Agricultural Production, Weather Variability, and Technical Change: 40 Years of Evidence from India*

Jeffrey D. Michler[a,b], Patrick S. Ward[c], and Gerald E. Shively[d]

[a] *University of Illinois, Urbana, USA*
[b] *International Crops Research Institute for the Semi-Arid Tropics, Bulawayo, Zimbabwe*
[c] *International Food Policy Research Institute, Washington, D.C., USA*
[d] *Purdue University, West Lafayette, USA*

December 2016

Abstract

Weather variability plays a key role in determining yields in agricultural production. However, Green Revolution technologies, in the form of fertilizers, irrigation, pesticides, and drought or flood tolerant seeds, help insulate farmers against the realization of uncertain weather events. This paper poses the question: to what extent have Green Revolution technologies mitigated the impact of rainfall variability on crop production? We use a long panel of parcel level data from six villages in India that covers 50 seasons from 1975 to 2014. In our descriptive analysis we generate several stylized facts about how agricultural production in the subcontinent has changed over the last 40 years. In a regression context, we develop an innovative application of multilevel modeling to long run production data. We conclude that Green Revolution technologies have reduced the amount of weather related risk faced by farmers, even when we account for greater amounts of variation in weather due to climate change.

JEL Classification: C11, D81, O12, O13, Q16, Q12
Keywords: Weather Risk, Agricultural Production, Technical Change, Multilevel Models, Rural India

*Corresponding author email: jmichler@illinois.edu. Department of Agricultural and Consumer Economics, University of Illinois, 1301 W. Gregory Drive, Urbana, IL 61801, United States. This paper and our thinking have been shaped by conversations with Frederi Viens, Garth Holloway, and Hans Binswanger-Mkhize. We are solely responsible for any errors or misunderstandings.

1 Introduction

Estimating the effect of weather on crop output has long been a concern in agricultural and applied economics (Wallace, 1920; Stallings, 1961; Shaw, 1964; Oury, 1965). Given the non-separability of household production and consumption in developing economies, the role of weather in crop production has also been of long standing interest in international development (Rosenzweig and Binswanger, 1993; Rosenzweig and Wolpin, 1993; Townsend, 1994). More recently, the popularity of weather index insurance as a micro-development intervention has refocused interest on weather and crop yields (Barnett and Mahul, 2007; Giné et al., 2007, 2008, 2012; Cole et al., 2013, 2014). Despite nearly a century of research, very few quantitative measures exist regarding the size or significance on weather and agricultural production. The lack of concrete measures is even more acute in developing countries, such as India. As a case in point, the frequently cited "fact" that rainfall variability accounts for 50 percent or more of variability in crop yields on the subcontinent comes from a 1976 report by the National Commission on Agriculture (NCA) for India.[1] The NCA arrived at this estimate via a linear regression "with yield as the dependent variable and total rainfall during the five crop growth phases as the independent variable" (National Commission on Agriculture, 1976). Apparently, no more recent or reliable estimates exist. With increased weather variability an ever more pressing concern, especially for those without adequate risk management tools, such as smallholders in the semi-arid tropics, new methods and new estimates are need for measuring weather's affect on agricultural production.

This paper poses a simple question: how large of a role does the weather play in determining variability of agricultural production in India? Finding an answer to this question is complicated by the role of technological change. Over short time horizons, technological change can safely be assumed to be constant. Once inputs, plot characteristics, and farmer ability have been controlled for, the remaining variation in yields across years will be due to weather (Michler et al., 2015). Over long time horizons, technology is not constant and must be accurately accounted for if the causal effects of weather on yield is to be identified.

To tie our results as closely as possible to the issue of weather risk in agricultural production, we use the ICRISAT household survey data from the same Indian villages studied by Townsend (1994) and Rosenzweig and Binswanger (1993). However, as a further extension, we use an expanded panel that covers 44 cropping seasons from 1976 through 2011. We focus our analysis on the ten most common crops: castor, chickpea, cotton, groundnut, maize,

[1] Parchure (2002), Barnett and Mahul (2007), and Giné et al. (2012) all cite this figure as motivation for the need for rainfall index insurance in South Asia.

paddy rice, pigeon pea, sorghum, soybean, and wheat. This gives us a total of 14,619 parcel-level observations come from 10,578 unique parcels operated by 766 distinct households in six villages.

We begin our analysis by examining descriptive evidence which generates several stylized facts about how agricultural production in the subcontinent has changed over the last 40 years. Most importantly, mean yields have increased and the variance in crop production, measured relative to the mean, has decreased. We find that much of the increase in yields is due to increases in yields for rice, maize, and wheat. Without the largest yield gains in these three crops, much of agricultural production in India has been static. Finally, we find that the increase in purchased inputs has exceeded the increase in yields. This stylized fact implies that farmers may be faced with a profit squeeze as prices for chemical inputs were rising at a more rapid clip than output prices over this 40 year period.

To examine the different sources of yield variability we use a multilevel/hierarchical regression framework. This approach more fully accounts for the covariance structure of the data than a standard regression framework and allows us to control for inputs and time trends at the parcel-level and also to isolate the amount of yield variance due to parcel-level effects, household-level effects, and seasonal effects, the latter of which we use as a proxy for weather events. Considering all sources of yield variance, we find that seasonal variation in weather makes up only a small fraction of overall variability in yield, accounting for only 3 percent of total variance in crop yields.

2 Agricultural Production in the VDSA

2.1 Data Source

To conduct our empirical analysis, we use household data from six villages in India, two from Andhra Pradesh and four from Maharashtra. These data were collected as part of the Village Level Studies/Village Dynamics Study of South Asia conducted by the ICRISAT (VDSA, 2015). Villages were surveyed starting in 1975 with half the villages being surveyed up until 1979 while the remaining half continued to be surveyed until 1984. At that point there was a gap until enumerators returned in 1989. A second, longer gap ensued after 1989 and the panel was not picked up again until 2001. From 2001 until 2008 semi-annual surveys were conducted in the villages. Starting in 2009 data collection was high frequency, with data collected every month. While numerous studies have made use of the low frequency data (1975-2008), we combine this older data with the newly available high frequency data covering the years 2009 through 2011. These data include monthly household observations on input purchases and labor expenditure for on-farm activities, and crop yields. We aggregate all data to the seasonal level. This results in 22 years of observations (44 seasons) for half the villages and 18 years (36 seasons) for the other half (See Table 1).

The VDSA has production data on over 100 different crops, much of it household fruit and vegetable production. We focus on the ten most common crops, which in aggregate account for 57 percent of the $25,567$ parcel level observations of crop output. These crops are: castor, chickpea, cotton, groundnut, maize, paddy rice, pigeon pea, sorghum, soybean, and wheat. This provides us with $14,619$ parcel-level observations. Among the ten crops, sorghum is the most common, accounting for 31 percent of observations. Next most common are rice and cotton, accounting for 16 and 15 percent of observations, respectively. Wheat is the fourth most commonly cultivated crop, making up 9 percent of observations. Castor, chickpea, groundnut, maize, pigeon pea, and soybean make up the remaining shares with between $4 - 6$ percent each.

Our $14,619$ parcel-level observations come from $10,578$ unique parcels operated by 766 distinct households, farming across 44 seasons, in 6 villages. We exploit this nested data structure in our empirical analysis.

2.2 Descriptive Evidence

We begin with a descriptive analysis of agricultural production in India in order to draw out several stylized facts regarding how technical change has affected the sensitivity of yield to variability in weather.

There is a high degree of seasonality in the data (see Table 2). Castor, cotton, pigeon pea, and soybean are all grown in *Kharif* , with planting at the start of the monsoon in June or July and ends with post-monsoon harvesting in November. Chickpea and wheat are cultivated during *Rabi* , with planting in December or January and harvest in April or May, prior to the start of the monsoon. Groundnut, maize, rice, and sorghum are all grown in both *Kharif* and *Rabi* . Statistics describing these data, by season, are presented in Table 3. Using the Mann-Whitney-Wilcoxon statistic we test if inputs usage in each season are drawn from the same population.[2] In pairwise comparisons we reject the null that inputs in each season

[2]In our case the Mann-Whitney-Wilcoxon is preferred to a standard t-test. This is due to the highly skewed, non-normal distribution of both input and output data. Unlike the t-test, the Mann-Whitney-Wilcoxon test does not require the assumption of a normal distribution and is nearly as efficient as a t-test when the underlying distribution is in fact normal. We also test for differences in the distribution of inputs across season using the Kolmogorov-Smirnov test, an alternative to the Mann-Whitney-Wilcoxon test. The

come from the same distributions. It is interesting to note, though, that most but not all input use is higher in *Kharif* compared to *Rabi* . Use of labor, fertilizer, and pesticide are all higher in *Kharif* . Pesticide use in *Kharif* is particularly high, mainly due to the cultivation of cotton and soybean in that season. The only input that is used more in *Rabi* then in *Kharif* is mechanization. This is due to the need for irrigation in the dry *Rabi* season. The need for irrigation also mean that parcel areas tend to be smaller in *Rabi* when compared to *Kharif* .

Table 4 presents summary statistics by crop. Without accounting for grain density, rice, wheat, and maize have the largest yields per hectare. Pigeon pea, chickpea, and castor have the lowest yields. Rice, groundnut, and cotton are the most labor intensive crops while pigeon pea, sorghum, and soybean used the least labor. Fertilizer usage is highest for rice and wheat by a large margin while legume and seed crops tend the need much less fertilizer. The one exception is soybean, which in general used more purchased inputs then other legumes. As with fertilizer, use of mechanization is much higher with rice, reflecting the need for intensive irrigation of *Rabi* paddy rice. Pesticide use is largest for cotton and soybean while almost no pesticide is used in the cultivation of maize and sorghum. Finally, soybean, cotton, castor, and sorghum have the largest average parcel size, with parcel area all over 2 hectares. Chickpea and maize parcels both tend to be under a hectare with paddy rice having the next smallest parcel size, again reflective of the need for more intensive cultivation of rice. This leads us to our first stylized fact

Stylized Fact 1 *Over time and on average, production in* Kharif *utilizes more labor, fertilizer, and pesticide while* Rabi *utilizes more mechanization. Cultivation in* Rabi *is more intensive in terms of land use then cultivation in* Kharif *.*

Figure 1 graphs mean seasonal yield by pooling crops and calculating the arithmetic mean of yields in each season. There is a pronounced nonlinear time trend in yields. The trend has a flattened S-curve shape long associated with the diffusion of technology and commented on in the context of agricultural production as early as Shaw (1964). While the increase in yields over time is interesting, and provides descriptive support for the impact of Green Revolution technologies, are primary focus is on how yield variability has changed over time. Figure 2 charts how variability in yield has changed over time.[3] Given that mean

K-S test results are equivalent to those obtained from the Mann-Whitney-Wilcoxon.

[3]Note that we graph the standard deviation of yield instead of the variance of yield. While our primary interest is on variance (the second moment) and not standard deviations, graphing the standard deviation allows us to present the results on the same scale as the mean.

yield increases over time it is unsurprising that variability in yield, in absolute terms, also increases. Similar to the nonlinear trend observed in mean yield, the standard deviation of yield has a flattened S-curve trend.

A naive reading of Figures 1 and 2 might lead one to conclude that the Green Revolution had little impact on crop resilience to weather variability. However, much of the increase in yield variability might simply be due to the overall increase in yields. To check this, we calculate the coefficient of variation, which is simply the ratio of the standard deviation to the mean. This allows us to measure variability in crop yields in each season relative to mean yields in that season (See Figure 3). When measured relative to mean yields, seasonal variation in yield displays a strong decreasing trend. From a maximum in *Kharif* 1976 of a standard deviation nearly double the mean, the coefficient of variation decreases to less then 1 (a standard deviation less than the mean). The rate of decrease in the coefficient of variation is decreasing over time, with the largest gains made in the early years of the panel (1976-1984). Together Figures 1 - 3 provide evidence for our second stylized fact:

Stylized Fact 2 *On average, yields in India are increasing while the variability of yields is decreasing.*

While it is clear that overall agricultural yield has increased over time, this result comes from the pooling of yields from all crop types. To gain a deeper understanding of crop production in India we can disaggregate production by crop type. Figure 4 displays average seasonal yields by crop type over time. By charting mean yields on the same scale for all crops we can quickly see which crops benefited most from technical change. By far the largest beneficiary was paddy rice, which was mean yields triple. Nearly as impressive were the gains made in maize and wheat, which saw yields double over the 40 year period. Cotton and groundnuts also saw significant yield increases while the remaining crops saw only marginal increases in yields over the nearly half century study period. Given the percentage increase in the yields of rice, maize, and wheat, and given that these three crops make up a third of crop observations, we can state our third stylized fact.

Stylized Fact 3 *The overall increase in total agricultural yield has been primarily driven by increases in paddy rice yields, with with improvements in maze and wheat yields playing secondary roles.*

We can dig in further to how technical change has impacted yields for each crop by graphing the distribution of yields for each crop in the first and final survey years. In

Figure 5 we overlay two kernel densities (1975 and 2011) for yields of some of the crops to compare the distributions over time.[4] These are unconditional, of course, and dont account for any inputs, seasonal effects or district effects. However, clear trends emerge even in this univariate analysis. Changes in yield distributions take one of two forms: a discrete shift in the distribution to the right and a complete change in the distribution. Crops in the first category (discrete shift) include castor, cotton, rice, pigeon pea, sorghum, and soybean. Distributions of yields for these crops in 1975 looks remarkably similar to distributions of yields in 2011, except that the mean and variance in 2011 are significantly larger than in 1975. For these crops, it appears that the production possibility frontier has shifted outward with little change to the variability in yields. In the second category (change in the distribution) are included chickpea, groundnut, maize, and wheat. For these crops, while mean yields have significantly increase, the variance in yields has decreased. Thus, gains in yields have been accompanied by a decline in the uncertainty in yields.

Stylized Fact 4 *The effect of technical change on crops has been of two types. For some crops, technical changes has increased both the mean and variance of yields. For other crops, technical change has resulted in an increase in mean and a decrease in variance of yields.*

Shifting from an investigation of crop-specific changes, we now focus on changes to the production relationship over time. Figures 6 charts the change in yields and input use over time. In each graph we plot yields and a specific input (labor, fertilizer, mechanization, or pesticide). We then fit the data to our time trend with a linear regression line. Examining the graph in the northwest quadrant, we see that, over the 44 season, yields have outpaced labor usage in crop production. Thus, in 2011 household in India, on average, get higher yields for less hours of work than they did in 1975. This signifies a significant increase in labor productivity in agricultural production over the last 40 years. For the remaining inputs, productivity has been decreasing. When comparing fertilizer (northeast quadrant), mechanization (southwest quadrant), and pesticide (southeast quadrant) use to yields, in all cases input use has increased at a greater rate then yields. The most striking example is the increase in fertilizer and pesticide use compared to yield. From 1975 to 2011 yields have more than doubled while fertilizer and pesticide use have increased by a factor of nine. In comparison, use of mechanization has tripled over the same time frame. This leads us to our final stylized fact.

[4] Note that soybean was not cultivated in any of the six study villages until 2002. Thus, the comparison for soybean is between yields in 2002 and 2011.

Stylized Fact 5 *On average, the increase in yields has exceeded the increase in labor usage. However, the increase in purchased inputs has exceeded the increase in yields.*

With these five stylized facts in mind, we now turn to outline the econometric framework which we utilize in our multivariate analysis.

3 Econometric Framework

3.1 Ordinary Least Squares

We begin by estimating a simple linear regression for yield. Let y_{it} denote the log of yield for parcel i at time t. We estimate

$$y_{it} = X_{it}\beta + \delta_1 D_t + \delta_2 D_t^2 + s_t + \epsilon_{it} \tag{1}$$

where X_{it} is a matrix of data, β is a vector of regression coefficients associated with various crops, D_t and D_t^2 are time trends with associated coefficients δ, and s_t is a seasonal fixed effect. We assume the error term is $\epsilon_{it} \sim \mathcal{N}(0, \sigma^2)$ so that $y_{it} \sim \mathcal{N}(X_{it}\beta + \delta_1 D_t + \delta_2 D_t^2 + s_t, \sigma^2)$, where β and σ^2 are regression and variance parameters which are time independent. The inclusion of a linear and quadratic time trend are used to control, in a simple way, for technical change over time.

We note two drawbacks associated with this linear estimation of the production function. First is that we are unable to discern the role weather plays in yield variability. Equation (1) allows us to estimate the impact of parcel-level inputs on yield at each point in time, control for time trends, and the impact of a seasonal weather dummy on yield. But these variables impact mean yield; the specification does not allow us to determine the share of seasonal variability in weather on the variance of yield (σ^2). A second drawback is that OLS limits our ability to control for additional clustered effects, such as parcel- or household-level effects. While changes in season clearly impact the effectiveness of parcel-level inputs, equally relevant effects may exist at the parcel- or household-level. Some households may be more efficient in their application of labor compared to others, while some parcels may be of better quality, resulting in less need for fertilizer. Even if it were computationally feasible to estimate season-specific, household-specific, and parcel-specific parameters using OLS, such grouped data would violate the assumption of independence for all data (Corrado and Fingleton, 2012).

3.2 The Multilevel Model

A multilevel or hierarchical modeling strategy addresses the first two drawbacks associated with the standard linear approach to estimation.[5] First, multilevel models offer a natural way to assess the role of seasonal changes in weather on variation in yields by explicitly modeling the variance, not just the mean of the data. This allows us to estimate the share of total variance attributed to variability in each group. In our case, a multilevel approach also allows us to disaggregate total variance in yields into its multiple sources, so as to measure the relative contribution of seasonality and weather risk in production. Second, a multilevel approach allows us to incorporate in our regressions intercepts for each grouping of the data without adding to the computational burden and without violating independence assumptions.

For expository purposes we start with an illustrative example of a simple two-level model in which realizations of yields are grouped within seasons. Let y_{nt} denote the log of an observed yield, n, realized in season t. We estimate

$$y_{nt} = X_n\beta + \delta_1 D_t + \delta_2 D_t^2 + \alpha_t + \epsilon_{nt} \tag{2a}$$

$$\alpha_t = \mu + \nu_t \tag{2b}$$

where X_n β, D, and δ are as previously defined, and α_t is a seasonal effect that is a function of an overall mean, μ, and a random disturbance term, ν_t. We assume that $\epsilon_{nt} \sim \mathcal{N}(0, \sigma_\epsilon^2)$, $\nu_t \sim \mathcal{N}(0, \sigma_\nu^2)$, and ϵ_{nt} is independent of ν_t.

In order to make our parameter of interest explicit, we can rewrite equations (2a) and (2b) in terms of probability distribution so that

$$y_{nt} \sim \mathcal{N}(X_{nt}\beta + \delta_1 D_t + \delta_2 D_t^2 + \mu, u_n) \tag{3}$$

with $u_n \equiv \sigma_\nu^2 + \sigma_\epsilon^2$. The above distribution is obtained by substituting (2b) into (2a) and using the independence of ϵ_{nt} and ν_t. Defining the regression equation in this way highlights the very specific dispersion structure of the residual, which is where our interest lies. It also allows us to easily define the intraclass correlation coefficient (ICC)

$$\rho = \frac{\sigma_\nu^2}{\sigma_\nu^2 + \sigma_\epsilon^2}, \tag{4}$$

[5] Gelman and Hill (2007) provide an introduction to multilevel analysis.

which is similar to the proportion of explained variance in an OLS regression.

The value of a multilevel model becomes obvious as we add additional levels. In our analysis we can view each observation on yield, y_n, as coming from a parcel group i; each parcel group as being nested within a household, h; and each household as observed within a season group, t.[6] We can write the multilevel model as:

$$\textbf{Level 1} \text{ (yields}, n = 14,619) : y_n = X_n\beta + \delta_1 D_t + \delta_2 D_t^2 + \alpha_{iht} + \epsilon_n \tag{5a}$$

$$\textbf{Level 2} \text{ (parcels}, i = 10,578) : \alpha_{iht} = \alpha_{ht} + \nu_{iht} \tag{5b}$$

$$\textbf{Level 3} \text{ (households}, h = 766) : \alpha_{ht} = \alpha_t + \nu_{ht} \tag{5c}$$

$$\textbf{Level 4} \text{ (seasons}, t = 44) : \alpha_t = \mu + \nu_t \tag{5d}$$

where again X_n is a matrix of input data and β is a vector of regression coefficients associated with the various crops.

Level 1 of the model estimates the log of yield as a function of inputs, applied by household h to that specific parcel i in the given season t and an idiosyncratic error term $\epsilon_n \sim \mathcal{N}(0, \sigma_1^2)$, where σ_1^2 is a constant variance parameter which we assume does not depend on i, h, or t. Each n observation comes from a parcel cluster i which we assign a unique intercept, α_{iht}. This parcel-level intercept allows the relationship between inputs and yield to differ across parcels depending on parcel-level characteristics. While many parcel-level characteristics, such as soil color, may be observable, many others are difficult to measure or costly to observe and thus remain unobserved to the econometrician. Such characteristics include soil micro-nutrients, grade, and aeration or composition. By including a unique intercept term for each parcel we can control for these parcel-level characteristics.

Level 2 of the model groups yields within parcels. Here parcel intercepts, α_{iht}, are a function of household characteristics, α_{ht}, and a random disturbance term, $\nu_{iht} \sim \mathcal{N}(0, \sigma_2^2)$, where σ_2^2 is a constant variance parameter. We assume that the ν_{iht} terms are independent of each other and that the vectors ϵ_n and ν_{iht} are independent. The household-level intercept allows variation in parcel-level production to be dependent on household characteristics. In most production regressions there is an attempt to control for unobserved household ability through proxy variables such as age or education. The multilevel approach allows us to control for any unobserved household-level characteristics by assigning each household a

[6] Note that i, h, and t can be understood as functions of n so that each unique data point corresponds to the index n, each data point can be identified with a unique parcel i, each parcel can be identified with a unique household h, and each household exists at unique time t.

unique intercept term without the need to rely on proxies. The use of a single disturbance term for all data points n corresponding to a given parcel group (iht) further enhances this control by imposing a covariance structure which is consistent with variation at the parcel group level.

Level 3 of the model groups parcels within households. Here household-level intercepts, α_{ht}, are a function of season, α_t, and a random disturbance term, $\nu_{ht} \sim \mathcal{N}(0, \sigma_3^2)$, where σ_3^2 is a constant variance parameter.[7] The season-level intercept allows variation in household-level efficiency to depend on seasonal weather events. While household ability is often viewed as time invariant, in fact it can be viewed as time dependent since household ability may improve, for example through education or experience. Additionally, household ability may be diminished or enhanced by changes in weather. Households with experience in dealing with drought conditions may find their ability diminished by flooding or cyclones. *A priori*, there is no reason to assume that seasonality or changes in weather have a constant or stationary effect on household characteristics. By allowing household-level intercepts to vary across season, we are relaxing the assumption that household characteristics are either time invariant or affected by weather in the same way each season.

Level 4 of the model defines season-level intercepts as a function of an overall mean, μ, and a random disturbance term, $\nu_t \sim \mathcal{N}(0, \sigma_4^2)$, where σ_4^2 is a constant variance parameter.[8] Having controlled for all other potentially relevant sources of yield variability we interpret seasonal variation as coming solely from weather events.

The model represented above by equations (5a)-(5d) and its dependence/independence structures can be summarized as a single model in terms of probability distribution with a special error structure which is a sum of independent disturbance terms with a nested dependence on indexes:

$$y_n \sim \mathcal{N}(X_n\beta + \delta_1 D_t + \delta_2 D_t^2 + \mu, u_n), \tag{6}$$

where u_n is a specific covariance matrix which is the sum of four covariance matrices corresponding to the disturbance vectors ϵ_n, ν_{iht}, ν_{ht} and ν_t from each level. Equation (6) makes clear that our parameters of interest are not the additive non-interacting scale terms (the α's) but the components of the error term (σ_1^2, σ_2^2, σ_3^2, σ_4^2). In particular, our model, whether represented by equations (5a)-(5d) or by equation (6), shows that the main goal in

[7] As in the case of the level-2 disturbance terms, the terms ν_{ht} are assumed to be independent of each other. The corresponding vector ν_{ht} is assumed to be independent of ϵ_n and ν_{iht}.

[8] The same group-level independence assumptions about ν_t holds as for those at levels 1, 2 and 3.

uncertainty quantification is to estimate the random disturbance terms. Indeed, this is the key to evaluating the share of variance in yield corresponding to each level in the hierarchy.

The ICC for the four-level model is the percentage of the total variance that is explained by the variance within clusters of groups. So, the correlation between realizations of yield within the same parcel is

$$\rho(parcel) = \frac{\sigma_2^2}{\sigma_4^2 + \sigma_3^2 + \sigma_2^2 + \sigma_1^2}. \tag{7}$$

The correlation between realizations of yield within the same household is

$$\rho(household) = \frac{\sigma_3^2 + \sigma_2^2}{\sigma_4^2 + \sigma_3^2 + \sigma_2^2 + \sigma_1^2}. \tag{8}$$

And, the correlation between realizations of yield within the same season is

$$\rho(season) = \frac{\sigma_4^2 + \sigma_3^2 + \sigma_2^2}{\sigma_4^2 + \sigma_3^2 + \sigma_2^2 + \sigma_1^2}. \tag{9}$$

By construction the ICC increases as we move to higher levels of aggregation. Thus, we also calculate each level's contribution to total variance in the model. This is simply the variance at each level divided by u_n, the total variance in yields.

4 Econometric Results

We focus on a two estimations of the production function and compare point estimates across the OLS and MLM specifications. These regressions rely on the same sample and contain the same set of inputs. To account for heterogeneous input response across crops, we allow all slope and intercept estimates to vary by crop. Table 5 presents regression result from the classical OLS regression as represented by equation (1). This regression model contains fixed effects for seasons but does not account for the multilevel structure of the data. Table 6 presents regression results from the MLE estimation of the multilevel model that explicitly account for the clustering of observations at the parcel, household and season levels. The OLS estimation of the production function generates point estimates that are broadly similar in sign, magnitude, and significance to those of the MLE multilevel regressions. Results from these two regressions point to a fairly robust set of basic patterns that are repeated with few exceptions. These include (i) positive and significant production relationships between yields and measured inputs (labor, fertilizer, mechanization and pesticides), (ii) generally increasing returns to labor, (iii) diminishing returns to purchased input use, and (iv) diminishing returns to technological change. We observe only one instance in which the point estimate for an input is negative and significant (for fertilizer in the case of chickpea). There are increasing returns to labor for all crops except groundnut, soybean, and wheat. Overall returns to scale appear to vary across crops and we can reject the hypotheses of equality of coefficients for each input across crops. For most crops, the coefficient on the time trend is positive and significant while the coefficient on the squared term is negative and significant. This provides evidence in support of the S-curve in Figure 1. There are some notable exceptions: neither time trends are significant in the case of chickpea and cotton and in the case of groundnut and wheat the linear trend is not significant while the nonlinear trend is positive and significant.

Table 7 reports estimated variance parameters (Panel A), ICCs (Panel B) and variance shares (Panel C) for the multilevel MLE regression. These statistics establish the key findings that inform our insights into the role of weather variability in agricultural production. We focus attention on Panel C of Table 7, which compactly summarizes the data expressed in the upper panels of the table. Similar to Townsend (1994) and Rosenzweig and Binswanger (1993) we find that idiosyncratic risk plays a larger role in determining observed yields then covariate risk. Unlike Townsend (1994) and Rosenzweig and Binswanger (1993), we are able to quantify these differences. Reading down the rows of the table allows us to assess the decomposition of variance and, by extension, the relative importance of each level in explaining overall variance in yields. We find that 10 percent of the total variance in yields comes from between-parcel differences, 5 percent comes from between-households differences, and 3 percent is attributed to between-season differences; 82 percent of the total residual is idiosyncratic noise. As a robustness check, we estimate the same model but with an additional level - village - to account for the observation that weather occurs at the village level. The addition of a village level does not change our results in a substantial way. An intuitive interpretation of these results is that much of the differences observed in yields reflects idiosyncratic shocks uncorrelated with any of our explanatory variables or our level effects. While 82 percent may make it seem like there is a large amount of unexplained variance in our model, the reference for this value is the residual and not the total variance. In our OLS model, the $R^2 = 0.959$. Thus, the value of the residual or the fraction of unexplained variance is only 0.041. It is this unexplained variance that the MLM allows us to decompose into shares of variance from each level. Of the total unexplained variance (not the total variance), 82 percent is random noise. Of the remaining 18 percent of unexplained variance, most is due to differences between parcels, such as soil quality. Household or farmer capability is relatively unimportant in explaining differences in yields. In other words, good farmers cannot make up for bad soil but bad farmers can still prosper if they have good soil. Only 3 percent of the variability in crop yield is due to seasonal weather variation. This basic pattern highlights the relatively small importance of between-season yield variance compared with between-parcel yield variance.

5 Conclusion

Despite long standing interest in the effect of weather on crop output few reliable estimates exist of the impact of weather variability on agricultural production. This deficit of knowledge is especially pressing in the developing world, where increased weather variability and a lack of adequate risk management tools is likely to affect millions of smallholders. We address this research gap using agricultural production data covering $14,619$ parcel level observations from India over nearly a forty year period. From a descriptive analysis we establish several stylized facts on the role technical change has played in agricultural production in India. Encouragingly, yields have increased substantially over time and this increase in yields has been accompanied by a decrease in the coefficient of variation. Using a multi-level/hierarchical regression framework, we estimate the different sources of yield variance. This approach controls for crop specific inputs and time trends and also isolate the amount of yield variability attributed to parcel-level effects, household-level effects, and seasonal weather effects. Having controlled for parcel level inputs and a time trend to account for technical change, we find that the remaining variance is mostly idiosyncratic. Only a small portion of the remaining variance (3 percent) is due to variability in weather. We conclude that technical change has reduced the amount of weather related risk faced by farmers, even when we account for greater amounts of variation in weather due to climate change.

References

Barnett, B. J. and O. Mahul (2007). Weather index insurance for agriculture and rural areas in lower-income countries. *American Journal of Agricultural Economics 89*(5), 1241–7.

Cole, S., X. Giné, J. Tobacman, P. Topalova, R. M. Townsend, and J. Vickery (2013). Barriers to household risk management: Evidence from India. *American Economic Journal: Applied Economics 5*(1), 104–35.

Cole, S., D. Stein, and J. Tobacman (2014). Dynamics of demand for index insurance: Evidence from a long-run field experiment. *American Economic Review 104*(5), 284–290.

Corrado, L. and B. Fingleton (2012). Where is the economics in spatial econometrics? *Journal of Regional Science 52*(2), 329–355.

Gelman, A. and J. Hill (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. New York: Cambridge University Press.

Giné, X., R. M. Townsend, and J. Vickery (2007). Statistical analysis of rainfall insurance payouts in Southern India. *American Journal of Agricultural Economics 89*(5), 1248–54.

Giné, X., R. M. Townsend, and J. Vickery (2008). Patterns of rainfall insurance participation in rural India. *World Bank Economic Review 22*(3), 539–66.

Giné, X., J. Vickery, L. Menand, and R. M. Townsend (2012). Microinsurance: A case study of the Indian rainfall index insurance market. In C. Ghate (Ed.), *Handbook of the Indian Economy*. New York: Oxford University Press.

Michler, J. D., F. G. Viens, and G. E. Shively (2015). Risk, agricultural production, and weather index insurance in village south asia. mimeo, Purdue University.

National Commission on Agriculture (1976). Report of the National Commission on Agriculture. Technical report, Ministry of Agriculture, New Delhi.

Oury, B. (1965). Allowing for weather in crop production model building. *American Journal of Agricultural Economics 47*(2), 270–283.

Parchure, R. (2002). Varsha bonds and options: Capital market solutions for crop insurance problems. mimeo, India.

Rosenzweig, M. R. and H. P. Binswanger (1993). Wealth, weather risk, and the consumption and profitability of agricultural investments. *The Economic Journal 103*(416), 56–78.

Rosenzweig, M. R. and K. I. Wolpin (1993). Credit market constraints, consumption smoothing, and the accumulation of durable production assets in low-income countries: Investments in bullocks in India. *Journal of Political Economy 101*(2), 223–244.

Shaw, L. H. (1964). The effect of weather on agricultural output: A look at methodology. *American Journal of Agricultural Economics 46*(1), 218–230.

Stallings, J. L. (1961). Weather and crop yields. *American Journal of Agricultural Economics 43*(5), 1153–1160.

Townsend, R. M. (1994). Risk and insurance in village India. *Econometrica 62*(3), 539–91.

VDSA (2015). Village Dynamics in South Asia (VDSA) database. Generated by ICRISAT/IRRI/NCAP in partnership with national institutes in India and Bangladesh. (http://vdsa.icrisat.ac.in).

Wallace (1920). Mathematical inquiry into the effects of weather on corn yield in the eight corn belt states. *U.S. Monthly Weather Review 48*(8), 439–446.

Table 1: Villages and Years of Data Collection

State	Villages	Years	Time Obs.
Andhra Pradesh	Aurepalle	1975-1984, 1989, 2001-2011	44
	Dokur	1975-1979, 1983, 1989, 2001-2011	36
Maharashtra	Kalman	1975-1979, 1983, 1989, 2001-2011	36
	Kinkhed	1975-1979, 1983, 1989, 2001-2011	36
	Shirapur	1975-1984, 1989, 2001-2011	44
	Kanzara	1975-1984, 1989, 2001-2011	44

Note: All villages were surveyed from 1975-1979, in 1989, and from 2001-2011. Up until 2009 surveys were conducted semi-annually. Starting in 2009 surveys were conducted monthly.

Table 2: Frequency of Crop Cultivation by Season

	Total	*Kharif*	*Rabi*
castor	918	918	0
chickpea	601	0	601
cotton	2,201	2,201	0
groundnut	617	403	214
maize	504	373	131
rice	2,371	1,521	850
pigeon pea	883	883	0
sorghum	4,492	974	3,518
soybean	681	681	0
wheat	1,351	0	1,351
total	14,619	7,954	6,665

Note: Table displays the frequency of observations of crops for each season. Cells with zero values signify that no observations for that specific crop come from the season in question.

Table 3: Descriptive Statistics by Season

	Total	*Kharif*	*Rabi*	MWW-test
labor (hrs/ha)	322.7	364.2	273.2	***
	(326.5)	(320.3)	(327.1)	
fertilizer (kg/ha)	56.05	59.39	52.06	***
	(79.18)	(74.86)	(83.86)	
mechanization (Rs/ha)	46.30	44.07	48.95	***
	(107.7)	(109.1)	(105.9)	
pesticide (Rs/ha)	99.45	165.2	20.91	***
	(331.0)	(430.7)	(87.40)	
parcel area (ha)	2.006	2.062	1.938	***
	(1.845)	(1.738)	(1.965)	
number of observations	14,619	7,954	6,665	
number of parcels	10,578	6,411	5,513	
number of households	766	702	567	
number of seasons	44	22	22	

Note: Table displays means of data for entire data set as well as by season with standard deviations in parenthesis. The final column presents the results of Mann-Whitney-Wilcoxon two-sample tests for differences in distribution. The test rejects the null that each of the inputs come from the same distribution. Results are similar if a t-test or a KolmogorovSmirnov test is used. Significance of MW-tests are reported in parentheses (*p<0.1; **p<0.05; ***p<0.01).

Table 4: Descriptive Statistics by Crop

	Castor	Chickpea	Cotton	Groundnut	Maize	rice	Pigeon Pea	Sorghum	Soybean	Wheat
yield (kg/ha)	147.4	197.8	303.6	411.4	564.7	1,505	139.0	240.7	500.6	722.9
	(111.6)	(203.8)	(356.3)	(401.6)	(609.1)	(750.7)	(169.1)	(278.0)	(248.93)	(468.7)
labor (hrs/ha)	216.2	191.2	375.6	524.7	289.0	752.6	127.4	159.0	175.0	280.2
	(99.92)	(219.1)	(172.7)	(444.2)	(204.5)	(432.2)	(107.0)	(147.4)	(78.60)	(266.2)
fertilizer (kg/ha)	18.84	11.16	66.56	33.37	46.90	152.5	7.776	14.27	51.94	101.1
	(22.29)	(30.28)	(59.09)	(64.13)	(63.84)	(107.5)	(20.96)	(34.45)	(21.64)	(72.46)
mechanization (Rs/ha)	1.829	9.757	11.84	43.14	26.61	221.1	3.259	4.565	10.27	35.92
	(5.577)	(23.73)	(22.17)	(63.78)	(38.09)	(174.0)	(9.778)	(21.03)	(7.689)	(44.23)
pesticide (Rs/ha)	29.37	21.18	365.6	30.86	1.464	82.53	34.06	2.292	493.2	13.05
	(74.55)	(94.75)	(670.6)	(100.3)	(32.73)	(170.9)	(155.4)	(28.34)	(490.9)	(46.70)
area planted (ha)	2.552	0.965	2.742	1.369	0.969	1.229	1.962	2.331	2.747	1.513
	(2.025)	(1.219)	(1.887)	(1.303)	(0.654)	(0.885)	(1.716)	(2.175)	(1.821)	(1.377)
number of observations	918	601	2,201	617	504	2,371	883	4,492	681	1,351
number of parcels	748	567	1,856	565	467	1,570	854	3,614	643	1,274
number of households	180	176	301	209	160	250	263	514	170	288
number of seasons	22	22	22	43	39	44	22	44	10	22
number of villages	3	6	5	6	4	6	6	6	4	6

Note: While results are presented in two columns, all coefficients and standard errors come from a single regression. This regression is the OLS estimate of the production function presented in Equation (1) and contains covariates and season dummies. Also included in the regression but not reported are ten crop specific intercepts. Standard errors are reported in parentheses (*p<0.1; **p<0.05; ***p<0.01).

Table 5: Results of OLS Production Function Regressions

	Dependent variable: log yield		
castor		rice	
log labor	1.708***	log labor	1.399***
	(0.092)		(0.056)
log fertilizer	0.084**	log fertilizer	0.193***
	(0.035)		(0.021)
log mechanization	0.225***	log mechanization	0.058***
	(0.056)		(0.016)
log pesticides	−0.005	log pesticides	0.036***
	(0.024)		(0.011)
time trend	−42.68***	time trend	−42.63***
	(5.116)		(5.139)
time trend^2	2.134***	time trend^2	2.158***
	(0.258)		(0.259)
chickpea		pigeon pea	
log labor	1.260***	log labor	1.421***
	(0.067)		(0.061)
log fertilizer	−0.001	log fertilizer	−0.201***
	(0.042)		(0.031)
log mechanization	0.059	log mechanization	0.208***
	(0.040)		(0.045)
log pesticides	0.065*	log pesticides	−0.020
	(0.040)		(0.028)
time trend	−42.93***	time trend	−42.34***
	(5.140)		(5.139)
time trend^2	2.171***	time trend^2	2.142***
	(0.259)		(0.259)
cotton		sorghum	
log labor	1.503***	log labor	1.493***
	(0.068)		(0.031)
log fertilizer	0.097***	log fertilizer	0.058***
	(0.019)		(0.012)
log mechanization	−0.001	log mechanization	0.165***
	(0.030)		(0.018)
log pesticides	0.117***	log pesticides	−0.025
	(0.014)		(0.026)
time trend	−42.88***	time trend	−42.46***
	(5.139)		(5.139)
time trend^2	2.161***	time trend^2	2.150***
	(0.259)		(0.259)
groundnut		soybean	
log labor	0.998***	log labor	0.399***
	(0.093)		(0.115)
log fertilizer	0.041	log fertilizer	0.054
	(0.026)		(0.047)
log mechanization	0.009	log mechanization	0.626***
	(0.030)		(0.103)
log pesticides	−0.008	log pesticides	−0.001
	(0.029)		(0.026)
time trend	−43.02***	time trend	−40.35***
	(5.139)		(5.241)
time trend^2	2.175***	time trend^2	2.092***
	(0.259)		(0.260)
maize		wheat	
log labor	1.712***	log labor	0.450***
	(0.101)		(0.057)
log fertilizer	0.050*	log fertilizer	0.277***
	(0.030)		(0.021)
log mechanization	−0.053	log mechanization	0.043*
	(0.037)		(0.025)
log pesticides	−0.012	log pesticides	0.087***
	(0.169)		(0.022)
time trend	−41.54***	time trend	−42.85***
	(5.140)		(5.140)
time trend^2	2.137***	time trend^2	2.170***
	(0.259)		(0.259)
Observations		14,619	
R^2		0.959	

Note: While results are presented in two columns, all coefficients and standard errors come from a single regression. This regression is the maximum likelihood estimate of the multilevel model and contains covariates and data clustered at the season, household, and parcel. Also included in the regression but not reported are ten crop specific intercepts. Standard errors are reported in parentheses (*p<0.1; **p<0.05; ***p<0.01).

Table 6: Results of MLM Production Function Regressions

Dependent variable: log yield			
castor		rice	
log labor	1.704***	log labor	1.449***
	(0.091)		(0.057)
log fertilizer	0.078**	log fertilizer	0.159***
	(0.035)		(0.022)
log mechanization	0.244***	log mechanization	0.069***
	(0.056)		(0.018)
log pesticides	−0.012	log pesticides	0.043***
	(0.024)		(0.011)
time trend	2.069*	time trend	3.186***
	(1.243)		(0.879)
time trend^2	−0.268***	time trend^2	−0.049
	(0.067)		(0.043)
chickpea		pigeon pea	
log labor	1.211***	log labor	1.424***
	(0.066)		(0.060)
log fertilizer	−0.002	log fertilizer	−0.185***
	(0.043)		(0.032)
log mechanization	0.064	log mechanization	0.225***
	(0.040)		(0.045)
log pesticides	0.084**	log pesticides	−0.023
	(0.040)		(0.028)
time trend	−0.864	time trend	4.279***
	(1.229)		(1.166)
time trend^2	0.083	time trend^2	−0.177***
	(0.068)		(0.059)
cotton		sorghum	
log labor	1.548***	log labor	1.457***
	(0.068)		(0.031)
log fertilizer	0.076***	log fertilizer	0.040***
	(0.020)		(0.012)
log mechanization	0.001	log mechanization	0.174***
	(0.030)		(0.018)
log pesticides	0.121***	log pesticides	−0.027
	(0.014)		(0.026)
time trend	−0.040	time trend	4.077***
	(0.950)		(0.750)
time trend^2	0.012	time trend^2	−0.122***
	(0.047)		(0.039)
groundnut		soybean	
log labor	0.985***	log labor	0.481***
	(0.091)		(0.115)
log fertilizer	0.043	log fertilizer	0.050
	(0.026)		(0.047)
log mechanization	0.028	log mechanization	0.585***
	(0.030)		(0.102)
log pesticides	−0.005	log pesticides	−0.004
	(0.029)		(0.026)
time trend	−1.345	time trend	26.622***
	(1.188)		(9.988)
time trend^2	0.128**	time trend^2	−0.699***
	(0.064)		(0.251)
maize		wheat	
log labor	1.690***	log labor	0.491***
	(0.099)		(0.057)
log fertilizer	0.040	log fertilizer	0.265***
	(0.030)		(0.022)
log mechanization	0.004	log mechanization	0.036
	(0.037)		(0.025)
log pesticides	−0.017	log pesticides	0.095***
	(0.168)		(0.022)
time trend	13.102***	time trend	0.507
	(1.315)		(1.046)
time trend^2	−0.237***	time trend^2	0.082*
	(0.068)		(0.049)
Observations		14,619	
Log Likelihood		-24,333	
Akaike Inf. Crit.		48,814	
Bayesian Inf. Crit.		49,376	

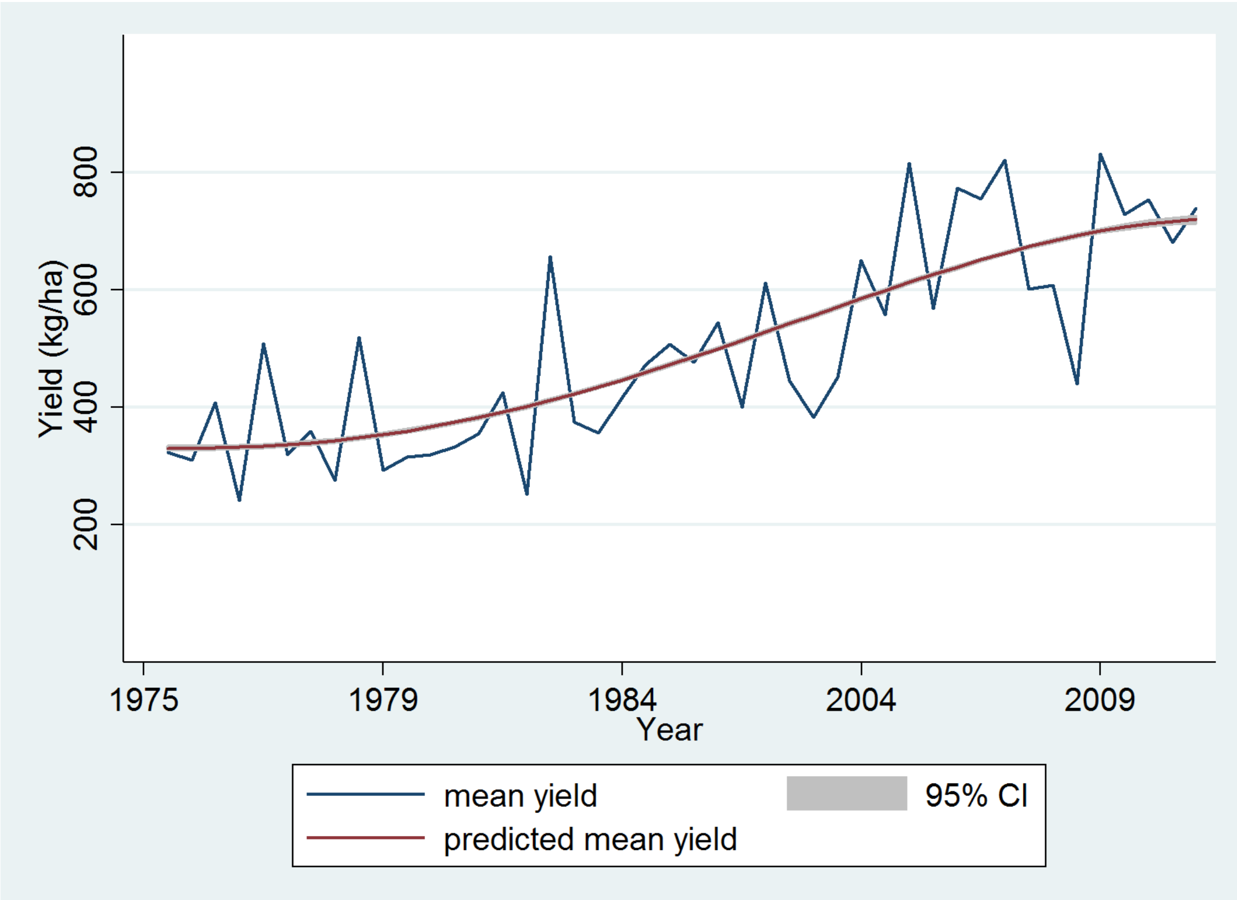
Note: While results are presented in two columns, all coefficients and standard errors come from a single regression. This regression is the maximum likelihood estimate of the multilevel model and contains covariates and data clustered at the season, household, and parcel. Also included in the regression but not reported are ten crop specific intercepts. Standard errors are reported in parentheses (*p<0.1; **p<0.05; ***p<0.01).

Table 7: Estimated Variance, ICCs, and Variance Shares from Multilevel Regressions

	w/o village	*w/ village*
Panel A: *Variance Parameter Estimates*		
parcel (σ_2^2)	0.161	0.161
household (σ_3^2)	0.091	0.063
season (σ_4^2)	0.042	0.040
village (σ_5^2)		0.060
idiosyncratic (σ_1^2)	1.385	1.383
Panel B: *Intraclass Correlation Coefficients*		
parcel	0.096	0.094
household	0.150	0.131
season	0.175	0.155
village		0.189
Panel C: *Shares of Variance From Each Level*		
parcel	10%	09%
household	05%	04%
season	03%	02%
village		03%
idiosyncratic	82%	81%

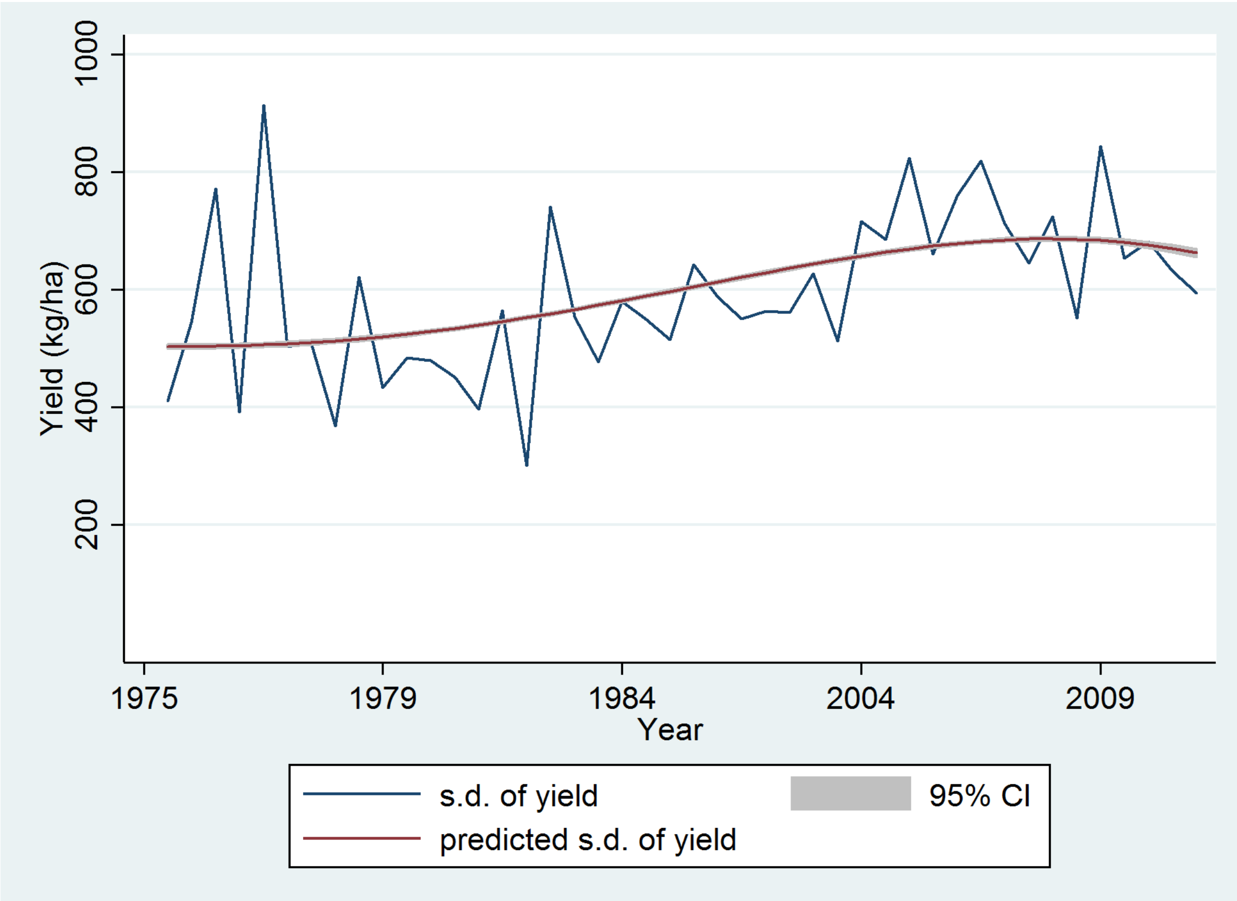
Note: In Panel A, estimates of the variance parameters on each level's residuals come from estimation of the model reported in in Table 6. Variances at levels 2, 3, and 4 ($\sigma_2^2, \sigma_3^2, \sigma_4^2$) represent the variance in crop yield that comes from the corresponding level. The final variance parameter (σ_1^2) corresponds to the idiosyncratic or unexplained portion of the model. Intraclass correlation coefficients in Panel B. Panel C decomposes the ICC into percent of variance accorded to each level.

Figure 1: Mean of Seasonal Yield



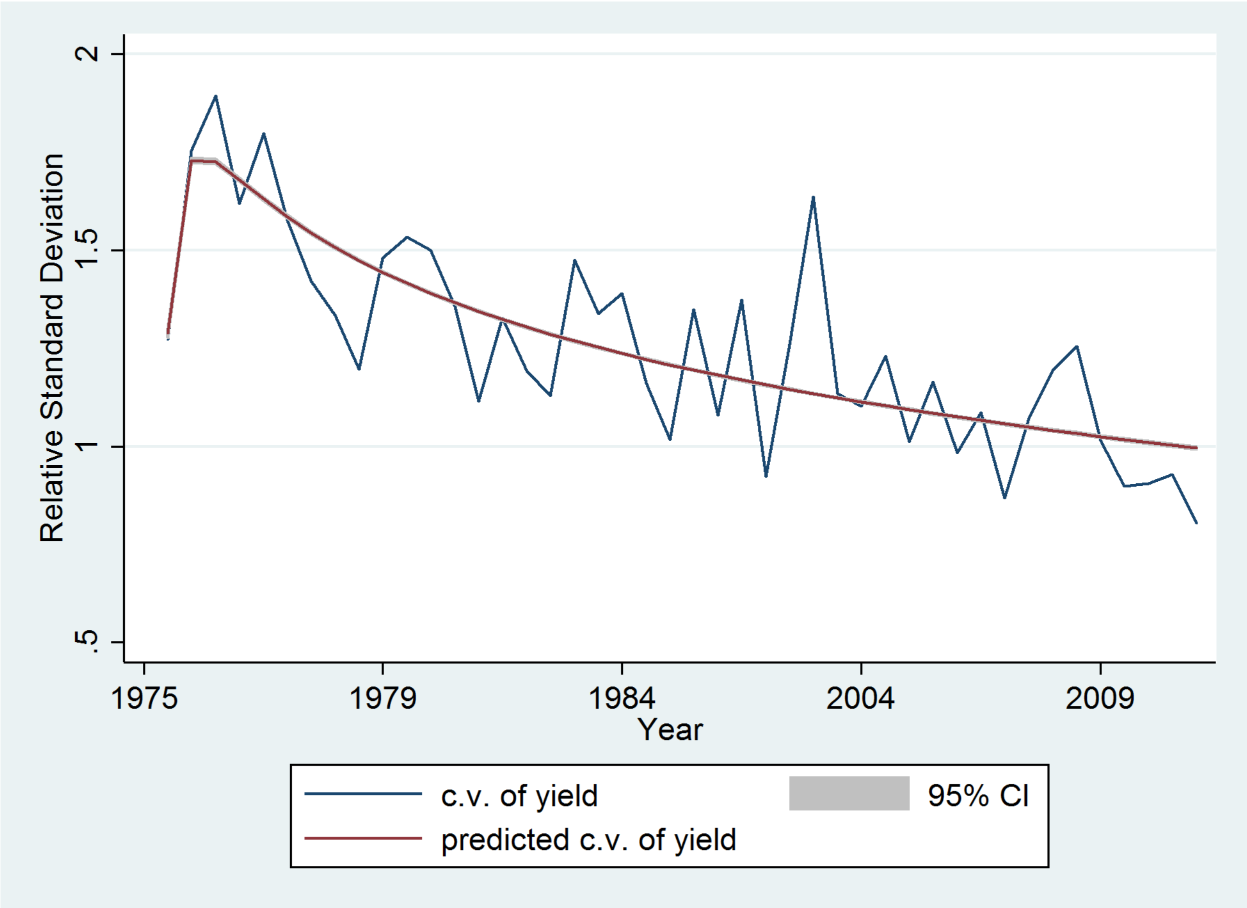
Note: Mean seasonal yield in each season is calculated as the arithmetic mean of output per hectare for all crops grown in that season.

Figure 2: Standard Deviation of Seasonal Yield



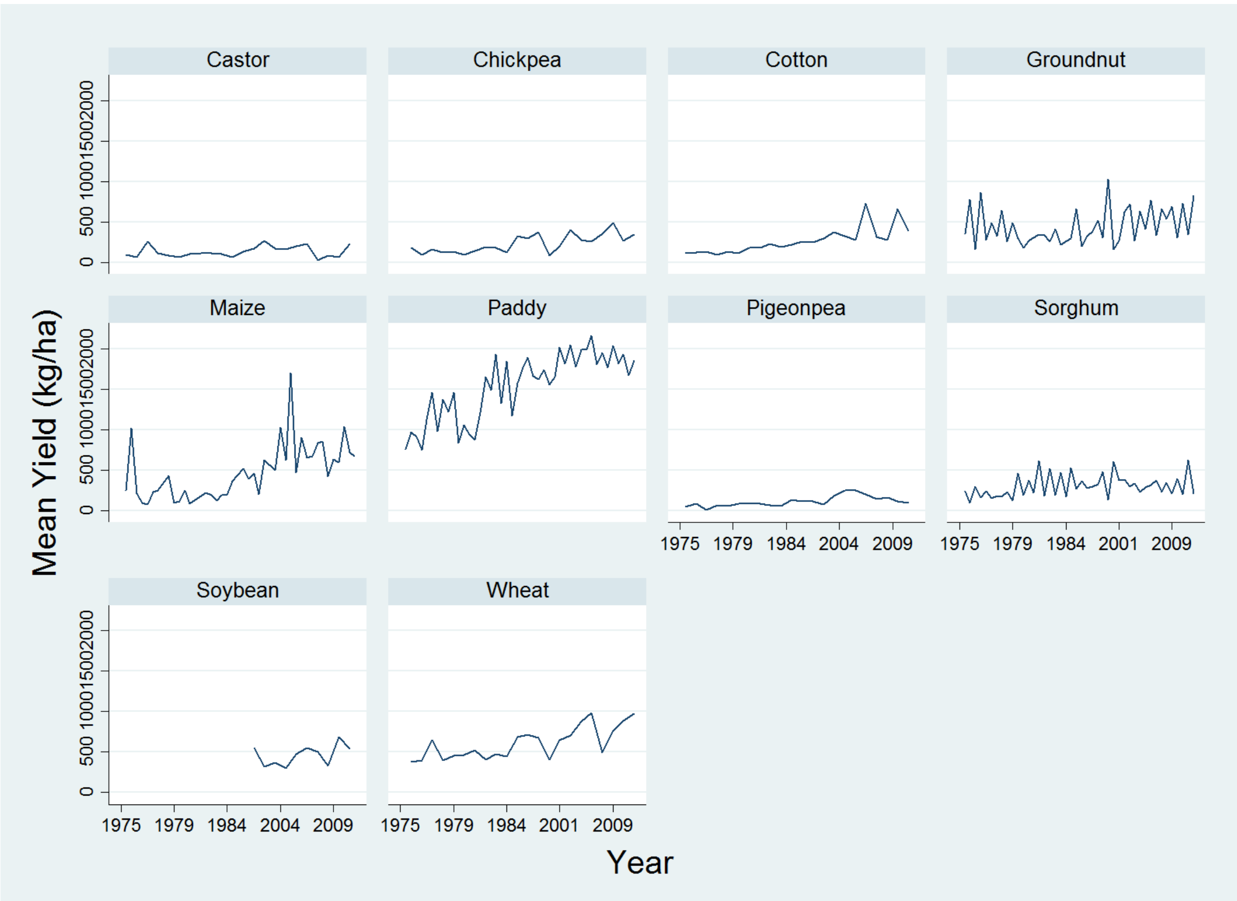
Note: Standard deviations of seasonal mean yield in each season is calculated as the standard deviation of output per hectare for all crops grown in that season.

Figure 3: Coefficient of Variation of Seasonal Yield



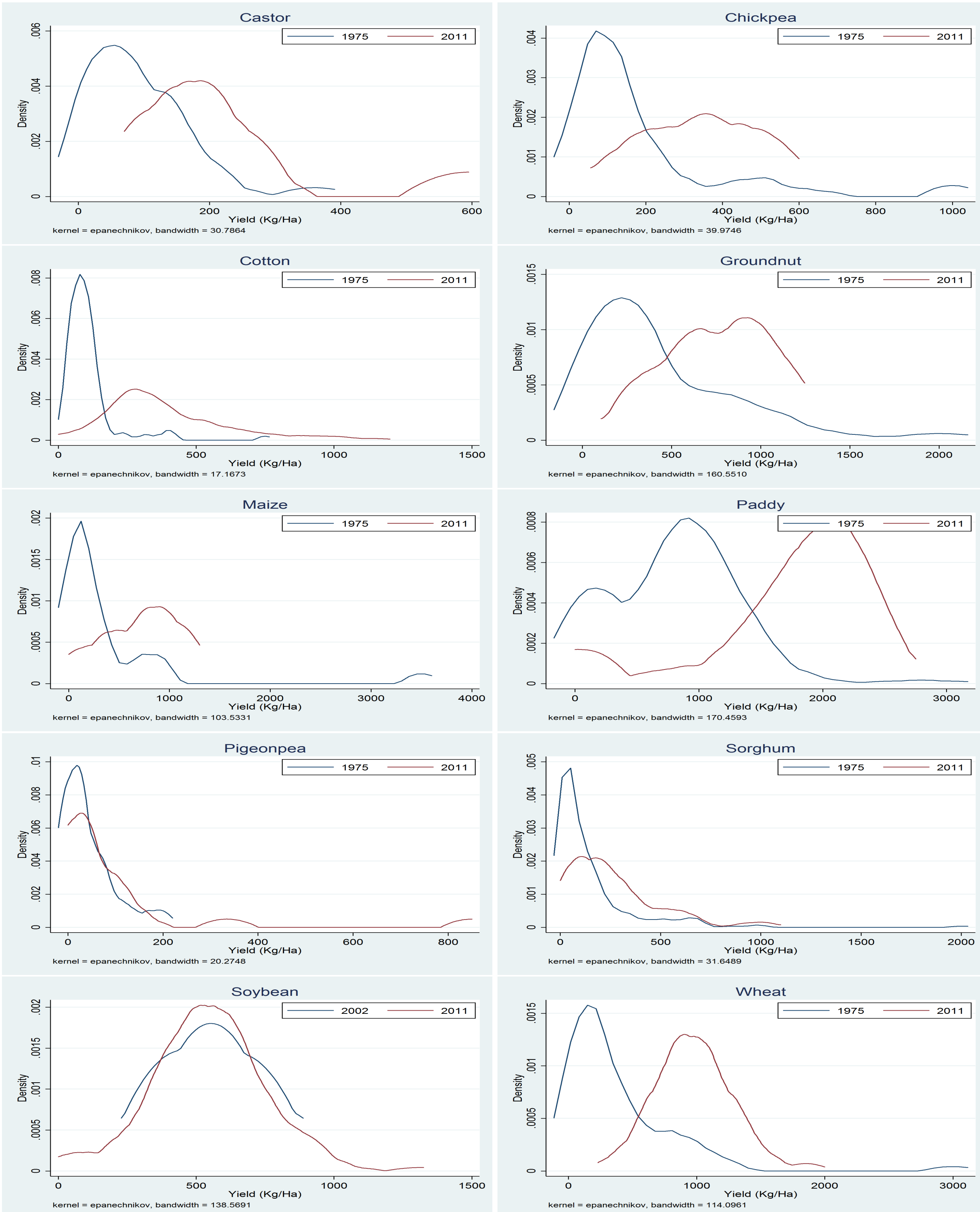
Note: Coefficient of variation of seasonal yield in each season is calculated as the ratio of the standard deviation to the arithmetic mean of output per hectare for all crops grown in that season.

Figure 4: Mean Seasonal Yield by Crop



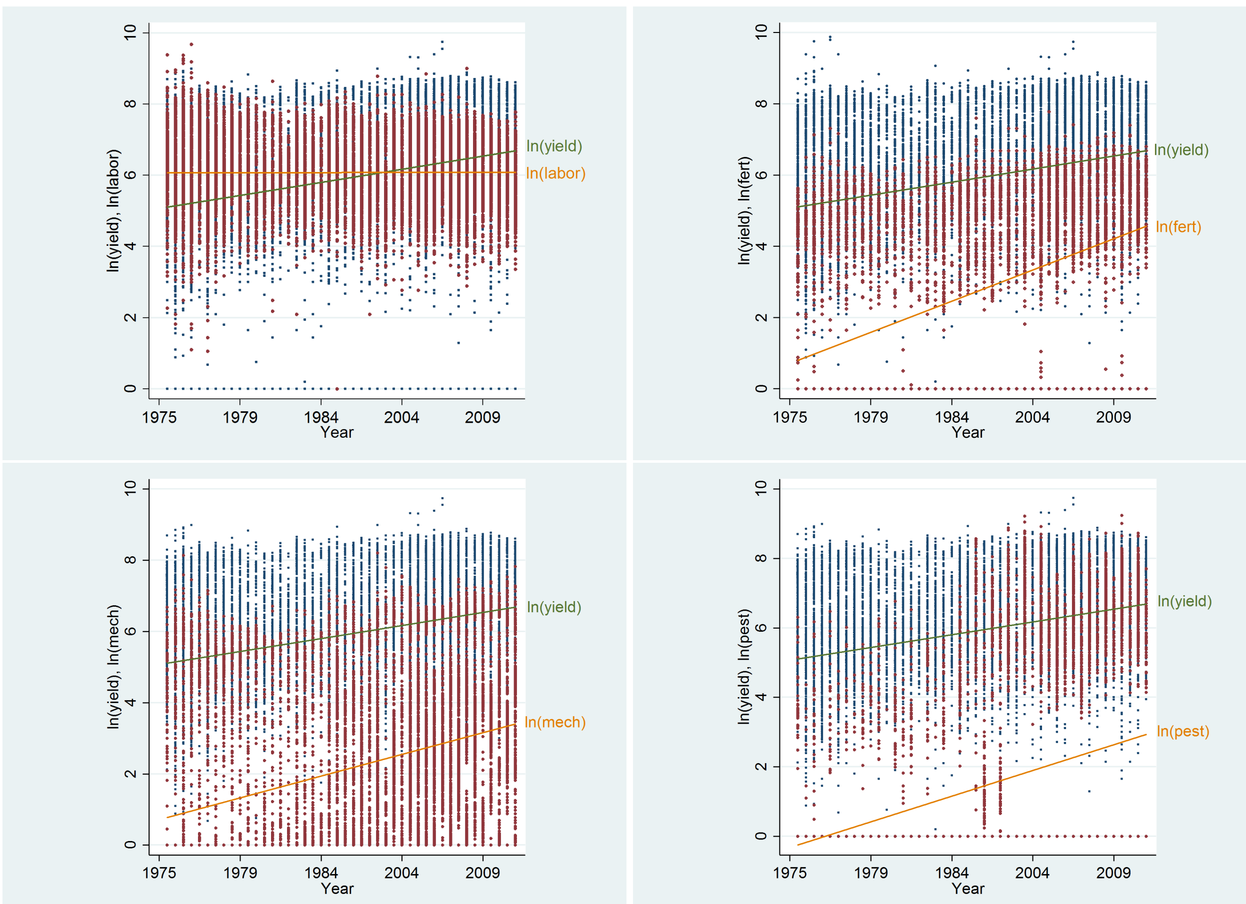
Note: Mean seasonal yield in each season is calculated as the arithmetic mean of output per hectare for each crop in that season. All graphs utilize the same scale.

Figure 5: Kernel Density of Yields by Crop (1975 & 2011)



Note: Graphs are kernel density plots by crop of yields in first and last year of survey (1975 & 2011). The exception is soybeans, which were not cultivated until 2002.

Figure 6: Yield and Input Use



Note: Hollow blue circles represent parcel level log of yield while hollow red diamonds represent parcel level log of labor, fertilizer, mechanization, and pesticide, respectively. The green line is the linear trend line of yields over time. The orange line is the linear trend of the specific input over time.