Estimating Multivariate Yield Distributions Using Nonparametric Methods

Qiujie Zheng

School of Economic Sciences, Washington State University

H. Holly Wang

Department of Agricultural Economics

Purdue University

Qinghua Shi

Aitai School of Economics and Management

Shanghai Jiaotong Univeristy

Selected Paper prepared for presentation at the American Agricultural Economics Association Annual Meeting, Orlando, FL, July 27-29, 2008.

Estimating Multivariate Yield Distributions Using Nonparametric Methods

Abstract

Modeling crop yield distributions has been an important topic in agricultural production and risk analysis, and nonparametric methods have gained attention for their flexibility in describing the shapes of yield density functions. In this article, we apply a nonparametric method to model joint yield distributions based on farm-level data for multiple crops, and also provide a way of simulation for univariate and bivariate distributions. The results show that the nonparametric models, both univariate and bivariate, are estimated quite well compared to the original samples, and the simulated empirical distributions also preserve the attributes of the original samples at a reasonable level. This article provides a feasible way of using multivariate nonparametric methods in further risk and insurance analysis.

Key words: yield distribution, multi-variate nonparametric, China, farm-level, risks

Estimating Multivariate Yield Distributions Using Nonparametric Methods

Introduction

Modeling crop yield is an important research topic in agricultural economics, because it provides a basis for studies on factor productivity, efficiency, risk and crop insurance, adoption of new technology, and many other topics. Vast amounts of literature can be found on crop yield modeling, such as Day (1965), Nelson and Preckel (1989), Taylor (1990), Moss and Shonkwiler (1990), Goodwin and Ker (1998), Wang and Zhang (2002), and Norwood, Roberts and Lusk (2004). Although yield modeling is only the first step in some of these studies which focus on other topics such as crop insurance or productivity analysis, most of them focus solely on the yield distribution itself.

Recently, nonparametric and semi-parametric methods have gained the attention of economists because of their flexibility in describing the yield distribution (Goodwin and Ker, 1998; Ker and Goodwin, 2000; Ker and Coble, 2003; Norwood, Roberts and Lusk, 2004). Not assuming a particular functional form at a priori, nonparametric method will select the shape of the yield density function that fits the data the best. In addition, nonparametric density estimation techniques offer a consistent approach to smoothing observations and building a continuous density estimation (Goodwin and Ker, 1998).

Most of the yield modeling research, especially the nonparametric yield models, are univariate, i.e., investigating one crop at a time and are often based on aggregated yield data at county or state levels. However, the farm based profitability and risk analysis call for joint yield distributions with multiple crops at the farm level. Although ways have been developed to impose correlations between two non-normally distributed variables such as in Taylor (1990), the joint distribution is not unique by applying this method.

The objective of this article is to use farm level yield data and nonparametric methods to model the univariate and multivariate yield distributions. We also provide an algorithm for nonparametric multivariate simulation, and simulate some bivariate distributions based on the estimation. Finally, we examine and evaluate the attributes of the simulated distributions and analyze the local idiosyncrasies for three crops, wheat, rice and corn, using Chinese farm level yield data.

Literature Review

When modeling crop yields, a long period of historical data is usually needed in order to provide an adequate sample size with annual observations and to capture the extremely low yield caused by severe natural disasters with a small probability. Therefore, a trend needs to be considered because the long run development of production technology tends to move the mean yield over time. Just and Weninger (1999) showed that errors in specification of trends can influence the identification of residual distribution. In particular, they may introduce skewness and nonnormal kurtosis to the otherwise normal errors. As a result, the trend needs to be identified and removed before the residuals can be used to determine the distribution.

There are basically three types of trend models in the literature. First, deterministic trend models are the most frequently used, including linear, quadratic and other polynomial trends, logarithm trend, and exponential trends. Other exogenous technical and economic variables can also be incorporated into the deterministic trend functions in yield models (Gallagher, 1987). Second, time series models using trend conditional on past yield observations such as Autoregressive and Moving Average (ARMA) and Integrated ARMA models (Goodwin and Ker, 1998). Third, a general stochastic trend model is also used (Moss and Shonkwiler; 1993).

Once the data is detrended, there are two primary approaches to represent the residual yield distributions: parametric and nonparametric distributions. Under parametric techniques, a parametric distribution is selected at a priori, and parameters of the distribution are estimated by fitting the data into the model. Normal distributions are the most conveniently used and can be found in early works like Botts and Boles (1958). Although skewed distributions are favored in more recent studies as discussed in the following, Just and Wenninger (1999) claimed that the finding of skewed yield distributions may be results of inappropriate detrending and failure to properly model heteroskedasticity. When using flexible polynomial trends for mean yield and yield variance, they found that normality is difficult to reject.

On the other hand, many studies have supported that crop yields are skewed. Day's work (1965) on yield distributions for cotton, corn, and oats found positive skewness. Gallagher (1987) noted soybean yields are nonsymmetric and negatively skewed, when he used a gamma distribution. Nelson and Preckel (1989) confirmed negative skewness in corn yield and assumed a beta distribution. Taylor (1990) estimated multivariate nonnormal probability distributions by fitting hyperbolic tangent transformations of normal variates. Moss and Shonkwiler (1993) and Wang et al. (1998) used an inverse hyperbolic sine transformation to incorporate negative skewness in a model of corn yields. Ramirez (1997) extended Moss and

Shonkwiler's model to allow heteroskedasticity and multivariate distributions for U.S. Corn Belt corn, soybean, and wheat yields. In addition, Atwood, Shaik and Watts (2003) demonstrated cases when normality is failed to be rejected while the distribution is actually non-normal.

The nonparametric density estimation technique may offer advantages in that no argument needs to be made between symmetric or skewed distributions at a priori. It can also represent multi-mode distributions with local idiosyncrasies that may not be reflected in parametric specifications. Goodwin and Ker (1998) used nonparametric density estimation procedures to evaluate county-level crop yield distributions and then to evaluate yield risk and insurance premium rates for wheat and barley. The results showed that nonparametric methods may offer improved accuracy and thus improve the performance of crop insurance programs. Ker and Goodwin (2000) employed empirical Bayes nonparametric kernel density estimator to estimate the conditional yield densities. They found such methodological improvements can significantly aid in ameliorating the data lack problem.

Ker and Coble (2003) mentioned the correctness problem of parametric estimators and inefficiency problem of nonparametric kernel estimator, and proposed a semiparametric estimator by undertaking two simulations. Norwood, Roberts and Lusk (2004) used a semiparametric model, which is parametric for the trend model and nonparametric for the residual distribution, to model crop yields with a nonparametric kernel smoother. They compared six yield densities based on the out-of-sample forecasting performance and concluded that the best model to forecast county yields is a semiparametric model. Following these results, we use semiparametric methods here.

Data

The empirical analysis in this article is based on farm level crop yields in China, the world's largest producer and consumer of several grain crops. Rice, wheat and corn are three main crops produced in the country. As China is currently developing its crop insurance programs, it is important to examine the yield risks of these crops. For individual farm households, the average acreage of production is very small and the yield is risky. No quantitative farm level crop yield analysis has been found.

A farm household survey was conducted tracking farm level crop yields for about 12 years in Shandong Province and the Yantze River delta area in China by National Rural Fixed Observation Office. Winter wheat and summer corn are planted in rotation within one year in Shandong province, while in the Yantze River delta area, wheat and rice are planted concurrently in different fields, often in rotation with other minor crops.

In Shandong province, seventeen villages were randomly chosen in which there are approximately 40 farm households each. In the Yantze River delta area, fifteen villages were chosen with five in each province of Shanghai, Jiangsu, and Zhejang, and there are about 100 farm households in each village. Those farms with less than three years of multiple crop yield data were dropped. Finally, we have twelve villages with 479 farms total in the Shandong province for the period between 1995 and 2006, and five villages each in Shanghai and Jiangsu of the Yantze River delta area with 527 farms total for the period between 1993 and 2005 with 1994 missing. All villages in Zhejiang province of the Yantze River delta area were dropped because wheat is rarely grown there.

Method

In this section, we discuss model estimation and simulation, compare empirically the univariate and bivariate yield distributions, and provide the algorithm for multivariate nonparametric simulation.

Trend and detrend

The semiparametric approach is taken when a deterministic yield trend is identified for each village. This is because the yield trend is determined by technology and agronomic conditions. The agronomic conditions are quite heterogeneous within a province but relatively homogeneous within a village, so is the technology. The linear trend model as in (1) is adopted after higher polynomial specifications are examined and dropped.

$$y_{jit} = \alpha_j + \beta_j t + \varepsilon_{jit} \qquad (1)$$

where y represents crop yield, subscript j, i, and t denotes village, household, and year, respectively .

Model (1) is estimated for each village j, separately. Because the production scale of each farm can be quite different in any village, we calculate the weighted average of farm yield for each year to represent the village yield using the corresponding farm's planted acreage as the weight. Then the time series village yield data are used to obtain the village trend parameter, β_j through model (1). Each farm is then assumed to follow the same slope of time trend, but having individual intercept coefficient, α_{ji} .

Univariate estimation and simulation

The detrended farm yields are considered iid samples for each farm that can be used to fit nonparametric distributions. We examine the univariate model estimation and simulation first for each crop individually.

The kernel function for a zero mean random variable *e*, the detrended yield in this case, is defined as follows:

$$f(e) = \frac{1}{nh}\sum_{i=1}^{n} K(\frac{e-e_i}{h}) \qquad (2)$$

where e_i is the ith observation; n is the number of observations for a particular farm; h is a bandwidth parameter, which determines the weight to assign to neighboring observations in constructing the density and corresponds to the amount of smoothing to be done; according to Silverman's modified rule-of-thumb method, we set $h = 0.9\sigma n^{-(1/5)}$, where σ is the smaller of standard deviation and interquartile range divided by 1.34; and $K(\cdot)$ is the standard normal probability density function.

Because the nonparametric nature of the distribution, numerical analysis based on simulated empirical distribution is often needed for risk analysis to serve the need of topics introduced at the beginning of the article. Here we simulate an empirical distribution with 100,000 samples for each crop in each farm household to illustrate the procedure.

We first define an interval around mean zero with upper and lower bounds defined by the sample residuals, assuming the probability for the random variable to take a value beyond this interval is zero. We then divide the interval into 200 equal segments. Based on the calculated distribution density from (2), we can determine the counts of random numbers among the total of 100,000 that fall into each of the segment. Uniformly distributed random numbers in each segment are generated for the calculated counts. Finally, we can convert the 100,000 random numbers into yields by adding back the trend for any year of interest.

Bivariate estimation and simulation

The bivariate estimation and simulation are more complicated than the univariate ones since we need to consider the covariance between the two crops and take the two crops yields as an $n \times 2$ matrix (Takada 2001).

First we need to standardize yield residual vector e_i into z_i with zero mean and identity variance covariance matrix as:

$$z_i = \Omega^{-\frac{1}{2}}(e_i - \overline{e_i}) \qquad (3)$$

where Ω is the covariance matrix and $\overline{e_i}$ is the mean for vector e_i.

The kernel function is as follows:

$$\tilde{f}(z) = \frac{1}{nh^d}\sum_{i=1}^{n}\phi(\frac{z - z_i}{h}) \qquad (4)$$

where *n* represents the number of sample data; *d* represents dimension of data which equals to 2 in bivariate analysis; *h* represents bandwidth which is set at $h = (\frac{4}{d+2})^{\frac{1}{d+4}} n^{-\frac{1}{d+4}}$; and $\phi(z)$ is the bivariate standard Gaussian Kernel function, $\phi(z) = \frac{1}{(2\pi)^{\frac{d}{2}}}\exp(-\frac{z'z}{2})$.

Further we need to calculate local bandwidth factors λ_i as:

$$\lambda_i = (\frac{\tilde{f}(z_i)}{g})^{-\alpha} \ with \ \alpha = \frac{1}{2} \qquad (5)$$

where $\log g = \frac{1}{n}\sum_{i=1}^{n}\log(\tilde{f}(z_i))$.

Then the Adaptive Kernel Density function is as follows:

$$\hat{f}(z) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{(h\lambda_i)^d}\phi(\frac{z-z_i}{h\lambda_i}) \qquad (6)$$

Finally we transform the estimate back to the original scale for e as:

$$\hat{f}(e) = (\det\Omega)^{-\frac{1}{2}}\hat{f}(z) \qquad (7)$$

Similar to the univariate simulation procedure, we first define a rectangle interval around mean (0, 0) in 2-dimension plane representing the detrended yields of two crops. The lower and upper bounds of each dimension are again determined by the original sample yield residuals. We then divide the interval into 50 by 50 equal rectangle segments. Based on the calculated distribution density from (7), we can determine the count of random pairs among the total 100,000 pairs that fall into each of the segment. This number of uniformly distributed random pairs is then generated within each segment. Finally, we can convert the 100,000 random pairs into yields by adding back the mean and the trends for any year of interest.

Results

The distributions of farm level crop yield are estimated and simulated for the year 2007. All yields in this article are measured by kilogram per hectare (kg/ha). Because there are more than 1,000 farms in our dataset, it is impossible to report the result individually. Therefore, we report the averaged sample attributes across farms for each village.

Trends

Table 1 represents the village trends of wheat and corn yield of Shandong Province. For the wheat yield, there are seven villages out of twelve that have a significant trend most of which are positive. The positive trends suggest that yields have an increasing tendency over time, thanks to technology development in irrigation, seed, and fertilizer. One village, 3715, has a significant negative trend. It is also possible that crop yields fall over time because of water scarcity and soil degradation. While for corn yield, ten villages have significant trends. The trend for the same village 3715 is negative and significant, confirming the environmental changes impact negatively on both crops.

Table 2 represents the Yantze River delta area wheat and rice yield trends. For wheat yield, four villages out of nine have significant positive trends. For rice yield, six villages have positive significant trend. The trends are moderate compared to Shandong province. This is because rainfall is abundant in this area make irrigation less contributive. No village has significant negative trends in this area. This reveals the environmental degradation, primarily water scarcity, is more serious in Northern China.

Moments from estimations

Tables 3.1 through 4.3 report the crop yield statistics from the sample, the estimation and the simulation, for Shandong Province and the Yantze River delta area, respectively. We list the statistics from both univaraite and bivariate models for comparison purpose. Tables 3.1 and 4.1 are the sample moment statistics, which means the statistics are calculated from the detrended yields directly. Tables 3.2 and 4.2 report the estimated results, which means the statistics are calculated from the theoretical definition based on the estimated density function of each crop individually. Tables 3.3 and 4.3 include the simulated results.

The first three moments and correlations are calculated for all farms and all crops but only reported at the village average level supported with the standard deviation for each village, which gives readers a measure of heterogeneity of the farms within each village. The first two columns are village average and standard deviation for farm yield means $\bar{\mu}_w$ and σ_{μ_w}. The second two columns are the village average and standard deviation for farm yield standard deviations $\bar{\sigma}_w$ and σ_{σ_w}, which measures the risks of farm yield. The next two columns are the village average and standard deviation of farm yield skewness, $\bar{S}_w$ and σ_{S_w}. They indicate the average level and the variability of the degree of farm yield symmetry. These six columns show up again for the second crop on the right side of the table. The right most two columns in the tables are the village average and standard deviation of the farm yield correlation between the two crops which are denoted as $\bar{\rho}_{wc}$ and $\sigma_{\rho_{wc}}$. These measures are not applicable to univariate estimation or simulation, because each crop yield is independently simulated. The subscripts, *c*, *w* and *r* of the variables denote corn, wheat and rice, respectively.

Comparing the Tables 3.1 with 3.2, and 4.1 with 4.2, we find that the nonparametric univariate density functions are estimated very well in that the theoretical first three moments are very close to those from the original samples. At the village average level, the yield means are very accurate with less than 0.05% difference. The yield standard deviations are also closely estimated with less than 0.5% difference. Even the third moments estimated are quite close to the sample.

The moments calculated from the estimated bivariate density function, the fourth panel, are also quite close to the sample statistics. Especially, the correlations between the two crops resemble the sample correlation very well. The differences between the estimated density and the sample are slighter large for the bivariate model than the univariate model.

Distribution moments from simulations

Tables 3.3 and 4.3 report sample statistics from the univariate simulated distributions in the top panel, and the bivariate simulated distributions in the bottom panel.

In general, the simulated results are quite close to the estimated ones. The univariate simulated yield means are very close to the estimated means, with most of the differences from the estimated means less than 0.5%, while the bivariately simulated means are also quite close to the estimated values with most of the deviations less than 3%. The reason for the slightly lower accuracy is truncation. For 10,000 simulated yields in the univariate model and 100,000 in the bivariate model, we drop a small number of negative or unreasonably large realizations, which may result in a small difference in yield means.

The simulated yield standard deviations are smaller than the sample or estimated ones. This is a result of our simulation algorithm. When we simulate uniformly distributed yields within in each interval, we lose some variability. The finer the intervals are allocated, the more variability can be simulated.[1] The truncation mentioned above also contributes to smaller standard deviations.

From the nonzero estimation of yield skewness, we know that the yield distribution is not symmetric. We have seven villages with negative skewness versus five of those with positive skewness. However, the size of the skewness is generally small. The univariate simulation exactly reflects the original estimated skewness. The bivariate simulations reflect the original estimated skewness reasonably well. Although a few signs are different, most of them are in line with the estimated ones.

The correlations are preserved very well in the simulated bivariate yield distributions. There is a strong positive correlation between wheat and corn in Shandong province when the two crops are planted in rotations. However, the correlations are small with different signs depending on the village in Yangtze River delta area, when the two crops are planted in different fields with different growing seasons. This means bivariate nonparametric models are needed in describing and simulating joint yield distributions.

Flexibility and Local idiosyncrasies

Figures 1 and 2 show the empirical distributions from simulated detrended wheat and rice yields based on univariate model for one farm in Jiangsu and another in Shanghai in the Yantze River delta area. Both of these two distributions have two humps that cannot be captured by most conveniently available parametric distributions. Distribution of each crop in each farm shows different shapes with single or multiple humps, symmetric or skewed, little or very kurtotic. Figure 3 shows the joint distribution from simulated wheat and rice yield of a farm in Jiangsu. The surface is not smooth enough to show the local idiosyncrasies clearly because of the sample size, however, we can tell the surface does not keep a consistent concavity. Both univariate and multivariate distributions show the power of non-parametric methods over parametric methods in describing local idiosyncrasies.

Crop yields in the two regions in China

Although not a focus of this article, we can briefly introduce the empirical results of Chinese corn, wheat and rice yield distributions from empirical results.

From tables 1 through 4, we can assess the crop yield patterns numerically for the two areas in China. Over the recent decades, wheat, corn, and rice yields all increase over time in the Northern Plains and Downstream Yantze River Plains, which are two of the most important grain production areas in China, with very few exceptions. The corn yield increases a lot faster than wheat and rice. Since rice is traditionally planted in regions with plenty of precipitation, its yield increase depends mainly on technology, especially the crop breeding ability. However, corn and wheat traditionally grow in Northern China with limited precipitation. Their yield increase in recent years comes not only from biological technology, but also from the adoption of irrigation and increasing use of fertilizer. As the underground water level falls in most areas in Northern China, it will be increasingly challenging to keep the yield trend in the future.

The simulated average 2007 farm level wheat and corn yields in each village range from 3268 to 7641 kilograms per hectare and 4528 to 8161 in Shandong, for wheat and rice range from 2823 to 4783 and 7118 to 9403 in Yantze River Delta. The wheat yield is a lot higher in Shandong than in the Yantze River Delta because the weather in Shandong is more suitable for wheat planting and harvesting than in the Yantze River Delta which has shorter summer daylight hours, more humidity, and shorter growing season allocated to wheat rotation.

On the other hand, the risks associated with yield, ranking from low to high, are rice in the Yantze River Delta, corn in Shandong, wheat in Shandong, and wheat in the Yantze River Delta. The corresponding average coefficients of variation are about 9%, 13%, 17%, and 18%, respectively. Because our rice data are the annual average yield across two rotations in the same field, the intra-year weather effects can be smoothed out, resulting in lower variation. The corn has the shortest growing season, and any adverse weather can have an unforgiving effect on yield. The reason that wheat yield is more risky in the Yangtze River delta area than in Shandong is, again, that because wheat is sensitive to excess moisture. Even though, all of the crop yield risks are not very high.

The wheat and corn correlation is quite high, around 0.5 in Shandong because of the rotation in a year. The good or bad weather can affect both crops in the year. On the other hand, hardly any correlation pattern is detected for wheat and rice in the Yantze River delta area. Although the two crops grow during the same year, they tend to be planted at different plots of fields with different soil and agronomic conditions, such as the availability of water. Furthermore, the level of precipitation can affect the two crops in different ways.

Conclusion

In this article, we apply nonparametric methods to estimate univariate and bivariate farm-level crop yield distributions and simulate a crop yield series. We use the Shandong Province and the Yantze River delta area farm-level data to represent Northern China Plains and Downstream Yantze River Plains, two most important food grain production areas with distinct agronomic conditions. The estimated density functions accurately represent the original sample, and the simulated empirical distributions also preserve the attributes of the original data quite well. The results indicate that nonparametric methods are suitable and flexible to estimate the crop yield distributions especially when multiple crops are considered with correlations, and the marginal distribution of each crop has multiple humps and local idiosyncrasies.

[1] We have tried 20 by 20 rectangles for bivariate simulation with 10,000 realizations, the variances are smaller. Then we use 50 by 50 rectangles and increase the size to 100,000, the variances are bigger.

Reference

Atwood, J., S. Shaik, and M. Watts. 2003. “Are Crop Yields Normally Distributed? A Reexamination.” American Journal of Agricultural Economics 85: 888-901.

Botts, R. R., and J. N. Boles. 1958. “Use of Normal-Curve Theory in Crop Insurance Rate Making.” Journal of Farm Economics 40: 733-40.

Day, R. H. 1965. “Probability Distributions of Field Crop Yields.” Journal of Farm Economics 47: 713-41.

Gallagher, P. 1987. “U.S. Soybean Yields: Estimation and Forecasting with Nonsymmetric Disturbances.” American Journal of Agricultural Economics 69: 798-803.

Goodwin, B. K., and A. P. Ker. 1998. “Nonparametric Estimation of Crop Yield Distributions: Implications for Rating Group-Risk Crop Insurance Contracts.” American Journal of Agricultural Economics 80: 139-53.

Just, R. E., and Q. Weninger. 1999. “Are Crop Returns Normally Distributed?” American Journal of Agricultural Economics 81: 287-304.

Ker, A. P., and K. Coble. 2003. “Modeling Conditional Yield Densities.” American Journal of Agricultural Economics 85: 291-304.

Ker, A. P., and B. K. Goodwin. 2000. “Nonparametric Estimation of Crop Insurance Rates Revisited.” American Journal of Agricultural Economics 83: 463-478.

Moss, C. B., and J.S. Shonkwiler. 1993. “Estimating Yield Distributions with a Stochastic Trend and Nonnormal Errors.” American Journal of Agricultural Economics 75: 1056-62.

Nelson, C. H., and P. V. Preckel. 1989. “The Conditional Beta Distribution as a Stochastic

Production Function." American Journal of Agricultural Economics 71: 370-378.

Norwood, B., M. C. Roberts, and J. L. Lusk. 2004. "Ranking Crop Yield Models Using Out-of-Sample Likelihood Functions." American Journal of Agricultural Economics 86(4): 1032-43.

Ramirez, O. A. 1997. "Estimation and Use of a Multivariate Parameteric Model for Simulating Heteroskedastic, Correlated, Nonnormal Random Variables: The Case of Corn Belt Corn, Soybean and Wheat Yields." American Journal of Agricultural Economics 79: 191-205.

Takada, T. 2001. "Bivariate Nonparametric Density Estimation of Stock Prices and Volume." Asia-Pacific Financial Markets 8(3): 215-236.

Taylor, R. C. 1990. "Two Practical Procedures for Estimation Multivariate Nonnormal Probability Density Functions." American Journal of Agricultural Economics 72: 210-17.

Wang, H. H. and H. Zhang. 2002. "A Model Based Classification for Cross Sectional Time Series Data." The Journal of Agricultural, Biological, and Environmental Statistics 7(1): 107-127.

Wang, H. H., S. D. Hanson, R. J. Myers, and J. R. Black. 1998. "The Effects of Crop Yield Insurance Designs on Farmer Participation and Welfare." American Journal of Agricultural Economics 80: 806-820.

Table 1. Shandong Wheat and Corn Yield Trend

Village ID	Number of Households	Wheat (kg/ha)		Corn (kg/ha)	
		β_0	β_1	β_0	β_1
3702	46	6141.26 (14.77)	44.72 (0.99)	5163.95 (9.78)	150.06** (2.61)
3703	40	4696.16 (4.47)	-48.92 (-0.43)	5073.09 (4.41)	-50.10 (-0.40)
3704	28	6204.19 (7.82)	-205.53 (-1.20)	4695.50 (4.83)	173.08 (0.83)
3705	40	5361.11 (22.86)	100.11*** (3.92)	3530.89 (20.28)	211.92*** (11.17)
3707	40	6026.41 (15.58)	16.39 (0.39)	6136.52 (13.05)	83.15* (1.62)
3708	39	4673.95 (13.86)	174.90*** (4.76)	3418.97 (7.80)	302.72*** (6.34)
3709	40	4611.84 (11.61)	165.36*** (3.58)	5346.79 (12.45)	162.01*** (3.25)
3711	45	5002.37 (12.03)	28.14 (0.62)	4658.48 (9.15)	146.32** (2.64)
3712	40	4304.48 (12.60)	75.66** (2.03)	4254.01 (16.37)	208.36*** (7.36)
3713	40	4882.58 (16.25)	128.47*** (3.92)	2917.12 (2.91)	332.50*** (3.04)
3714	41	4983.26 (17.87)	173.68*** (5.86)	2113.63 (1.52)	405.93** (2.74)
3715	40	5850.77 (15.74)	-83.85** (-2.12)	6765.32 (10.44)	-140.35** (-2.04)

Note: ***, ** and * denote trend significance at 1%, 5% and 15% level respectively.

Numbers in the parentheses are the t-values of the estimates.

Table 2. Yantze River delta area Wheat and Rice Yield Trend

Village ID	Number of Households	Wheat (kg/ha)		Rice (kg/ha)	
		β_0	β_1	β_0	β_1
3101	96	3885.48 (8.45)	-64.43 (-0.91)	6182.93 (21.23)	89.43** (2.00)
3102	26	2312.30 (7.51)	165.99** (2.88)	6494.79 (9.70)	208.29* (1.66)
3103	75	3031.09 (6.73)	13.53 (0.20)	6592.69 (23.26)	57.99* (1.33)
3104	19	3278.17 (6.63)	70.98* (1.27)	6504.53 (31.79)	45.67** (1.97)
3207	96	3669.14 (12.39)	35.68 (1.06)	7650.61 (36.22)	99.47*** (4.15)
3208	60	3012.20 (6.95)	112.77** (2.29)	7618.34 (14.81)	44.67 (0.77)
3209	62	3653.37 (6.04)	-8.16 (-0.12)	7367.16 (22.84)	6.59 (0.18)
3210	52	3723.85 (7.21)	-58.67 (-1.00)	6806.01 (25.89)	58.65** (1.97)
3211	40	1955.38 (6.14)	173.54 (5.00)***	7505.42 (15.22)	-38.79 (-0.72)

Note: ***, ** and * denote trend significance at 1%, 5% and 15% level respectively.

Numbers in the parentheses are the t-values of the estimates.

3101 to 3104 are villages in Shanghai, and 3207 to 3211 are villages in Jiangsu province.

Table 3.1. Shandong Wheat and Corn Yield Sample Statistics

Village ID	Wheat						Corn						Wheat and Corn	
	$\bar{\mu}_w$	σ_{μ_w}	$\bar{\sigma}_w$	σ_{σ_w}	$\bar{S}_w$	σ_{S_w}	$\bar{\mu}_c$	σ_{μ_c}	$\bar{\sigma}_c$	σ_{σ_c}	$\bar{S}_c$	σ_{S_c}	$\bar{\rho}_{wc}$	$\sigma_{\rho_{wc}}$
3702	6801.202	403.72	892.233	390.213	0.145	0.876	7418.296	384.074	1080.815	452.517	0.127	0.835	0.531	0.333
3703	4153.334	492.381	1431.633	358.351	-0.26	0.627	4528.274	518.033	1634.109	425.473	-0.139	0.609	0.776	0.172
3704	3267.627	1063.599	892.202	1573.613	-0.076	0.475	7497.473	1545.866	1499.649	2167.009	0.01	0.676	0.126	0.716
3705	6928.482	710.097	841.629	435.444	0.112	0.641	6780.449	557.062	829.335	292.2	0.123	0.603	0.562	0.332
3707	6388.284	607.54	1033.662	479.324	0.135	0.872	7502.581	696.263	1165.382	515.951	0.124	0.815	0.477	0.37
3708	7399.242	561.755	970.561	425.113	0.003	0.635	8161.251	723.374	1430.61	865.818	0.235	0.81	0.422	0.342
3709	7121.884	444.621	797.501	253.902	-0.319	0.661	7807.883	436.381	880.937	317.632	-0.02	0.543	0.494	0.299
3711	5524.797	460.83	933.348	822	-0.071	0.897	6863.844	499.467	1270.308	404.063	-0.078	0.598	-0.08	0.35
3712	5508.789	465.766	655.827	207.094	0.078	0.586	7430.553	364.935	898.468	369.426	-0.143	1.061	0.061	0.311
3713	6846.792	267.7	761.369	310.463	-0.06	1.016	7995.644	450.716	1682.266	432.401	-0.44	0.765	0.206	0.24
3714	7640.503	326.623	702.607	370.056	-0.072	0.663	8133.366	393.683	2241.334	695.353	-0.958	0.645	0.104	0.364
3715	4596.322	173.505	613.721	194.243	-0.112	0.571	4751.547	574.614	1155.273	436.116	0.451	0.525	0.231	0.294

Table 3.2. Moments from Estimated Density Functions for Shandong Wheat and Corn Yield

Village ID	Wheat						Corn						Wheat and Corn	
	$\bar{\mu}_w$	σ_{μ_w}	$\bar{\sigma}_w$	σ_{σ_w}	$\bar{S}_w$	σ_{S_w}	$\bar{\mu}_c$	σ_{μ_c}	$\bar{\sigma}_c$	σ_{σ_c}	$\bar{S}_c$	σ_{S_c}	$\bar{\rho}_{wc}$	$\sigma_{\rho_{wc}}$
Univariate														
3702	6802.271	405.974	922.955	395.689	0.146	0.74	7417.262	382.967	1111.028	453.569	0.136	0.673	N/A	N/A
3703	4151.521	490.944	1509.274	376.523	-0.202	0.461	4528.871	519.742	1724.692	456.919	-0.113	0.448	N/A	N/A
3704	3278.396	1116.078	861.849	1491.308	-0.045	0.291	7500.093	1610.501	1440.228	2079.276	0.00	0.419	N/A	N/A
3705	6930.117	709.45	889.007	467	0.084	0.489	6780.823	556.921	879.974	310.372	0.079	0.443	N/A	N/A
3707	6387.271	604.24	1079.034	492.325	0.109	0.705	7501.621	689.885	1221.943	532.304	0.085	0.619	N/A	N/A
3708	7399.824	562.11	1023.138	446.586	0.005	0.525	8161.754	723.732	1488.148	849.096	0.197	0.708	N/A	N/A
3709	7121.967	444.671	842.782	266.197	-0.237	0.507	7806.629	437.517	924.363	327.776	-0.021	0.399	N/A	N/A
3711	5524.173	460.471	961.767	784.534	-0.05	0.822	6863.477	499.435	1338.065	420.727	-0.074	0.468	N/A	N/A
3712	5508.143	464.553	693.122	191.978	0.076	0.502	7430.63	364.69	935.373	368.017	-0.122	0.995	N/A	N/A
3713	6846.771	267.944	795.231	308.438	-0.033	0.867	7994.699	450.41	1763.602	457.933	-0.352	0.634	N/A	N/A
3714	7640.197	328.506	732.512	376.806	-0.035	0.459	8124.705	397.231	2325.573	695.449	-0.663	0.451	N/A	N/A
3715	4595.527	173.867	653.092	207.28	-0.073	0.394	4752.557	571.156	1216.556	450.06	0.353	0.46	N/A	N/A
Bivariate														
3702	6791.278	383.842	935.839	413.26	0.08	0.348	7413.321	369.792	1132.081	470.326	0.025	0.321	0.523	0.326
3703	4151.442	493.072	1499.972	415.151	-0.087	0.269	4527.937	511.379	1714.617	494.624	-0.064	0.269	0.766	0.172
3704	3258.167	2302.92	851.82	1553.705	-0.081	0.19	7925.33	3701.412	1338.718	2126.911	-0.044	0.226	0.126	0.71
3705	6911.716	701.249	887.608	455.047	0.074	0.294	6767.406	544.912	883.853	306.98	0.078	0.269	0.554	0.324
3707	6400.891	626.774	1067.395	493.913	0.027	0.364	7507.171	700.413	1206.17	529.886	0.099	0.374	0.467	0.366
3708	7390.186	549.659	1046.123	456.003	0.043	0.318	8146.27	702.14	1525.363	845.225	0.095	0.33	0.413	0.337
3709	7129.124	443.13	851.643	264.734	-0.163	0.33	7806.803	429.091	947.536	346.072	-0.038	0.274	0.485	0.298
3711	5521.692	437.049	967.25	744.037	-0.037	0.362	6859.323	489.98	1343.375	429.199	0.023	0.302	-0.079	0.339
3712	5504.234	460.923	708.815	221.934	0.054	0.284	7427.199	355.698	958.155	383.265	-0.053	0.426	0.058	0.3
3713	6849.052	257.637	819.489	298.702	-0.101	0.467	7997.425	432.371	1848.906	484.44	-0.2	0.353	0.198	0.23
3714	7639.806	320.386	699.828	328.782	-0.022	0.313	8155.269	397.226	2159.094	563.045	-0.352	0.262	0.103	0.351
3715	4593.618	171.872	639.584	201.699	0.015	0.227	4745.06	567.194	1188.635	404.226	0.185	0.27	0.224	0.289

Table 3.3. Sample Statistics from Simulated Shandong Wheat and Corn Yield Distributions

Village ID	Wheat						Corn						Wheat and Corn	
	$\bar{\mu}_w$	σ_{μ_w}	$\bar{\sigma}_w$	σ_{σ_w}	$\bar{S}_w$	σ_{S_w}	$\bar{\mu}_c$	σ_{μ_c}	$\bar{\sigma}_c$	σ_{σ_c}	$\bar{S}_c$	σ_{S_c}	$\bar{\rho}_{wc}$	$\sigma_{\rho_{wc}}$
Univariate														
3702	6802.777	431.292	685.752	292.078	0.181	0.877	7401.697	382.072	825.054	333.21	0.168	0.762	N/A	N/A
3703	4116.449	494.172	1121.01	303.175	-0.195	0.489	4526.639	546.399	1290.013	375.698	-0.117	0.457	N/A	N/A
3704	3325.57	1399.551	530.428	890.031	-0.046	0.28	7516.006	2011.405	870.273	1256.346	0.002	0.402	N/A	N/A
3705	6951.315	707.462	667.664	348.079	0.079	0.521	6791.79	562.583	664.107	233.976	0.071	0.464	N/A	N/A
3707	6382.135	594.565	787.162	347.497	0.154	0.822	7497.264	665.061	899.71	378.608	0.104	0.667	N/A	N/A
3708	7407.685	573.199	795.435	348.918	0.017	0.571	8169.23	708.609	1141.476	621.893	0.205	0.828	N/A	N/A
3709	7119.122	448.597	635.41	189.459	-0.24	0.548	7791.138	456.932	697.398	251.419	-0.024	0.414	N/A	N/A
3711	5502.706	408.257	718.586	578.579	-0.061	1.047	6862.282	510.853	1010.558	317.05	-0.084	0.536	N/A	N/A
3712	5502.78	461.921	533.299	150.518	0.095	0.567	7428.347	347.469	713.517	274.847	-0.133	1.254	N/A	N/A
3713	6839.577	275.324	610.2	222.458	-0.003	0.991	7979.252	465.153	1370.125	346.267	-0.386	0.688	N/A	N/A
3714	7631.957	347.067	521.162	248.061	-0.041	0.485	7986.663	433.295	1621.871	432.704	-0.71	0.486	N/A	N/A
3715	4580.056	188.877	488.176	151.617	-0.069	0.409	4776.659	561.35	904.233	303.341	0.39	0.563	N/A	N/A
Bivariate														
3702	6897.394	633.456	749.492	368.451	-0.036	0.163	7604.26	823.005	952.582	459.187	-0.039	0.22	0.556	0.337
3703	4072.284	670.074	1204.672	350.611	-0.035	0.163	4456.518	1015.565	1356.311	418.364	-0.04	0.273	0.771	0.171
3704	3449.062	1766.964	532.732	908.798	-0.02	0.102	7543.081	2736.19	867.578	1257.909	-0.008	0.099	0.127	0.709
3705	6971.436	794.377	708.483	360.87	0.05	0.163	6809.915	665.872	696.48	265.339	0.051	0.239	0.549	0.332
3707	6319.353	801.834	848.257	396.785	0.052	0.195	7615.257	1114.807	948.914	433.911	-0.039	0.25	0.466	0.386
3708	7487.54	739.934	837.811	391.732	0.047	0.229	8479.476	1646.151	1295.579	809.566	0.02	0.239	0.44	0.363
3709	7046.054	524.102	698.267	238.27	-0.054	0.177	7913.965	589.02	756.913	310.074	-0.153	0.243	0.487	0.309
3711	5652.613	1641.594	817.837	824.55	-0.001	0.198	6805.806	821.494	1050.504	381.027	-0.016	0.261	-0.073	0.353
3712	5550.936	690.073	547.964	193.893	-0.057	0.293	7340.972	860.707	817.716	390.412	0.029	0.229	0.075	0.335
3713	6882.219	606.658	661.623	263.944	-0.15	0.269	7739.517	1115.788	1531.16	446.104	-0.064	0.388	0.22	0.247
3714	7653.823	474.104	527.899	265.492	-0.226	0.194	7247.137	669.345	1754.773	386.335	-0.026	0.235	0.126	0.379
3715	4564.119	242.734	499.797	162.037	0.111	0.165	5038.531	683.641	939.271	313.414	-0.01	0.177	0.227	0.303

Table 4.1. Yantze River Delta Area Wheat and Rice Yield Sample Statistics

Village ID	Wheat						Rice						Wheat and Rice	
	$\bar{\mu}_w$	σ_{μ_w}	$\bar{\sigma}_w$	σ_{σ_w}	$\bar{S}_w$	σ_{S_w}	$\bar{\mu}_r$	σ_{μ_r}	$\bar{\sigma}_r$	σ_{σ_r}	$\bar{S}_r$	σ_{S_r}	$\bar{\rho}_{wr}$	$\sigma_{\rho_{wr}}$
Sample														
3101	2883.833	290.575	747.843	379.018	-0.162	0.841	7516.781	357.191	706.829	542.159	-0.114	0.801	-0.264	0.485
3102	4783.31	567.869	615.464	374.284	0.105	0.554	9403.285	516.007	665.807	255.723	0.033	0.409	0.2	0.639
3103	3256.317	802.9	1036.115	617.776	0.033	0.632	7676.03	732.529	1004.498	770.927	-0.019	0.704	-0.095	0.53
3104	4334.06	351.693	644.48	462.664	0.052	0.898	7152.321	331.108	553.225	379.331	0.112	0.804	0.21	0.527
3207	4235.015	289.939	791.063	264.209	-0.453	0.747	9207.545	417.674	723.338	455.443	-0.117	0.82	0.005	0.349
3208	4662.805	293.868	854.737	254.408	0.02	0.564	8350.593	323.892	970.135	284.023	-0.355	0.675	-0.257	0.266
3209	3564.594	453.947	1226.724	933.271	-0.346	0.616	7427.622	531.631	1040.916	1246.333	0.149	0.756	-0.33	0.349
3210	2822.928	299.719	963.097	370.632	-0.106	0.484	7783.176	340.056	689.81	484.006	-0.022	0.896	-0.26	0.405
3211	4599.937	372.452	747.918	208.705	0.173	0.67	7117.657	468.038	1251.337	554.658	-0.306	1.064	0.111	0.293

Table 4.2. Moments from Estimated Density Functions for Yantze River Delta Area Wheat and Rice

Village ID	Wheat						Rice						Wheat and Rice	
	$\bar{\mu}_w$	σ_{μ_w}	$\overline{\sigma}_w$	σ_{σ_w}	$\overline{S}_w$	σ_{S_w}	$\bar{\mu}_r$	σ_{μ_r}	$\overline{\sigma}_r$	σ_{σ_r}	$\overline{S}_r$	σ_{S_r}	$\overline{\rho}_{wr}$	$\sigma_{\rho_{wr}}$
Univariate														
3101	2882.653	294.456	775.457	398.156	-0.104	0.55	7517.248	363.297	731.989	560.81	-0.076	0.522	N/A	N/A
3102	4784.048	570.962	600.636	352.524	0.07	0.344	9403.035	515.773	652.956	249.241	0.02	0.246	N/A	N/A
3103	3256.308	801.854	1032.143	610.105	0.02	0.398	7676.468	735.262	1001.087	768.537	-0.01	0.445	N/A	N/A
3104	4334.445	357.731	656.942	476.086	0.035	0.587	7153.326	332.99	564.405	392.549	0.077	0.52	N/A	N/A
3207	4234.439	290.259	831.651	274.221	-0.31	0.514	9206.596	418.192	758.086	454.977	-0.075	0.61	N/A	N/A
3208	4662.521	293.89	905.497	257.669	0.022	0.436	8350.823	322.48	1011.427	270.808	-0.301	0.593	N/A	N/A
3209	3563.785	463.727	1296.226	961.035	-0.235	0.426	7429.701	551.05	1069.207	1260.646	0.127	0.603	N/A	N/A
3210	2823.18	300.928	1014.629	385.597	-0.068	0.324	7781.147	343.585	720.174	498.367	-0.012	0.677	N/A	N/A
3211	4600.778	294.456	775.457	398.156	-0.104	0.55	7116.456	467.074	1290.112	532.022	-0.235	0.957	N/A	N/A
Bivariate														
3101	2881.449	274.128	723.742	372.661	-0.066	0.359	7519.173	338.041	679.156	496.004	-0.013	0.332	-0.26	0.474
3102	4780.051	564.62	541.627	294.35	0.003	0.224	9398.609	501.271	587.596	219.409	0.041	0.123	0.197	0.63
3103	3257.164	798.712	936.271	527.613	-0.001	0.219	7674.629	722.208	903.309	667.185	0.018	0.227	-0.094	0.518
3104	4331.744	343.091	588.471	424.51	0.062	0.372	7149.912	320.643	503.072	338.131	0.056	0.348	0.201	0.517
3207	4242.017	287.482	806.427	257.581	-0.215	0.306	9205.689	407.898	743.769	426.165	-0.022	0.353	0.004	0.336
3208	4660.908	291.452	937.422	280.861	-0.011	0.274	8356.552	314.178	1062.996	287.828	-0.128	0.293	-0.249	0.26
3209	3569.637	426.868	1242.147	847.005	-0.289	0.298	7431.251	552.004	1058.326	1145.466	0.078	0.315	-0.317	0.341
3210	2824.143	291.181	970.787	359.707	-0.054	0.257	7782.56	319.533	692.955	471.905	0.011	0.35	-0.256	0.396
3211	4599.46	368.133	812.82	217.767	0.076	0.299	7122.423	447.864	1340.553	561.644	-0.128	0.421	0.105	0.281

Table 4.3. Sample Statistics from Simulated Yantze River Delta Area Wheat and Rice Yield Distriubtions

Village ID	Wheat						Rice						Wheat and Rice	
	$\bar{\mu}_w$	σ_{μ_w}	$\overline{\sigma}_w$	σ_{σ_w}	$\overline{S}_w$	σ_{S_w}	$\bar{\mu}_r$	σ_{μ_r}	$\overline{\sigma}_r$	σ_{σ_r}	$\overline{S}_r$	σ_{S_r}	$\overline{\rho}_{wr}$	$\sigma_{\rho_{wr}}$
Univariate simulation														
3101	2875.578	350.102	536.732	282.18	-0.104	0.591	7509.895	449.821	503.857	381.766	-0.083	0.565	N/A	N/A
3102	4791.899	596.908	375.834	212.531	0.073	0.344	9404.711	520.629	416.436	162.519	0.024	0.246	N/A	N/A
3103	3260.251	812.632	670.008	387.344	0.024	0.405	7676.037	777.988	648.767	496.05	-0.01	0.458	N/A	N/A
3104	4342.042	426.581	439.39	326.371	0.034	0.634	7167.99	380.775	376.23	263.469	0.081	0.57	N/A	N/A
3207	4217.408	306.84	604.226	192.483	-0.322	0.545	9191.939	438.287	555.479	324.742	-0.069	0.686	N/A	N/A
3208	4654.793	296.668	699.353	197.086	0.025	0.46	8345.033	313.508	784.159	204.871	-0.311	0.693	N/A	N/A
3209	3542.755	618.326	940.506	634.694	-0.234	0.469	7437.18	753.705	784.719	857.283	0.156	0.724	N/A	N/A
3210	2820.761	334.75	735.819	269.451	-0.067	0.331	7767.277	397.863	516.952	345.423	-0.005	0.789	N/A	N/A
3211	4616.075	376.448	603.657	154.734	-0.132	0.585	7105.279	444.903	977.198	381.97	-0.257	1.133	N/A	N/A
Bivariate simulation														
3101	2940.991	691.338	566.272	304.815	0.003	0.209	7466.719	840.293	528.648	434.251	-0.061	0.254	-0.261	0.478
3102	4761.532	529.573	383.522	218.417	0.006	0.077	9415.004	564.514	419.32	163.989	-0.022	0.141	0.196	0.638
3103	3265.146	806.853	684.226	400.35	0.003	0.137	7661.047	974.456	663.06	508.459	-0.006	0.148	-0.087	0.529
3104	4347.998	676.524	461.576	353.029	0.046	0.234	7228.766	798.486	382.698	267.7	0.058	0.281	0.22	0.508
3207	4163.872	469.066	650.028	228.65	-0.036	0.246	9242.866	880.479	603.938	390.994	-0.157	0.233	0.004	0.349
3208	4806.918	540.706	744.955	269.938	-0.089	0.208	8114.486	764.109	902.994	337.004	-0.012	0.221	-0.274	0.294
3209	3522.199	1401.438	949.638	669.88	0.1	0.203	7739.679	1936.009	914.932	884.074	-0.19	0.176	-0.402	0.372
3210	2776.741	562.93	765.349	277.063	0.003	0.22	7798.264	772.72	578.367	438.698	-0.048	0.189	-0.275	0.421
3211	4721.633	526.899	642.763	197.325	-0.072	0.296	6798.088	1188.128	1160.181	606.382	0.087	0.248	0.112	0.324

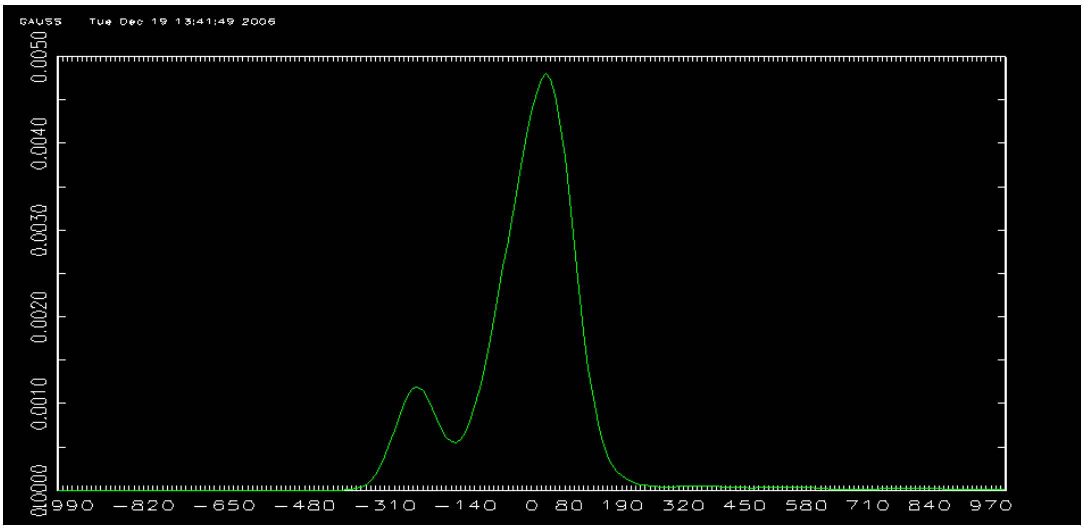

Figure 1. Simulated detrended wheat yield for farm 3207776

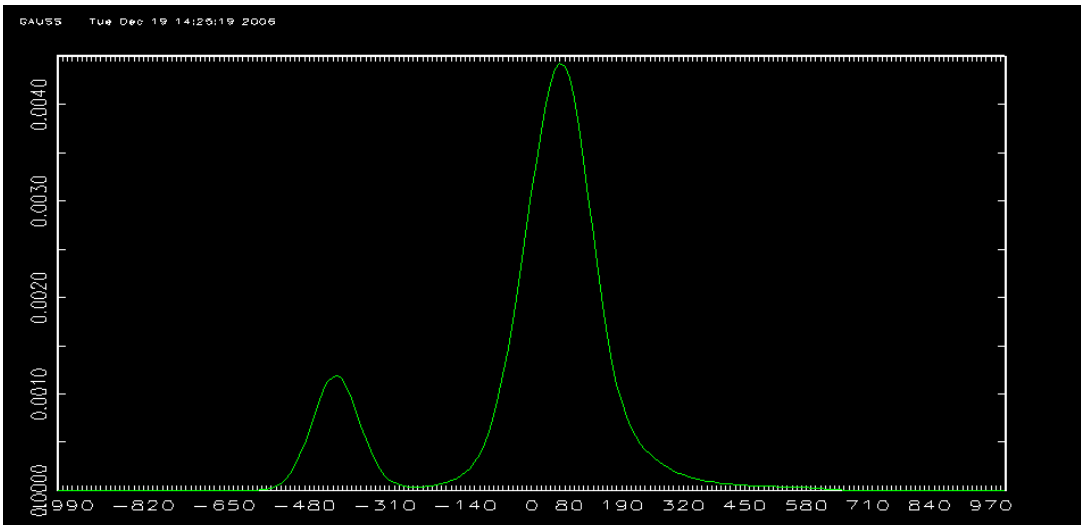

Figure 2. Simulated detrended rice yield for farm 3103794

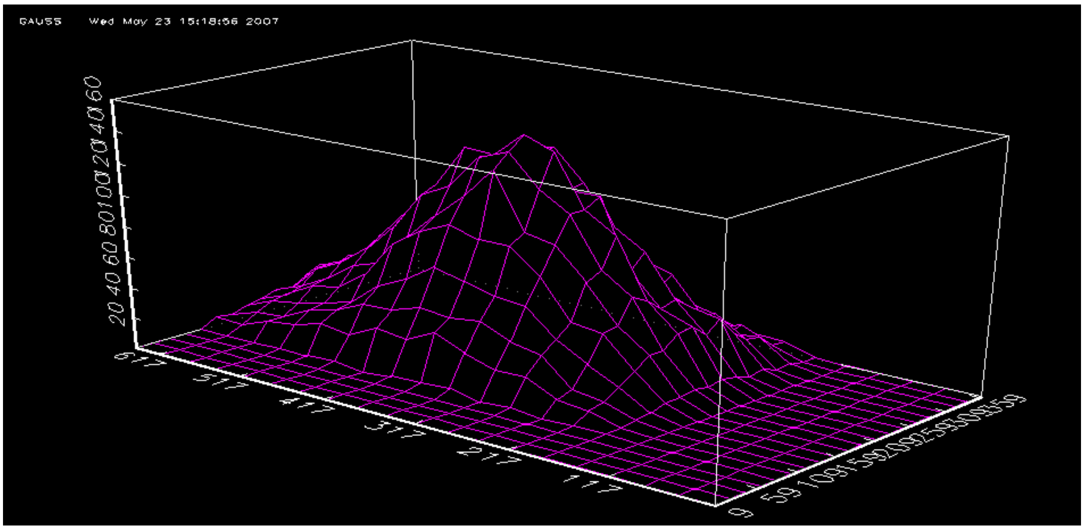

Figure 3. Simulated joint distribution of wheat and rice yields for farm 3211030