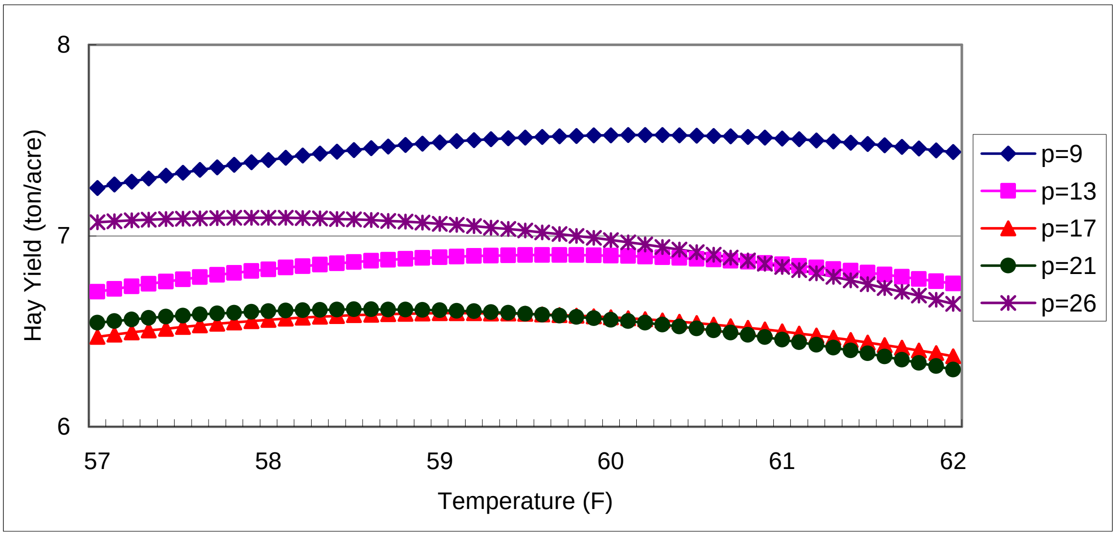
4-2 (e): Pinal Predicted Cotton Yield on Precipitation (Year 2010)



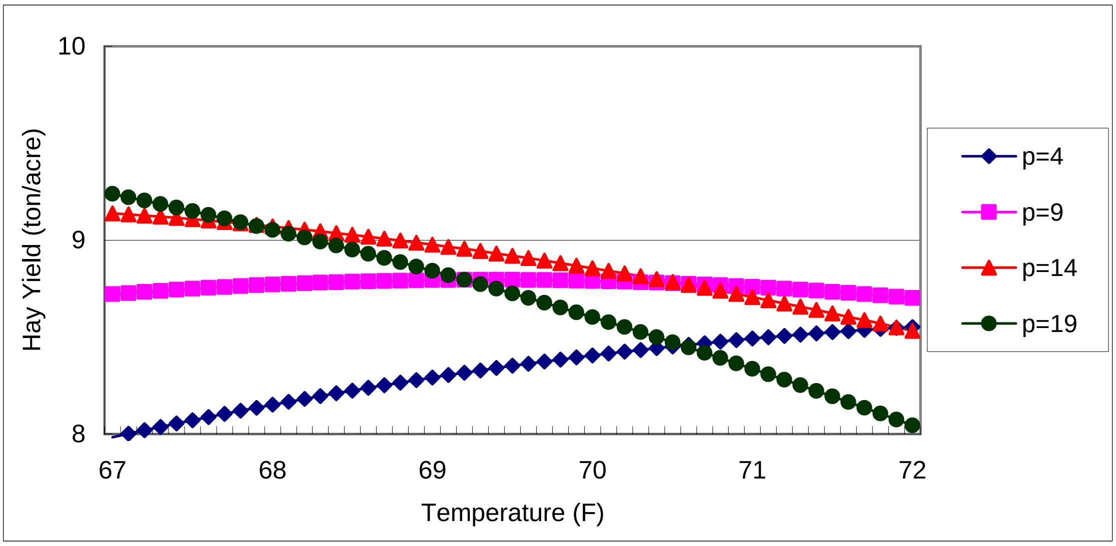
4-2 (f): Yuma-La Paz Predicted Cotton Yield on Precipitation (Year 2010)



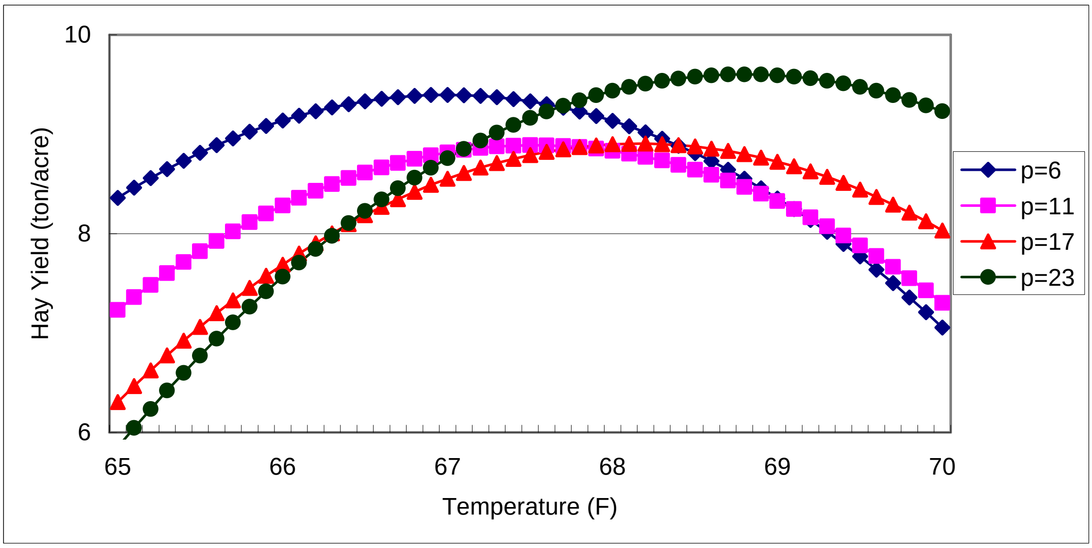
4-3 (a): Cochise Predicted Hay Yield on Temperature (Year 2010)



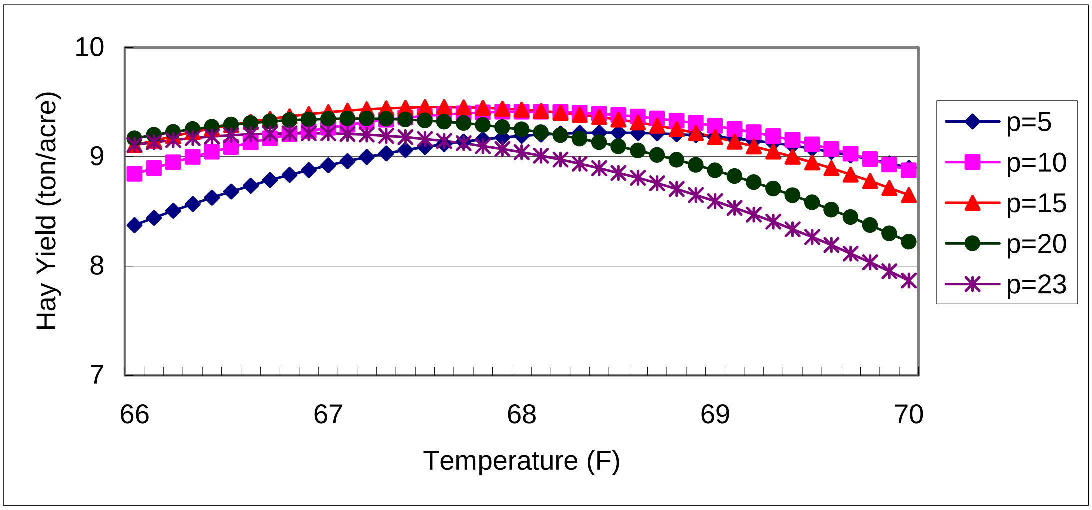
4-3 (b): Graham Predicted Hay Yield on Temperature (Year 2010)



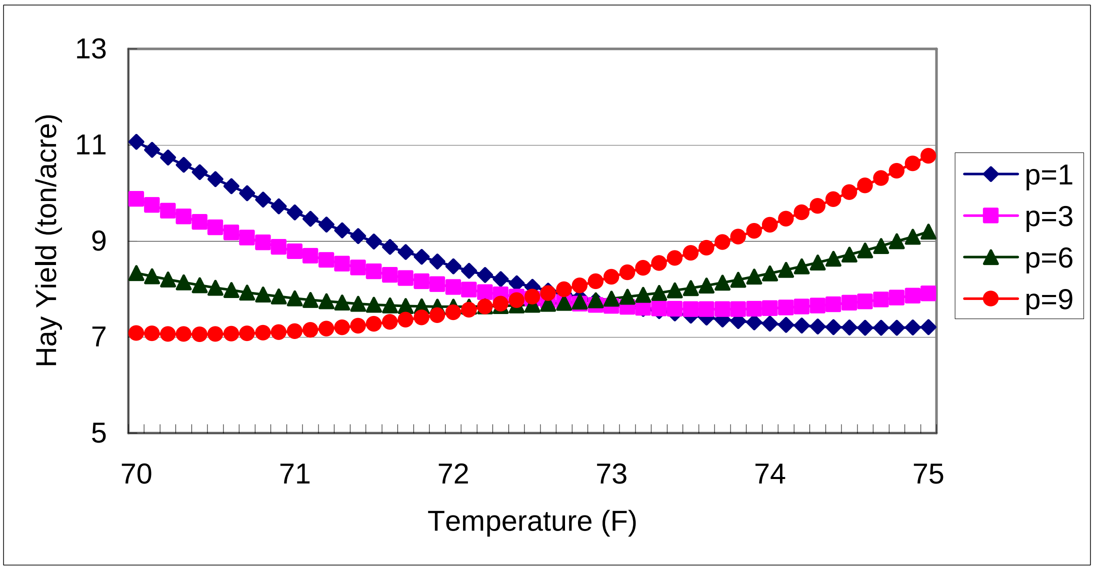
4-3 (c): Maricopa Predicted Hay Yield on Temperature (Year 2010)



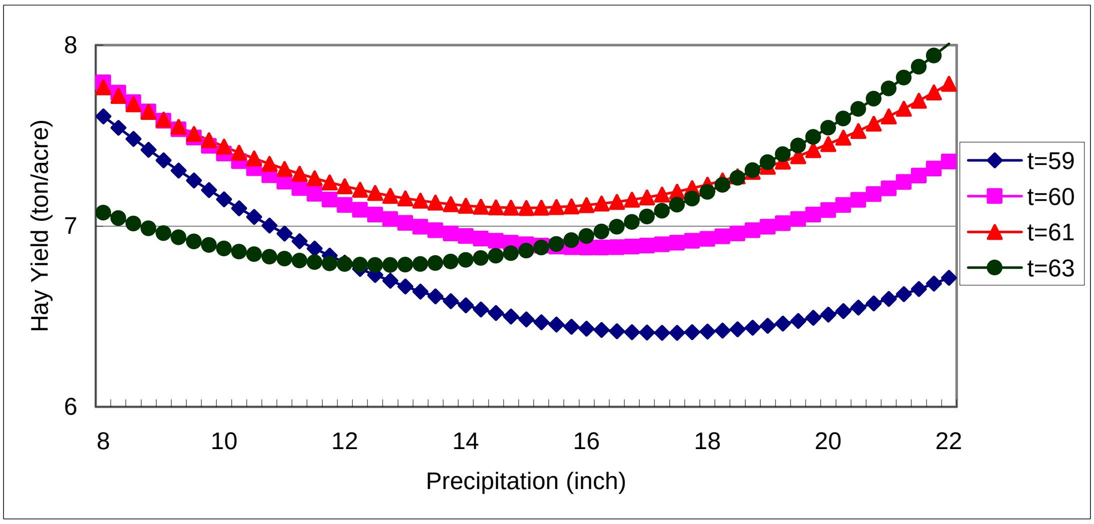
4-3 (d): Pima Predicted Hay Yield on Temperature (Year 2010)



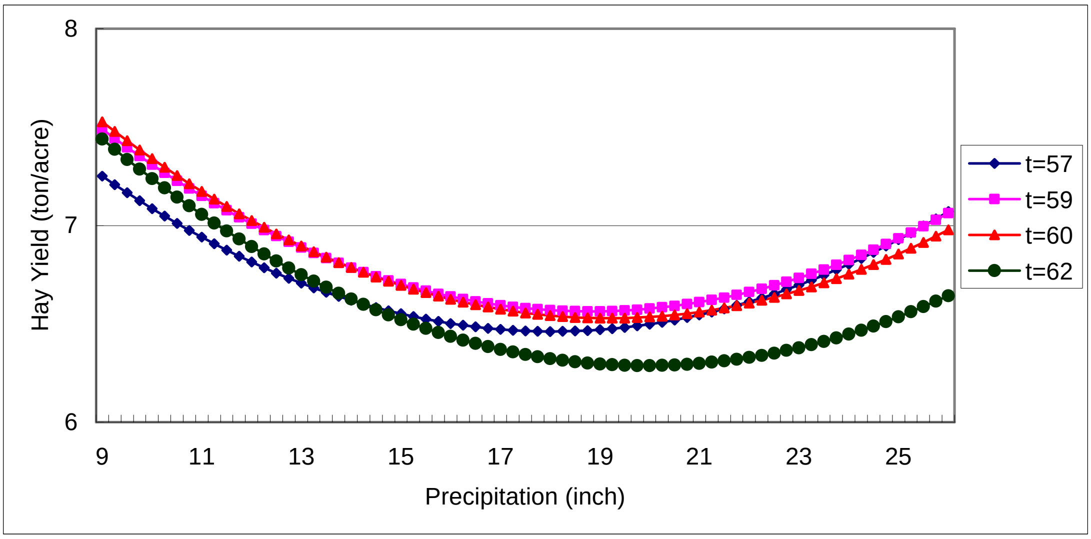
4-3 (e): Pinal Predicted Hay Yield on Temperature (Year 2010)



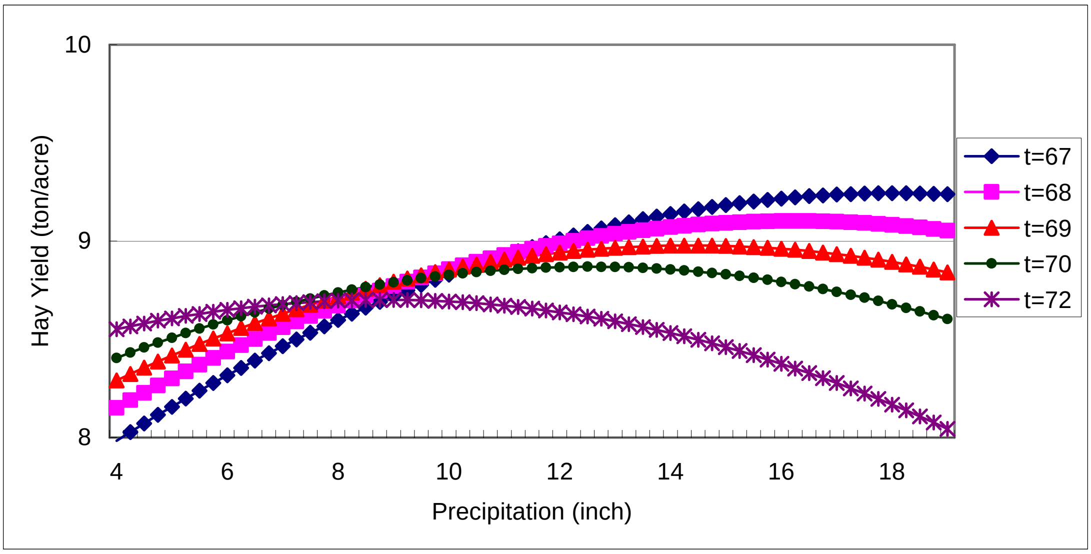
4-3 (f): Yuma-La Paz Predicted Hay Yield on Temperature (Year 2010)



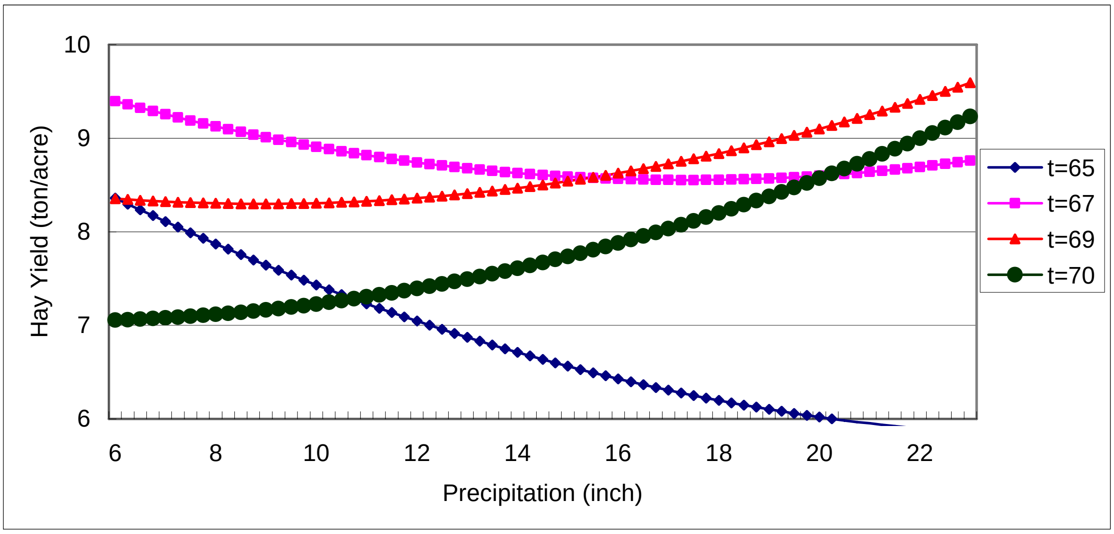
4-4 (a): Cochise Predicted Hay Yield on Precipitation (Year 2010)



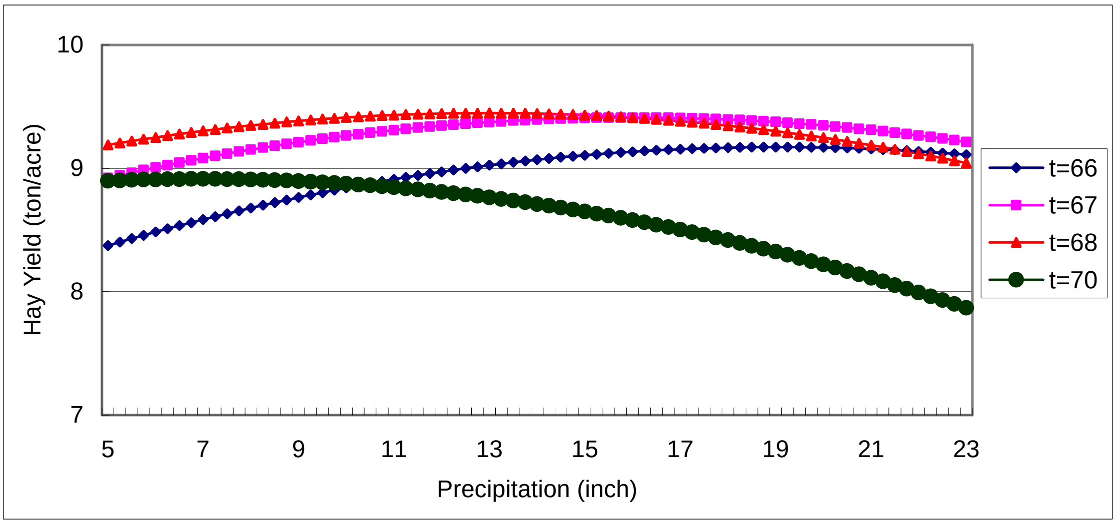
4-4 (b): Graham Predicted Hay Yield on Precipitation (Year 2010)



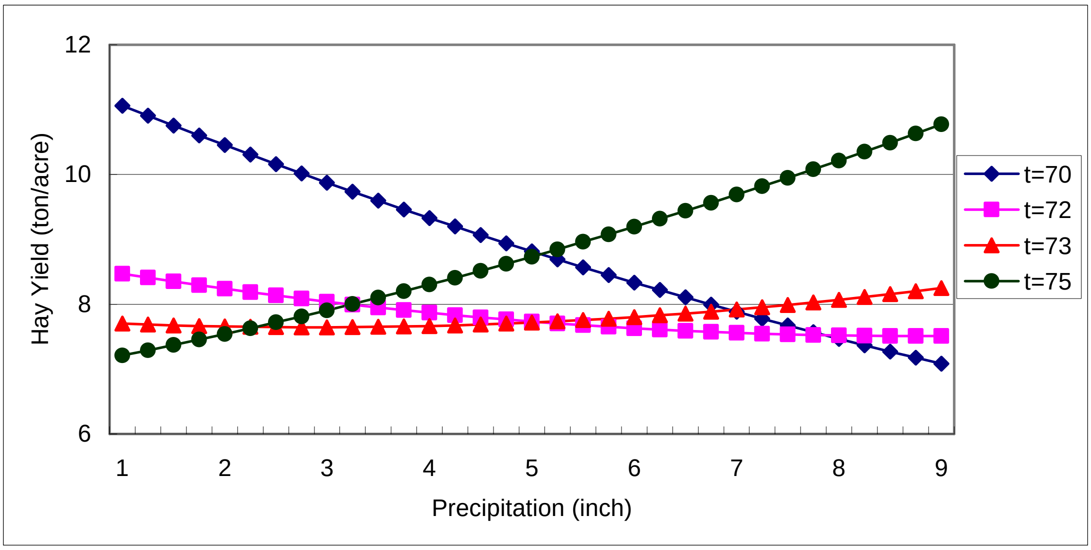
4-4 (c): Maricopa Predicted Hay Yield on Precipitation (Year 2010)



4-4(d): Pima Predicted Hay Yield on Precipitation (Year 2010)



4-4 (e): Pinal Predicted Hay Yield on Precipitation (Year 2010)



4-4 (f): Yuma-La Paz Predicted Hay Yield on Precipitation (Year 2010)

4.3 Marginal Effects

As mentioned in discussing the estimate of coefficients, because of the quadratic functional form it is convenient to look at the predicted curves and marginal effect instead of individual estimates of coefficients. Predicted curves are more visual in showing marginal effect and change of yield, but it is not precise (in a numerical sense) and it is hard to tell the significance of marginal effect. So within next two subsections the numerical marginal effect and its significance will be discussed. The standard error and hence t statistics values are approximated by delta methods.

4.3.1 Marginal Effect on Yield

From basic model equation 2.1 to 2.3, we know that the crop yield is separated into two parts: mean function (deterministic part) and variance function (stochastic part). And different function has different explanatory variables, which means we have to discuss the marginal effect for each function respectively. In this section, we discuss the marginal effects of temperature and precipitation on mean yield. Table 4.5 reports the marginal effects at different quartiles of explanatory variables and t statistic value as a measure of statistical significance. The quartiles are calculated based on the historical observation on temperature and precipitation in time period 1965-2008. For example, when the marginal effect of temperature on cotton yield is calculated at 25% quartile (T=60.38 F) the value of precipitation is holding at its median level (P = 12.44 inch), and then marginal effect comes as 22.2304 lb/(acre*F) with high significance (t=3.3502). Similarly, when the marginal effect of precipitation on cotton yield is calculated at different quartiles the value of temperature is holding at the median level (T=67.1 F). Other than making the reference solely based on the marginal effect at median or mean level, we look at the marginal effect at different quartile levels. By this way we are able to tell the change of marginal effect with respect to the change of explanatory variables.

For cotton yield, the marginal effect of precipitation varies from positive to negative as precipitation level increases. At higher precipitation level (>15 inch) it has significant negative marginal effect, which means higher level of precipitation is harmful to the cotton productivity even though it is necessarily beneficial at lower level. On the contrary, the marginal effect of temperature is ambiguously positive with high significance as temperature increases. For hay yield, the marginal effect of precipitation is significantly negative at low level of precipitation. When the precipitation level is higher, the marginal effect becomes positive and significant. On the other hand, the marginal effect of temperature varies from negative at low temperature level to positive at mid and higher temperature level. And the marginal effect of temperature on hay yield is consistently of high significance. To give a summary from another way, the marginal effects indicate: (1) both cotton and hay adapts to temperature factors better than to precipitation factors; (2) generally temperature has positive marginal effect on cotton and hay yield, but this positive marginal effect varies as temperature and precipitation level change; (3) in some cases, temperature also has negative marginal effect on crop yield. For example, temperature has a significant negative marginal effect on hay yield at relative low temperature level (table 4.5).

Table 4.5 Marginal Effect of Variables on Mean of Crop Yield

	Quartiles of Variables			Marginal Effects of Variables on Mean of Yield			
Crops	Quartiles	P	T	P (t values)		T (t values)	
Cotton Upland (pound/acre)	Minimum	4.14	57.19	13.2366	1.3861	19.8512	1.7621
	25%	10.34	60.38	1.3927	0.3134	22.2304	3.3502
	Median	12.44	67.10	-2.6318	-0.7977	27.2405	5.8315
	75%	15.71	68.95	-8.8835	-2.4997	28.6165	3.9732
	Maximum	25.73	71.99	-28.0404	-2.4092	30.8840	2.6543
Hay All (ton/acre)	Minimum	4.14	57.19	-0.0211	-0.4948	-0.2460	-4.9043
	25%	10.34	60.38	0.0115	0.6218	-0.0999	-3.3830
	Median	12.44	67.10	0.0226	1.5332	0.2078	10.3172
	75%	15.71	68.95	0.0398	2.0053	0.2923	9.2945
	Maximum	25.73	71.99	0.0926	1.5076	0.4316	8.4325

Table 4.6 Marginal Effect of Variables on Variance of Crop Yield

Crops	Quartiles of Variables					Marginal Effects of Variables on Yield Variance							
	Quartiles	VP	VT	P	T	VP (t values)		VT (t values)		P (t values)		T (t values)	
Cotton Upland (lb/acre)	Minimum	0.16	136.06	4.14	57.19	-3043.41	-1.5732	-304.25	-2.0046	270.41	1.4196	-147.21	-0.4875
	25%	0.67	171.56	10.34	60.38	-2731.02	-1.7236	-189.07	-2.7188	325.96	1.1404	-141.94	-0.5043
	Median	1.15	186.72	12.44	67.1	-2463.32	-1.8959	-154.31	-3.2013	347.33	1.0689	-131.46	-0.5437
	75%	1.63	205.12	15.71	68.95	-2223.72	-2.1042	-120.58	-4.0662	383.33	0.9740	-128.72	-0.5556
	Maximum	4.34	242.25	25.73	71.99	-1245.63	-5.3576	-73.32	-8.2674	518.59	0.7654	-124.33	-0.5765
Hay All (ton/acre)	Minimum	0.16	136.06	4.14	57.19	-0.0634	-1.3164	-0.0041	-1.9991	0.0100	11.4901	0.0000	0.0093
	25%	0.67	171.56	10.34	60.38	-0.0566	-1.4471	-0.0028	-2.6115	0.0182	4.2149	0.0000	0.0093
	Median	1.15	186.72	12.44	67.1	-0.0509	-1.5981	-0.0024	-3.0017	0.0223	3.3978	0.0000	0.0093
	75%	1.63	205.12	15.71	68.95	-0.0458	-1.7823	-0.0020	-3.6606	0.0306	2.6030	0.0000	0.0093
	Maximum	4.34	242.25	25.73	71.99	-0.0251	-4.9963	-0.0013	-6.4069	0.0808	1.5086	0.0000	0.0093

4.3.2 Marginal Effect on Yield Variance

As we mentioned in last section, the yield function can be separated into two parts. In this subsection we look at the marginal effect on the second part: yield variance. The calculation method for marginal effect on yield variance is same as the method for calculation of marginal effect on mean of yield. However, for the variance function we have four (linearly) independent variables: temperature (T), precipitation (P), within year variance of temperature and precipitation (VP, VT). The marginal effect and t statistic value are reported in table 4.6. For the variables VP and VT, all of the marginal effects are negative without regard to significance. As a matter of fact, this observation is consistent with the basic rule of plant growth. A higher within year variance of temperature and precipitation means that, given the aggregate level the related resources (sunshine, water) are distributed less evenly over 12 months in the year. Therefore, the negative marginal effect means that as the variation of temperature and precipitation becomes larger, the crop yield variation become smaller.

As it indicates by table 4.6, for both cotton and hay yield variance, the marginal effect of within year variance of precipitation is generally not significant. The only exception is that, the marginal effect becomes significant at extremely high levels of variance of precipitation. And then, for the marginal effect of within year variance of temperature, it is highly significant for both cotton and hay at all quartiles. And the absolute value of marginal effects keeps decreasing, and here the law of diminishing marginal utility is applicable. All of these observations are consistently related with the observations in previous section that both cotton and hay adapt to temperature factors more than precipitation factors.

For the mean temperature and precipitation which we have been using in both mean and variance functions, temperature level is not significant at all for both cotton and hay yield variance and its marginal effect is relatively small (not significantly different from 0). On the other hand, the precipitation level is more powerful in accounting for the yield variation of cotton and hay. For cotton yield variance, the marginal effect is only slightly significant at low precipitation level. However, for hay yield variance, the precipitation level has a highly significant positive marginal effect on yield variance. This last observation could be an interesting story to tell. As yearly precipitation level increases, the hay yield could be more unstable and of more variation. Since agricultural areas in Arizona generally have a higher irrigation rate (table 3.4), so this response of hay yield on precipitation change might has interaction with the irrigation resources associated with hay production.

4.4 Correlation of Crop Yields

When we discuss the estimates in table 4.4, we have already mentioned part of the correlation between cotton and hay yield. In this research, we factorize the error term into two parts. The first part is the error that can be interpreted by climatic variables (heteroscedasticity); another part is the random white noise which follows a normal distribution. The correlation coefficients being estimated in all SUR model are the correlation among these random errors, which in other words are the heteroscedasticity-corrected correlation among crop yields.

In table 4.4, we report that, overall, the correlation between yield of cotton and hay is positive and highly significant. For each of Individual counties, it is highly correlated in Cochise, Maricopa (slightly) and Yuma-La Paz. And in other three counties, the correlation is not evident. The explanation for this could be that, as it indicates by table 3.3, in Cochise, Maricopa and Yuma-La Paz County cotton and hay production are equally important. For other three counties, at least by looking at the production percentage, cotton is more important than hay production, especially for Graham.

Table 4.7 Correlation of Cotton Yield among Different Counties

Yield Correlation	Cochise	Graham	Maricopa	Pima	Pinal	Yuma-La Paz
Cochise		0.4734 (5.6961)	0.0711 (0.7586)	0.2482 (2.9711)	0.3009 (3.3224)	0.2203 (2.5487)
Graham			0.3054 (3.2657)	0.0387 (0.4287)	0.2914 (3.2828)	0.3616 (3.7741)
Maricopa				0.1929 (2.2337)	0.5998 (8.3875)	0.4635 (5.8206)
Pima					0.4890 (6.6655)	0.3806 (4.5817)
Pinal						0.4106 (4.8786)
Yuma-La Paz						

* t statistic value is reported in parentheses.

Table 4.8 Correlation of Hay Yield among Different Counties

YieldCorrelation	Cochise	Graham	Maricopa	Pima	Pinal	Yuma-La Paz
Cochise		0.2376 (1.8094)	0.0980 (0.8589)	-0.2789 (-2.9564)	-0.0161(-0.1423)	-0.1423 (-1.2004)
Graham			-0.0608(-0.6105)	0.0393 (0.4069)	0.0069 (0.0732)	-0.1315 (-1.3321)
Maricopa				0.3254 (3.8745)	-0.0624 (-0.6779)	0.4142 (4.9891)
Pima					0.0351 (0.3831)	0.1544 (1.6451)
Pinal						0.2814 (2.9813)
Yuma-La Paz						

t statistic value is reported in parentheses.

Figure 4.5: County Map of Southern Arizona

Equation (2-5-1) shows that the SUR model can also be constructed for each individual crop, in this kind of SUR model each equation represents a region where the crop is growing.

Via this type of SUR model, we are able to examine the correlation of yield for certain crop in different regions. In our case, we have two corps (cotton and hay) and six counties, which can be modeled into two SUR model. In table 4.7, we report the correlation of yield for cotton. In general, the crop yields among all six counties are highly correlated. The only exceptions are correlation between Cochise and Maricopa, and correlation between Pima and Graham. In table 4.8 the correlation of yield for hay is reported, we can observe that: (1) Graham is positively correlated with its neighbor county Cochise; (2) Pima is highly correlated with its neighbor counties Cochise and Maricopa; (3) Yuma-La Paz, as a leading county in hay production, is highly correlated with other two leading counties (Maricopa and Pinal) in hay production.

To summarize, we can come out three brief conclusions for the correlation of crop yields. First, the yield of cotton and hay tends to be highly correlated, which can be at least explained by similar response of two crops to climatic conditions and impacts from similar agricultural policy. Second, all of significant correlations are positive correlation except for the hay yield correlation between Pima and Cochise. The negative correlations are generally not significantly different from zero. Again, this may indicate that yield of cotton and hay is influenced by similar shocks and factors. The last conclusion is that, as it is shown in both table 4.7 and 4.8, the correlation of cotton yield among different counties is more evident than the same correlation for hay. A possible explanation for this observation could be that, upland cotton is a more specific crop and of consist yield behavior. However, for hay, we include different kinds of hay in this research, which covers alfalfa hay (main), and all other kinds of hay. So the diversity in hay category could be a reason for the heterogeneity in the correlation of hay yield.

Chapter 5: Conclusion

This research uses a stochastic SUR model to capture and differentiate the impacts of both long term climate change and short term climatic variation on crop yields. The model is estimated across seven main farming counties in Arizona for time period 1965-2008. We observe that, overall, temperature is highly significant in explaining the mean change of both cotton and hay yield; yearly precipitation level is only slightly significant in explaining the mean change of cotton yield, while it is highly significant in explaining the mean change of hay yield. For county level SUR model, the estimates show a lot of heterogeneity. From the marginal effect of temperature and precipitation we find that, for cotton yield, the marginal effect of precipitation varies from positive to negative as precipitation level increases. On the contrary, the marginal effect of temperature is ambiguously positive with high significance as temperature increases. For hay yield, the marginal effect of precipitation is significantly negative at low and mid level of precipitation. When the precipitation level is higher and close to maximum, the marginal effect becomes positive but it is not significantly different from none impact. On the other hand, the marginal effect of temperature varies from negative at low temperature level to positive at mid and higher temperature level with consistent high significance.

Although the impact of short term climatic variation on crop yields has not generally been considered by literature, short term climatic variation is a very important factor in driving crop yield fluctuation. A contribution of this research is that we account for the variance change of crop yield by introducing within year variance of temperature and precipitation into stochastic part of model. The results display that within year variance of temperature and precipitation is generally more significant than average temperature and precipitation themselves. And these new measures show a smoothing effect on crop yield, which means as within year variance of climatic conditions goes up the variation of crop yield will be smaller. This makes sense for the fact that, higher within year variance means temperature or precipitation are more averagely distributed across 12 months within the year, which is beneficial for crop growth. Another finding is that, as within year variance of climatic conditions increases their impact on the variance of crop yield tends to decrease. For mean temperature and precipitation which we have been using in both mean and variance functions, temperature level is not significant at all for both cotton and hay yield variance and its marginal effect is relatively small (not significantly different from 0). On the other hand, the precipitation level is more powerful in accounting for the yield variation of cotton and hay.

Instead of estimating the yield correlation of crops directly by calculating the covariance matrix among residuals, we propose an alternative way to estimate the correlation matrix. The estimated correlation matrix displays the relationship among the unobserved structural errors in the system beyond the observed errors associated with the heteroskedasticity of each equation. In general, first we find that, the yield variation of cotton and hay tends to be highly correlated, which can be at least explained by similar response of two crops to climatic conditions and impacts from similar agricultural policy. Second, all of significant correlations are positive correlation. There are some negative correlations, but which is not significantly different from none correlation. Again, this may indicate that yield variation of cotton and hay comes from similar shocks and factors. The last observation is that the correlation of cotton yield variation among different counties is more evident than the same correlation for hay. A possible explanation for this observation could be that, upland cotton is more specific crop and of consist yield behavior. However, for hay, we include different kinds of hay in this research, which covers alfalfa hay (main), and all other kinds of hay. So the diversity in hay category could be a reason for the heterogeneity in the regional correlation of hay yield variation.

Unfortunately, under the framework of a SUR model with identical number of observations, it is not possible to expand both the cross section and time series dimension of the data. For example, for cotton the production data has been collected since 1935 in most of counties, but the production data of hay has only been available after 1965. So expanding our method to a SUR model with unequal observations and heteroskedasticity becomes an alternative and feasible way to improve this research in future work. As we have already mentioned in chapter 2 that, another way to implement the yield forecasting is to employ a multivariate GARCH model. Which could be an interesting framework to follow in future research, and it could be used as a comparison for the results from SUR model. However, the implication of this research may still be valuable and practical in: (1) resource allocation in agriculture production, especially the water resource; (2) optimization of planting structure (crop choosing) in different regions conditional on the available resources; (3) management of risk in both agricultural production and agribusiness based on yield forecasting.

Reference

Adams, R. M., B. H. Hurd, S. Lenhart, and N. A. Leary. “Effects of Global Change on Agriculture: an interpretative review.” *Journal of Climate Research* 11(1) (1998): 19-30.

Adams, R. M., C. Chen, Bruce A. McCarl, and R. F. Weiher. “The Economic Consequences of ENSO Events for Agriculture.” *Climate Research* 13 (1999): 165-172.

Aitken, A. C. “On Least Squares and Linear Combinations of Observations.” *Proceedings of the Royal Society of Edinburgh* 55 (1935): 42-48.

Aitken, A. C., and H. Silverstone. “On the Estimation of Statistical Parameters.” *Proceedings of the Royal Society of Edinburgh* 61 (1942): 186-194.

Baltagi, Badi H. *Econometrics*, Spriner, 1998, 255-257.

Bauwens, L., S. Laurent, and J. V. Rombouts. “Multivariate GARCH models: a survey.” *Journal of Applied Econometrics* 21 (2006): 79-109.

Chen, C., Bruce A. McCarl, and David E. Schimmelpfennig. “Yield Variability as Influenced by Climate: A Statistical Investigation.” *Climatic Change* 66, No. 1-2 (2004): 239-261.

Deschênes, O., and M. Greenstone. “The Economic Impacts of Climate Change: Evidence from Agricultural Output and Random Fluctuations in Weather.” *American Economic Review* 97, Issue: 1 (2007): 354-385.

Ferris, J. N. “Forecasting World Crop Yields as Probability Distributions”. International Association of Agricultural Economists in its series 2006 Annual Meeting, Queensland, Australia, No. 25649, 12-18, August 2006.

Gardner, Bruce L. *American Agriculture in the Twentieth Century: How It Flourished and What It Cost*, Cambridge, MA: Harvard University Press, 2002.

Hwang, H. “Estimation of a Linear Sur Model with Unequal Numbers of Observations.” *The Review of Economics and Statistics* 72, No. 3 (1990): 510-515.

Isik, M., and S. Devadoss. “An Analysis of the Impact of Climate Change on Crop Yields and Yield Variability.” *Applied Economics* 38 (2006): 835-844.

Just, R. E., and R. D. Pope. “Stochastic Specification of Production Functions and Economic Implications.” *Journal of Econometrics* 7 (1978): 67-86.

Kennedy, P. *A guide to econometrics*, 3rd edition, Oxford: Blackwell, 1992.

Kmenta, J. *Elements of Econometrics,* Second Edition, Macmillan Publishing Co., 1986, 637- 647.

McCarthy, J. J., O. F. Canziani, N. A. Leary, D. J. Dokken, and K. S. White. *Climate Change 2001: Impacts, Adaptation, and Vulnerability, Intergovernmental Panel on Climate Change*, Cambridge University Press, 2001.

Mearns, L. O., C. Rosenzweig, and R. Goldberg. “Mean and Variance Change in Climate Scenarios: Methods, Agricultural Applications and Measures of Uncertainty.” *Climate Change* 35 (1997): 367-396.

Mendelsohn, R., W. D. Nordhaus, and D. Shaw. “Climate Impacts on Aggregate Farm Value: Accounting For Adaptation.” *Agricultural and Forest Meteorology* 80 (1996): 55-66.

Mendelsohn, R., W. D. Nordhaus, and D. Shaw. “The Impact of Global Warming on Agriculture: A Ricardian Analysis.” *American Economic Review* 84, No.4 (1994): 753-771.

Riha, S. J., D. S. Wilks, and P. Simoens. “Impact of Temperature and Precipitation Variability on Crop Model Predictions.” *Climatic Change* 32, No.3, (1996): 293-311.

Rubin D. B. “Inference and Missing Data.” *Biometrika* 63 (1976): 581-592.

Rubin D. B. *Multiple Imputation for Nonresponse in Surveys,* New York: John Wiley & Sons, Inc., 1987.

Schmidt, P. “Estimation of Seemingly Unrelated Regressions with Unequal Numbers of Observations.” *Journal of Econometrics* 5 (1977): 365-377.

Seo, S. N., and R. Mendelsohn. “An Analysis of Crop Choice: Adapting to Climate Change in South American Farms.” *Ecological Economics* 67 (2008): 109-116.

Srivastava, V. K., and D. E. A. Giles. *Seemingly Unrelated Regression Equations Models: Estimation and Inference*, New York: Marcel Dekker Inc., 1987.

Stallings, James L. “A Measure of the Influence of Weather on Crop Production.” *Journal of Farm Economics* 43, No. 5 (1961): 1153-1160.

Thomas, C. S., P. W. Skinner, A. D. Fox, C. A. Greer, and W. D. Gubler. “Utilization of GIS / GPS-based Information Technology in Commercial Crop Decision Making in

California, Washington, Oregon, Idaho, and Arizona." *Journal of Nematology* 34(3) (2002): 200-206.

Wit, A. J. W. de, and C.A. van Diepen. "Crop growth modeling and crop yield forecasting using satellite-derived meteorological inputs." *International Journal of Applied Earth Observation and Geoinformation* 10, Issue 4 (2008): 414-425.

Zellner, A. "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias." *Journal of the American Statistical Association* 57(298) (1962): 348-368.