INTERNATIONAL FOOD
POLICY RESEARCH INSTITUTE
sustainable solutions for ending hunger and poverty
Supported by the CGIAR

IFPRI Discussion Paper 01170

April 2012

Measuring the Contribution of Bt Cotton Adoption to India's Cotton Yields Leap

Guillaume P. Gruere

Yan Sun

Environment and Production Technology Division

INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE

The International Food Policy Research Institute (IFPRI) was established in 1975. IFPRI is one of 15 agricultural research centers that receive principal funding from governments, private foundations, and international and regional organizations, most of which are members of the Consultative Group on International Agricultural Research (CGIAR).

PARTNERS AND CONTRIBUTORS

IFPRI gratefully acknowledges the generous unrestricted funding from Australia, Canada, China, Denmark, Finland, France, Germany, India, Ireland, Italy, Japan, the Netherlands, Norway, the Philippines, South Africa, Sweden, Switzerland, the United Kingdom, the United States, and the World Bank.

AUTHORS

Guillaume P. Gruere, International Food Policy Research Institute
Research Fellow, Environment and Production Technology Division
g.gruere@cgiar.org

Yan Sun, International Food Policy Research Institute
Research Analyst, Environment and Production Technology Division
y.sun@cgiar.org

Notices

IFPRI Discussion Papers contain preliminary material and research results. They have been peer reviewed, but have not been subject to a formal external review via IFPRI's Publications Review Committee. They are circulated in order to stimulate discussion and critical comment; any opinions expressed are those of the author(s) and do not necessarily reflect the policies or opinions of IFPRI.

Copyright 2012 International Food Policy Research Institute. All rights reserved. Sections of this material may be reproduced for personal and not-for-profit use without the express written permission of but with acknowledgment to IFPRI. To reproduce the material contained herein for profit or commercial use requires express written permission. To obtain permission, contact the Communications Division at ifpri-copyright@cgiar.org.

Contents

Tables

Figures

ABSTRACT

While a number of empirical studies have demonstrated the role of Bt cotton adoption in increasing Indian cotton productivity at the farm level, there has been questioning around the overall contribution of Bt cotton to the average cotton yield increase observed these last ten years in India. This study examines the contribution of Bt cotton adoption to long- term average cotton yields in India using a panel data analysis of production variables in nine Indian cotton-producing states from 1975 to 2009. The results show that Bt cotton contributed 19 percent of total yield growth over time, or between 0.3 percent and 0.4 percent per percentage adoption every year since its introduction. Besides Bt cotton, the use of fertilizer and the increased adoption of hybrid seeds appear to have contributed to the yield increase over time. However, if official Bt cotton adoption contributed to increased yield after 2005, unofficial Bt cotton might also have been part of the observed increase of yields starting in 2002, the year of its official introduction in India.

Keywords: Bt cotton, India, panel data, multiple imputation

ACKNOWLEDGMENTS

The authors would like to thank Sangeeta Bansal and Anchal Arora for their support in collecting the Cost of Cultivation data and Tingju Zhu for providing rainfall data. They would like to further thank the International Cotton Advisory Committee for sharing international data on cotton. The study was supported by the South Asia Biosafety Program, a project funded by the U.S. Agency for International Development. Views expressed in this article are the authors’ alone.

1. INTRODUCTION

In the past 10 years, India, formerly self-sufficient or net- cotton- importing nation, has become the world's second producer and exporter of cotton, by doubling its production in five years (for example, Gruere, Mehta-Bhatt and Sengupta 2008). While the area of production has increased relatively marginally, cotton productivity jumped significantly from 2002 onward, in coincidence with the official introduction of transgenic- insect- resistant Bt cotton. Many observers believe that the adoption of Bt cotton was actually the engine of cotton productivity growth in India. For example, on October 19, 2011, India's minister of agriculture, Sharad Pawar, noted that seed cotton yields had gone from 1.5 quintal (150 kilograms per hectare) to 5 quintal (500 kilograms per hectare) thanks to genetically modified cotton adoption (Press Trust of India, 2011).

Indeed, despite ongoing controversies among civil society groups, more than 25 farmer surveys in Indian cotton-producing states have demonstrated the overall positive impact of the technology on yields, even if with significant variance across locations, varieties, and over time (for example, Karihaloo and Kumar 2009; Rao and Dev 2010; Huang et al. 2011; Raney and Matuschke 2011). Together with total insecticide use reduction (for example, International Cotton Advisory Committee 2010), and despite higher seed costs, farmers in these studies were found to benefit significantly from adopting Bt cotton (for example, Finger et al. 2011; Gruere and Sengupta 2011). Furthermore, additional empirical studies in India have shown that Bt cotton reduced the use of pesticides, and thereby had positive health effects on cotton farmers (Kouser and Qaim 2011), that additional harvests resulted in increased women's labor opportunities (Subramanian and Qaim 2009), that it did contribute to poverty reduction (Subramanian and Qaim 2010), and that it is likely not related to a claimed increase in Indian farmer suicides (Gruere and Sengupta 2011).

At the national level, apart from these micro studies, there are still some questionings around the actual contribution of Bt cotton to the observed increase in average yields, compared to other factors. At a seminar in November 2009 on the use of socioeconomic assessment in biosafety decision making in New Delhi, an Indian cotton specialist emphasized that besides Bt cotton, increased hybrid adoption had played a significant role in the observed yield increase. Almost all Bt cotton sold in India is hybrid, but hybrid cotton adoption started long before its introduction (for example, Basu and Paroda 1995), so Bt cotton adoption could have accelerated the use of high- yielding varieties. Others have noted that the observed yield growth might also have been at least partially due to favorable weather conditions, increased extension activities in cotton, or the recent increase in cotton prices leading to more investment in input. Some studies also contend that the role of Bt cotton was limited in scope, because of the use of low-quality varieties, bad expression of the protein, growing insect resistance, or other nonrelated issues (for example, Blaise and Kranthi 2011; Glover 2010).

So if there is an apparent relative agreement about the role of Bt cotton in increasing cotton productivity at the farm level in the peer reviewed economics literature, at the macro level, there has been questioning around the contribution of Bt cotton to the observed average yield increase. An international cotton expert recently noted that it was remarkable to see that the results of farm- level surveys about Bt cotton seemed to coincide with macro- level production increases—a feature that is not generally observed in cases of technology adoption.[1] But is this only a coincidence? This study examines this question, which to our knowledge, has not been specifically addressed in the literature, by using a state-level panel data analysis of average cotton yields in India during the past three decades. The focus is on the nine cotton- producing states[2] and the best available national and state data from 1975 to 2009 are used.

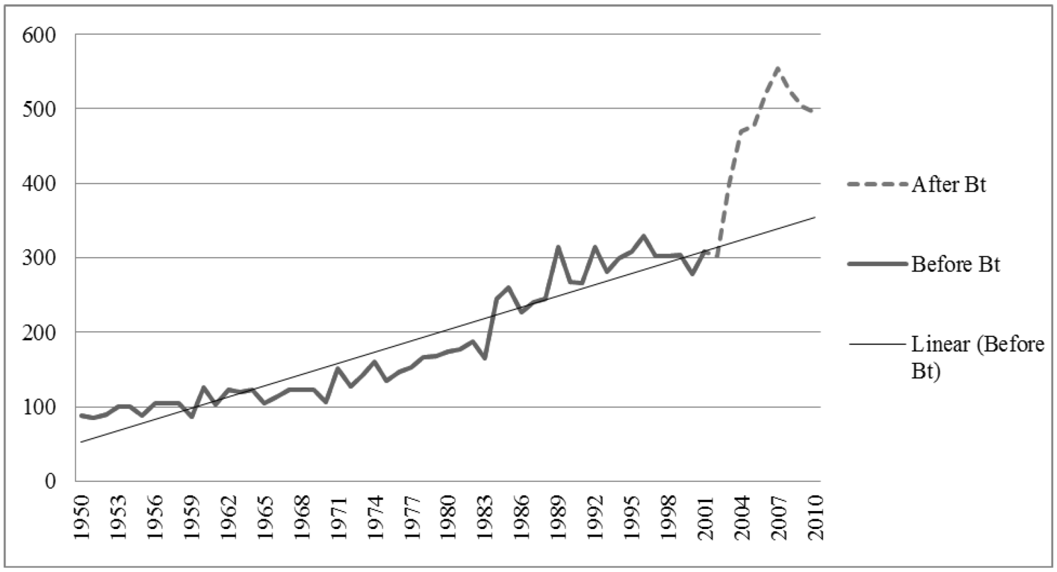
The following section examines the evidence around Bt cotton and average yields, testing for a structural break in the national figures. A panel data analysis is then conducted using an incomplete data set (unbalanced panel). A multiple imputation process is then used to provide a complementary set of estimations with a balanced panel data set. The last section provides some general conclusions.

[1] Meeting with John Baffes, Development Prospect Group, World Bank, March 17, 2011.
[2] Andhra Pradesh, Gujarat, Haryana, Karnataka, Madhya Pradesh, Maharashtra, Punjab, Rajasthan, and Tamil Nadu.

2. BT COTTON AND NATIONAL AVERAGE YIELDS

As shown in Figure 2.1, average cotton yields increased steadily over time in India, in an almost linear fashion until 2002, reaching a level close to 300 kilograms per hectare. Starting in 2003, average yields increased dramatically, jumping to more than 500 kilograms per hectare in four years, and then remaining close to that figure in more recent years. The role of Bt cotton appears significant because its adoption started in 2002, but given that the adoption was not fast the first couple of years, perhaps other factors contributed to the average yield growth.

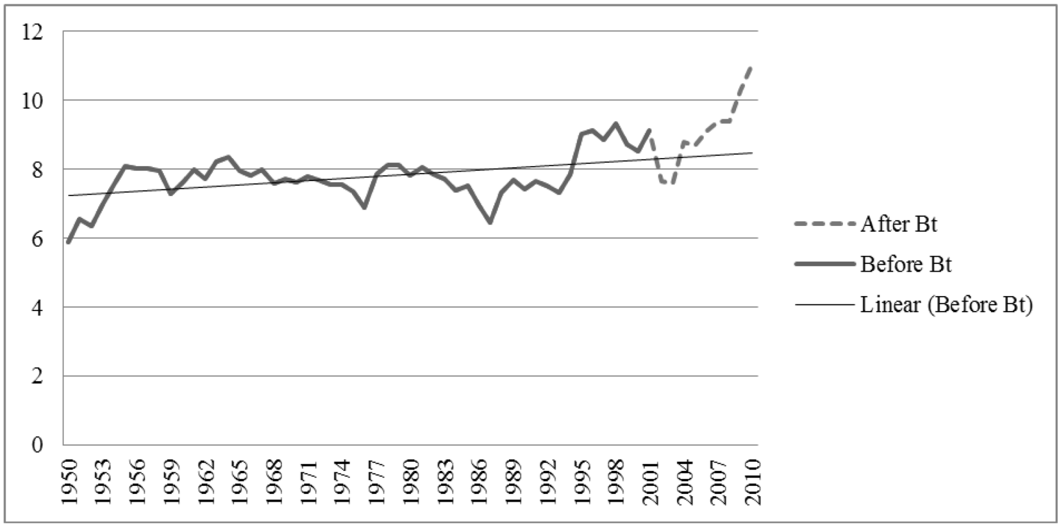
Figure 2.1—Average cotton yields in India, 1950–2010



Source: Indian Cotton Advisory Board, obtained from International Cotton Advisory Committee.

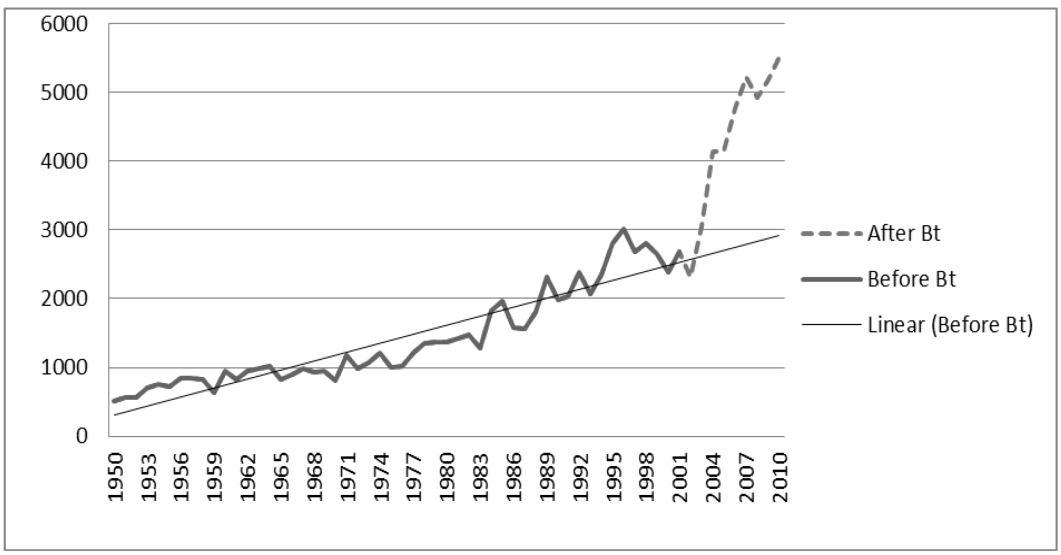
Because average yields are defined as the ratio of production to area, this figure could be due to changes in both factors or primarily one of them. Figures 2.2 and 2.3 show the total area and total production of cotton during the same period. While the production of cotton increased in a very similar fashion average yields, with an acceleration during the post- 2002 era, cotton production area has remained relatively constant over time, increasing only marginally after 2002, and increasing again around 2007. This last phase of acceleration, associated with expansion in less productive land, is reported to have contributed to the stabilization or slight decrease in average yields observed after 2007 (Subramani 2011). Still, Figure 2.3 shows that, undeniably, cotton production increased rapidly after 2002 in this slow expanding area.

Figure 2.2—Total cotton production area (in millions of hectares) in India, 1950–2010



Source: Indian Cotton Advisory Board, obtained from International Cotton Advisory Committee.

Figure 2.3—Total cotton production (in thousands of metric tons), 1950–2010



Source: Indian Cotton Advisory Board, obtained from International Cotton Advisory Committee.

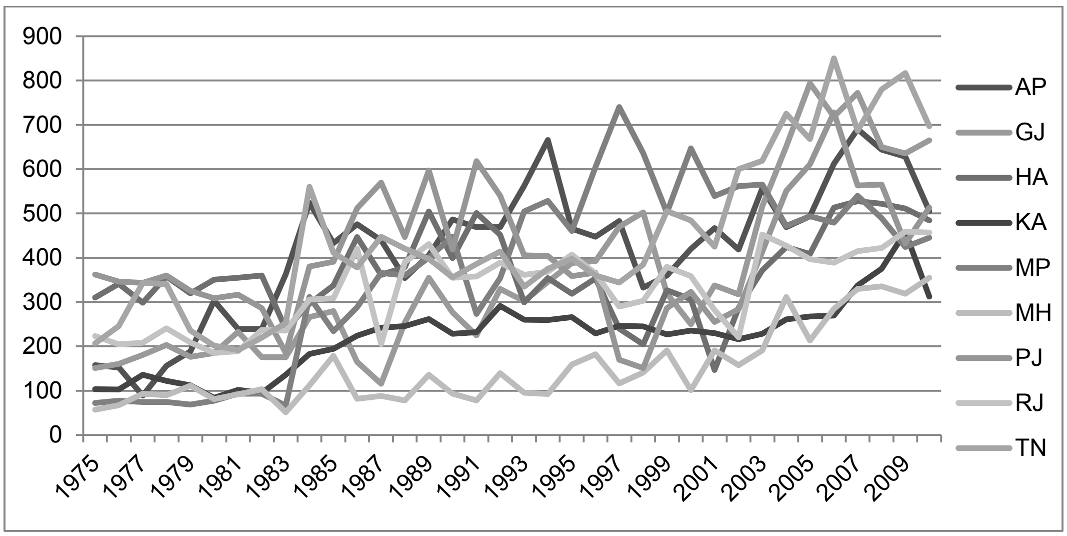
Can the yield growth starting in 2002 define a structural break? We tested this hypothesis with a linear regression, using a simple Chow test (Chow, 1960) on a log-linear model, including only the year of production as the independent variable. We ran three ordinary least squares regressions: (1) for the complete data set, (2) for before 2002, (the year of the official introduction of Bt cotton), and (3) for after 2002 (included); we derived the test statistic for the null hypothesis of equality between before-Bt and after- Bt coefficients as

$$Ctest = \frac{SS_C - (SS_B + SS_A)/k}{((SS_B + SS_A)/(N_B + N_A - 2k))},$$

where SS_C (respectively, SS_A and SS_B) is the sum of square errors from the regression with the complete data set (the before- and after- Bt cotton truncated data sets), k is the number of regressors, and N_A and N_B are the sample sizes of the two subsamples. In our case, we used data from 1950 to 2010, so $N = 61$, $N_B = 52$, $N_A = 7$, and $k = 2$. The Chow test result is 6.36, which is then used for a statistical F test with degrees of freedom = (2, 57). The p -value is found to be equal to .0032; that is, we reject the hypothesis. In other words, the model is consistent with a structural change.[3]Thus, these results indicate that there has been an unambiguous change in average yield trends in the period around 2002–/2003.

Still, the observed national trend masks a significant heterogeneity across states. Figure 2.4 shows average yields in the nine cotton-producing states of India. As seen in this figure, while Gujarat, Maharashtra, and Tamil Nadu have known a remarkable yield increase, Madhya Pradesh actually reduced its average yield after 2002, and other states do not appear to present a significant yield growth in this period. Figure 2.5 shows the evolution of production area. Of the three main cotton-producing states, Maharashtra, Gujarat, and Andhra Pradesh, only Gujarat increased its cotton cultivation area significantly. Last, Figure 2.6 shows production in all the states, and the pattern appears to be much clearer, with Gujarat, Maharashtra, and Andhra Pradesh leading the production growth after 2002 and with other states following on a much smaller scale.

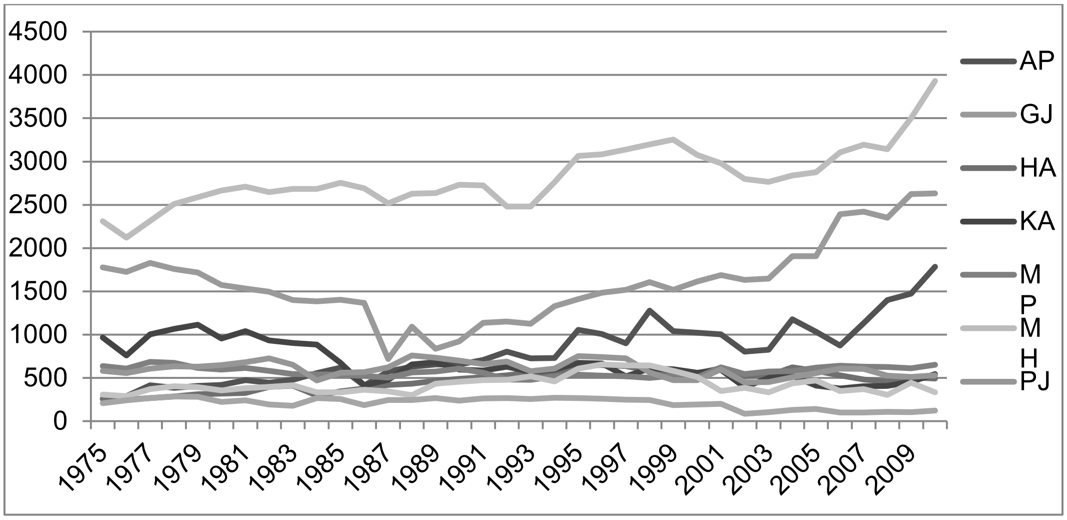
Figure 2.4—Average yields (kg/ha) in the nine cotton-producing states, 1975–2010



Source: Indian Cotton Advisory Board, obtained from International Cotton Advisory Committee.
Note: AP = Andhra Pradesh; GJ = Gujarat; HA = Haryana; KA = Karnataka; MP = Madhya Pradesh; MH = Maharashtra; PJ = Punjab; RJ = Rajasthan; TN = Tamil Nadu.

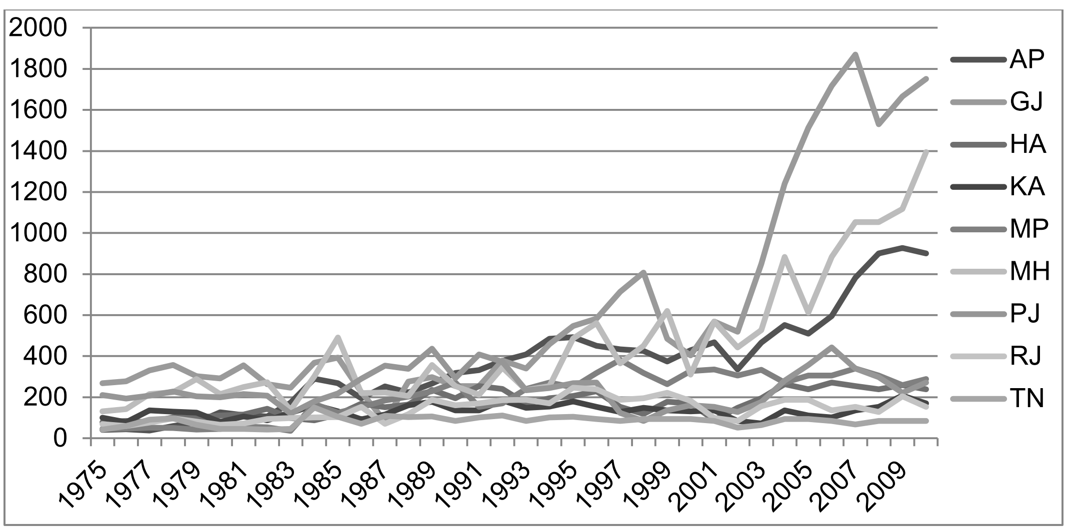
[3] The same test statistics are confirmed when area is added as an explanatory variable (the only variable for which we have 61 value); we find that $Ctest = 4.66$, which leads to an F (3, 55) test p value of .0057, rejecting the null hypothesis of no structural break at the 1 percent level. Adding irrigation rates (proportion of irrigated cotton area to the total cotton area), which drop five recent points, still results in rejecting the null hypothesis at the 5 percent% level, ($Ctest = 3.51$, $F(4, 48)$,) $p = .0137$. Adding annual rainfall or fertilizers does not change this result either while reducing the degrees of freedom of the denominator.

Figure 2.5—Production area (thousand ha) in the nine cotton-producing states, 1975–2010



Source: Indian Cotton Advisory Board, obtained from International Cotton Advisory Committee.
Note: AP = Andhra Pradesh; GJ = Gujarat; HA = Haryana; KA = Karnataka; MP = Madhya Pradesh; MH = Maharashtra; PJ = Punjab; RJ = Rajasthan; TN = Tamil Nadu.

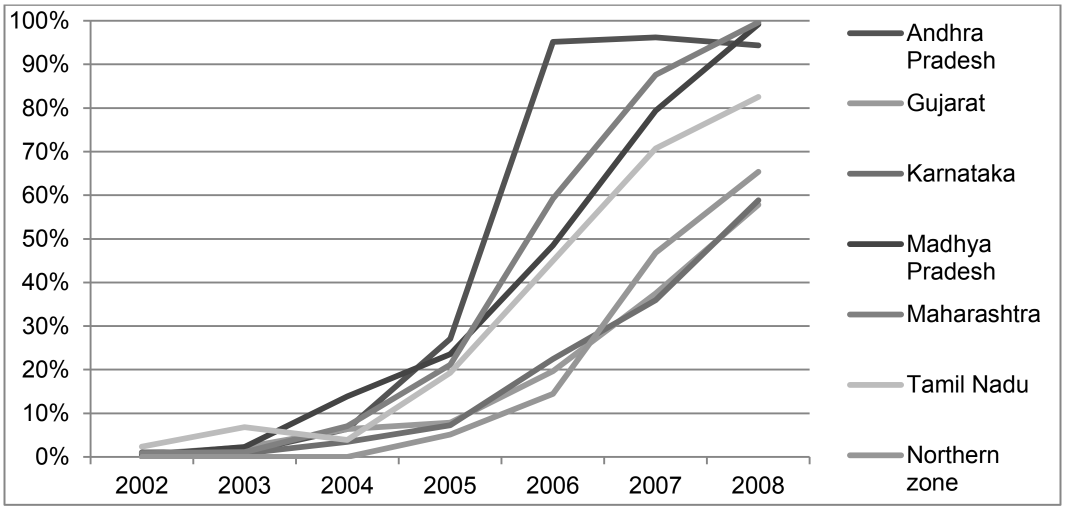
Figure 2.6—Production of cotton lint (thousand metric tons) in the nine cotton- producing states, 1975–2010



Source: Indian Cotton Advisory Board, obtained from International Cotton Advisory Committee.
Note: AP = Andhra Pradesh; GJ = Gujarat; HA = Haryana; KA = Karnataka; MP = Madhya Pradesh; MH = Maharashtra; PJ = Punjab; RJ = Rajasthan; TN = Tamil Nadu.

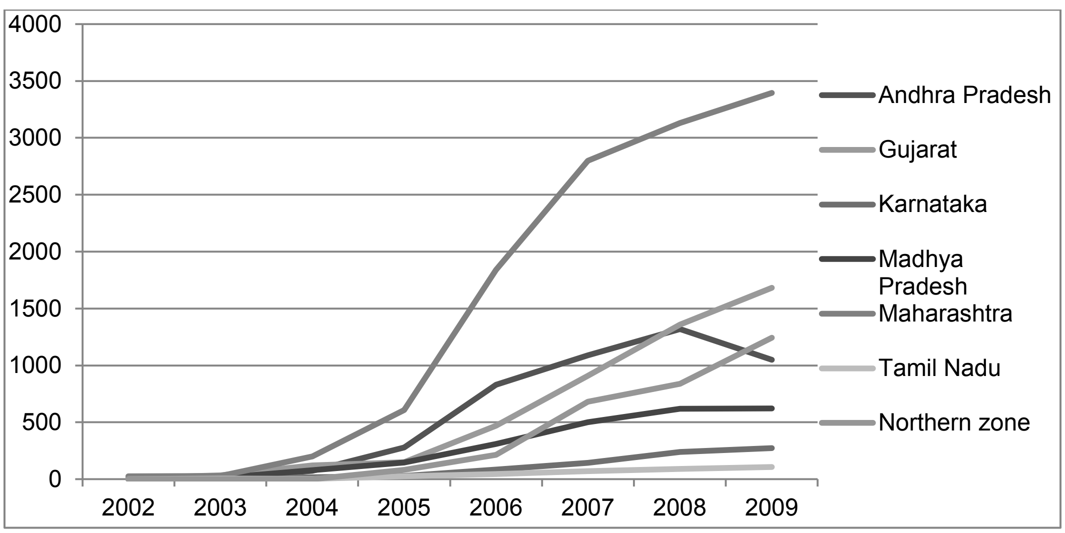
Bt cotton can have had a significant role in the yield growth of these three major states only if its adoption rate was sufficiently large. Figures 2.7 and 2.8 show the official adoption rates and adoption area per state. It is well known that illegal cotton adoption spread out in the west, especially in Gujarat (Pray, Bengali, and Ramaswami 2005; Lalitha, Pray, and Ramaswami 2008), so these figures underestimate adoption there, but they do provide an indication of the dynamics of adoption. Certain observers argued that the jump of yield occurred during 2002–2004, when adoption was low, and it appears that they may be correct at least for some states. Maharashtra's average yields increased mostly between 2002 and 2004, a period in which it had less than 10 percent adoption, and then between 2005 and 2007, when its adoption exploded from 20 percent to 90 percent. Andhra Pradesh's largest increase in yields occurred in 2002–2003, with very low adoption, and then again in 2005–2007, with adoption jumping from 20 percent to 95 percent. Gujarat's main increase in average yields and production occurred in 2003–2005, when official adoption was less than 10 percent but actual adoption was probably much more. Lalitha, Pray, and Ramaswami (2008) report that the area under illegal Bt cotton seeds in Gujarat exceeded the one with legal seeds from 2002–2003 until 2005–2006.[4]

Figure 2.7—Bt cotton official adoption (in percentages) in each state, 2002–2008



Source: Derived from Choudhary and Gaur (2010).
Note: Northern zone is the average adoption for Haryana, Punjab, and Rajasthan.

Figure 2.8—Area of official adoption (in thousands of hectares) of Bt cotton per state, 2002–2009



Source: Choudhary and Gaur (2010).
Note: Northern zone is the average adoption for Haryana, Punjab, and Rajasthan.

[4] Furthermore, the fact that adoption in 2008 was only 60 percent in Gujarat compared to 90 percent elsewhere suggests a potential bias of around 30 percent.

These figures therefore raise the question of an alternative reason for the early jump in yields in 2002–2005 (observed in Figure 2.1). If major cotton states such as Andhra Pradesh and Maharashtra increased yields with very small adoption rates, what contributed to this early jump in yields? Does illegal adoption of Bt especially in Gujarat (Lalitha, Pray, and Ramaswami 2008) explain this discrepancy? Moreover, was the observed structural break in 2002 really associated with Bt cotton? Further analysis shows that the Chow test does not reject the presence of a break in 2005,[5] when Bt cotton started to be largely adopted (according to official figures), but the role of other factors remains to be determined in a more quantitative setting, as addressed in the next section.

[5] With year only, *CTest*(2, 57) = 7.34, $p = .0015$; adding area, *Ctest*(3, 55) = 5.02, $p = .0038$. We can't add irrigation or rain (variables available only before 2005), but using fertilizers (available only until 2007) with areas does result in rejecting the null hypothesis.

3. PANEL DATA ANALYSIS USING INCOMPLETE DATA SET

To assess the role of Bt and other factors, we use a conventional Cobb-Douglas functional form,

$$lnY_{it} = \alpha_i + \sum \beta_k X_{ik} + \gamma BT_{it} + \varepsilon_{it} \quad (1)$$

where Y_{it} is the yield in each state i during each year t, X_{it} is the explanatory factors, BT_{it} is the indicator variable for the use of Bt cotton, ε_{it} is the error terms, and α_i is the state- specific effects (Cameron and Trivedi 2005). We include key productivity- influencing variables: inputs such as synthetic fertilizers, manure, hybrids, pesticides, labor, irrigation facilities, and weather- related variables (rainfall) as well as output-related factors, that is, lagged cotton prices and area of cultivation.[6] For each of the input variables, we assume that an increased use would increase productivity, except in the case of pesticide, where long- run pesticide use could also decrease with Bt cotton use. We further expect to have a potentially positive effect of lagged price on cotton yields and a negative effect of area on yields.

In the case of Bt cotton, we use two alternative regressors, a dummy variable equal to one each year after the introduction of Bt cotton (initial year included), and the official adoption rate in percentage (as shown in Figure 2.7.[7] By comparing results of regressions with each of these two variables, we aim to separate the two hypotheses formulated above, that is, (hypothesis A) that the early jump in yields (2002–2005) was not due to Bt cotton (given its low adoption) and (hypothesis (B) that it was in fact in part due to Bt cotton's illegal or unofficial varieties spread, and therefore not reported in the official adoption rates. Furthermore, this comparison can also help explain whether the productivity growth may have been due to both types of adoption (hypothesis C), to one type but not the other (hypothesis D), or to neither (hypothesis E). Table 3.1 synthesizes possible interpretations to different regression results.

Table 3.1—Determining the role of unofficial versus official variety adoption based on analysis

Coefficient on Bt Dummy Variable	Coefficient on Adoption Rate	Possible Interpretation
Significant	Significant	Two successive increases explained by unofficial and then official Bt cotton adoption (hypotheses B and C)
Significant	Nonsignificant	Unofficial Bt leading the jump in productivity, official adoption marginal effects (hypotheses B and D)
Nonsignificant	Significant	Other factors played a role early, official Bt adoption contributed later (hypotheses A and D)
Nonsignificant	Nonsignificant	Bt was not a major contributor to the jump in productivity (hypotheses A and E)

Source: Authors.
Note: Significant coefficients are expected to be positive.

Descriptive statistics for the variables used in the regression are presented in Table 3.2. As shown in this table, none of the variables, except the Bt dummy are continuously available, that is, representing nine states for the 35-year span from 1975 to 2009. Despite all our efforts, consulting the major Indian databases (including Official Statistics of India on IndiaStat.com) on the Internet and in New Delhi, and

[6] We left out extension expenditures for lack of sufficiently comprehensive data. Tests with incomplete series (multiple imputation) of data based on Rani (2007)—number of cotton frontline demonstrations from 1996–1997 to 2006–2007 by cotton zone rather than state—still found that extension expenditures did not have a significant effect on yields, and that they did not add any explanatory power to the selected models.

[7] This second option is similar to Ramaswami, Pray, and Kelley (2002), where the authors use the proportion of a district's crop area seeded with high-yielding varieties.

asking economists and agriculture specialists in India, we came to the conclusion that no such data set exists. Instead, a significant amount of data is available for many variables from official sources in segments (covering only a few years or a few states).

Table 3.2—Descriptive statistics

Variable Name	Description	Observations	Minimum	Maximum	Mean
Dependent variable					
Log of cotton yields[a]	Log of cotton yield (100 kilograms per hectare)	187	0.93	3.19	2.23
Independent variables					
Bt cotton adoption					
Dummy of adoption[b]	Dummy: 1 = Bt cotton was introduced in the state	315	0	1	0.20
Adoption rate[b]	Proportion of Bt—cotton area cultivated to the total cotton area	306	0.00	0.99	0.06
Rainfall deviation	Absolute value of rainfall deviation in the growing season	297	0.03	388.27	117.54
Fertilizer use[a]	Fertilizer input (kilograms per hectare)	187	16.11	308.46	97.62
Manure use[a]	Manure input (Quintals per hectare)	174	0.00	96.10	17.29
Seed cost[a]	Cost of seed (Rupees. per hectare)	191	28.14	3,779.76	797.22
Human labor[a]	Quantity of farm human labor input (man-hours/hectare)	186	264.06	1,686.94	810.87
Pesticides[c]	Pesticide consumption for the state (metric tons)	201	645.00	13,650.00	3,965.2 6
Cotton price index, lagged[c]	Cotlook price index	261	41.80	94.30	67.98
Cotton area[c]	Cotton area (1,000 hectares)	306	85.00	3,254.00	888.66
Irrigation rate[c]	Proportion of irrigated cotton area to the total cotton area	108	0.03	1.43	0.50

Sources: [a] Directorate of Statistics (1991, 1996, 2011). [b]. Choudhary and Gaur (2010). [c]. IndiaStat.com.

The most complete data set available was the Cost of Cultivation data set, published by the Department of Statistics of the Ministry of Agriculture (Directorate of Statistics 2011), which is an annual report at the state level of the key input quantities and costs for major commodities, based on representative production surveys conducted by extension officers in each state. This data set, which includes categories of input quantities and cost, is available quasi-completely (and was recently added to the Internet) for the nine states between 1996 and 2009, but only in segments for earlier years (for example, limited states and/or limited to 1975–1984). We used the yield level from this data set for consistency purposes, since input factors are taken from the same data set.

In the case of prices, area,[8] rainfall, irrigation, and adoption, we obtained estimates from other sources. In the case of hybrid cotton, we asked several cotton experts both inside and outside India were not able to find a variable that would continuously cover the last decade (IndiaStat.com provides only partial data before 1996). Since this variable was thought to be important, we used the seed cost variable of the Cost of Cultivation data as a proxy for hybrid adoption. Given the large price difference with open pollinated varieties, and the fact that until 2009 virtually all Bt cotton was hybrid, we believe that an

[8] The use of an alternative variable—area relative to long- term average for 1975–2009—for cotton area to represent area expansion over time derived similar results and is therefore not presented.

increased average seed cost would inevitably characterize an increased adoption in cotton hybrids. In the case of rainfall, rather than annual absolute precipitation, we used absolute deviation from the mean during the critical months of the crop season for each state, following a procedure described in the appendix, to emphasize seasons of excessive drought or excessive precipitation.

Given that the gaps in each variable did not generally coincide, especially for those not from the Cost of Cultivation data set, the resulting panel is very limited in size if we include all variables. In particular, the variable on irrigation constrains the whole data set significantly, reducing it to a very small subset of the maximum possible of 187 (dependent variable). To preserve degrees of freedom, while avoiding compromising with the key productivity factors, and noting that rainfall deviation would provide an imperfect indicator of water availability, we kept all variables except irrigation, which led to a balanced panel data set of 110 observations.

Two alternative models were estimated, with the Bt dummy and the adoption rate, as proposed above. The results are shown in Table 3.3. We tested for heteroscedasticity and did not reject the null hypothesis of homoscedasticity; therefore, we did not correct standard errors.[9] We used the Hausman test (Hausman 1978) to determine the best model to use in each case (for example, see Cameron and Trivedi 2005). In the first specification (Bt dummy), the test rejected the null hypothesis that the selected regressors are strictly exogenous with respect to the state- specific unobserved effects, and a fixed effect estimator is used to obtain consistent estimates. Conversely, the null hypothesis is not rejected in the second model, and we use a random effect estimator.[10]

Table 3.3—Summary of panel regressions of complete cases

Variable	Dummy of Adoption			Adoption Rate		
	Coefficient	*t* Statistics	*p* Value	Coefficient	*t* Statistics	*p* Value
Bt cotton adoption	0.1899	2.94***	.004	0.3942	2.02**	.043
Rainfall deviation	–0.0004	–1.52	.133	–0.0007	–1.63	.103
Fertilizer	0.0035	3.89***	.000	0.0030	2.60***	.009
Manure	0.0011	0.41	.685	0.0003	0.11	.911
Seed cost	0.0001	3.19***	.002	0.0002	2.93***	.003
Human labor	0.0008	4.56***	.000	0.0003	1.54	.122
Pesticide	0.0002	–1.34	.182	0.0001	3.69***	.000
Cotton price index, lagged	0.0032	1.57	.121	–0.0001	–0.03	.973
Cotton area	–0.0001	–0.71	.483	–0.0001	–1.44	.151
Constant	1.0692	4.73***	.000	1.4742	6.21***	.000
Number of observations	110			110		
Number of groups	9			9		
F statistics	23.5			3.34		
Probability > *F*	.000			.000		
R-squared (within/overall)	.6969			.4849		
Hausman test: χ^2	36.92			3.34		
Panel data model	Fixed effect			Random effect		

Source: Results from the analysis.
Note: **Significant at 5 percent. ***Significant at 1 percent.

[9] Specific results are available upon request.
[10] It is important to note these differences which somewhat limit the value of pairwise comparison across models, but we still can observe consistency in results.

4. MULTIPLE IMPUTATION ESTIMATIONS

A number of approaches can be used to address the issue of incomplete or missing data, each with its advantages and drawbacks. Here, we use multiple imputation (for example, see Rubin 1987, 1996; Schafer 1997), to represent the uncertainties around the missing data. This method has the advantage of drawing valid statistical inferences from a completed data set. Introducing appropriate random error into the imputation process helps generate unbiased estimates of the parameters, and repeated imputation allows us to get good estimates of the standard errors.

The principle of multiple imputations is to derive estimates from a small set of repeated regressions, each using a data set completed with imputed values (Rubin 1987). More specifically, the process follows three steps (for example, as explained in Marchenko 2009 or Allison 2000): (1) imputation: a small number of data sets (3–10) are completed based on prior distribution of probabilities; (2) completed-data analysis: each of the data sets is analyzed, following the primary specification; and (3) pooling: these results are combined following Rubin's (1987) combination rules (for example, sample averages of estimates).

One of the critical assumptions for obtaining valid inferences from multiple imputation is that missing values should be missing at random (Little and Rubin 1987), that is, that their absence does not convey any information about the data set. In our case, the data set is based on statistical agricultural production, and missing data (for specific years and states) a priori comply with this requirement. The results of multiple imputation are validated if the used imputation method (step 1) follows Rubin's (1987) recommendations, which we do, and if the model used in step 2 that is, the original analysis, without multiple imputation) is statistically valid (Rubin 1987).

In our case, we first define the imputed data and the dependent and independent variables containing missing values that need to be imputed: log of cotton yield, fertilizer, manure, seed cost, human labor, pesticide, and irrigation rate as shown in Table 4.1. We do not impute international price or rainfall deviation, but we include irrigation rate to include in an extended model analysis. The program then imputes 10 versions of the missing values using multivariate normal regressions.[11] Multivariate normal regressions are computed using an iterative Markov chain Monte Carlo method. Last, the new panels are estimated with nonimputed variables and pooled to provide a consistent estimate of each coefficient. We run the same models as with the incomplete panel (with fixed effect and random effect),[12] with correction for possible heteroscedasticity during the data-imputation process.

Table 4.1—Values missing by variable

Variable	Complete	Incomplete	Imputed	Total
Log of cotton yield	187	128	128	315
Fertilizer	187	128	128	315
Manure	174	141	141	315
Seed cost	191	124	124	315
Human labor	186	129	129	315
Pesticide	201	114	114	315
Irrigation rate	108	207	207	315

Source: Authors.

[11] We choose multivariate normal regressions due to the arbitrary missing pattern and continuous feature of our missing variables.

[12] If the imputation process is correctly generated and the missing variables are missed at random, we can assume that the unobserved errors are consistent with the original incomplete data set.

The results of the regression using the variables selected in the incomplete panel (original model) are shown in Table 4.255.[13] Several key differences with the incomplete panel emerge. First, Bt cotton adoption rate has a positive effect on yields, but the dummy variable does not. Second, seed costs—but also fertilizer use, manure, and pesticide use—have a positive significant effect on yields, regardless of the specification. Thus, we also find that hybrids and key agronomic outputs contributed to increase yields over time, as expected, but that Bt cotton introduction itself may not have had a significant effect if considering 27 years (1981–2007) of completed panel data.[14] This second set of estimation is therefore consistent with the third possible interpretation of the yield jump in Table 3.1; Bt cotton adoption contributed significantly to yield increase but that the early productivity jump was not overall due to Bt cotton informal introduction but rather to other factors (hypotheses A and D). Furthermore, the results suggest that Bt cotton contributed an average 0.33 percent of yield increase per percentage adoption per state since its introduction, a figure broadly consistent with the earlier result with the same model.

Table 4.2—Summary of panel regressions of multiple imputations (original model)

Variable	Dummy of Adoption			Adoption Rate		
	Coefficient	*t* Statistics	*p* Value	Coefficient	*t* Statistics	*p* Value
Bt cotton adoption	0.1419	1.41	.166	0.3358	2.04**	.046
Rainfall deviation	–0.0000	–0.18	.857	–0.0000	–0.14	.885
Fertilizer	0.0036	2.88**	.011	0.0036	3.29***	.002
Manure	–0.0010	–0.33	.747	–0.0020	–0.67	.504
Seed cost	0.0005	2.95***	.005	0.0004	1.86*	.064
Human labor	0.0001	2.47**	.022	0.0002	2.72**	.009
Pesticide	0.0000	2.38**	.030	0.0001	2.78**	.007
Cotton price index, lagged	0.0025	1.07	.291	0.0017	0.79	.432
Cotton area	0.0001	0.68	.501	–0.0001	–1.44	.161
Constant	0.9238	3.96***	.000	1.1633	3.77***	.000
Number of observations	243			243		
Number of groups	9			9		
F statistics	13.41			12.61		
Probability > *F*	.000			.000		
Panel data model	Fixed effect			Random effect		

Source: Authors' estimations
Note: ***", "**", "*" represents significance at 1%, 5%, and 10%, respectively.

Because of its importance, we ran an extended model that also included irrigation rates (as imputed). The results of the regression are shown in Table 4.3. Once again, we find the adoption rate, but not the dummy variable, to be significant and positive, confirming the hypothesis that Bt cotton adoption had a later effect and that the early jump in yields was not driven by unofficial varieties (hypotheses A and D in Table 3.1). Other factors matter consistently across the model, such as fertilizer, seed cost, and human labor. The effects of pesticides are not significant, but irrigation percentage rates do contribute to increases in cotton productivity as expected. Including irrigation reduces slightly the annual contribution of Bt cotton to the overall yield growth, reaching around 0.29 percent per percentage adoption in each

[13] Note that the inclusion of the cotton price index and the rainfall variables, together available only from 1981 to 2007, limits the number of observations to 9 × 27 = 243 instead of a total possible of 315 (Table 1.4).

[14] See footnote 13.

state. In comparison, a 1 percent increase in irrigation does increase yield by 0.53 to 0.60 percent in each state. Naturally these are average annual effects over a 27-year time span, and irrigation rate likely increased significantly over time even before Bt cotton was introduced. But once again it emphasizes the partial even if significant effect of Bt cotton adoption on cotton yield growth.

Table 4.3—Summary of panel regressions of multiple imputations (extended model with irrigation rate)

Variable	Dummy of Adoption			Adoption Rate		
	Coefficient	*t* Statistics	*p* Value	Coefficient	*t* Statistics	*p* Value
Bt cotton adoption	0.1243	1.33	.257	0.2866	2.31**	.029
Rainfall deviation	–0.0000	–0.12	.914	0.0000	0.00	.996
Fertilizer	0.0039	3.46**	.069	0.0040	3.80***	.001
Manure	0.0016	0.65	.559	0.0016	0.69	.490
Seed cost	0.0002	3.11**	.050	0.0002	3.65***	.001
Human labor	0.0006	3.50**	.016	0.0005	3.01***	.003
Pesticide	0.0000	0.16	.886	0.0000	0.34	.736
Cotton price index, lagged	0.0024	1.13	.320	0.0016	0.85	.399
Cotton area	0.0001	0.79	.476	0.0000	0.22	.824
Irrigation rate	0.5349	4.93**	.018	0.6075	5.36***	.000
Constant	0.6085	2.72**	.059	0.7439	3.85***	.000
Number of observations	243			243		
Number of groups	9			9		
F statistics	13.64			23.44		
Probability > *F*	.000			.000		
Panel data model	Fixed effect			Random effect		

Source: Authors' estimations.
Note: ***", "**", "*" represents significance at 1%, 5%, and 10%, respectively.

5. CONCLUSIONS

This study examines the contribution of Bt cotton to the observed increase in average cotton yields in India starting in 2002. A number of farm-level studies have shown that, on average, Bt cotton has been a yield-increasing technology, but their results vary widely by location and over time and are based on limited, albeit statistically representative, samples. In contrast, this study looked at state-level adoption of Bt cotton and its effects on average yields since its introduction using a longitudinal data set from the Indian Ministry of Agriculture.

We first looked at yields, production, and area trends over time. At the national level, we confirmed the presence of a structural break in yields in 2002. We then found that in major cotton states, the first large observed increase in average yields occurred between 2002 and 2004, at a time when the official rates of adoption of Bt cotton were still low, and a second increase occurred when adoption started to increase in 2005. This prompted us to formulate the hypothesis that Bt cotton might have had no role years during 2003-05 yield leap, as compared to other factors, or that if it had, it was because of unofficial adoption of Bt varieties, as observed in Gujarat. We further hypothesized that Bt cotton contributed to a second yield leap later on with increasing adoption rates.

To test these hypotheses and assess the potential contribution of Bt cotton, we then used a set of linear panel estimations of average cotton yields over time in the nine main cotton- producing states of India, including key input and output variables. We used two models to address the hypotheses, one with a dummy variable for Bt and the other with the adoption rate, and ran regressions on a limited balanced panel and a complete data set using multiple imputations.

Our results show that Bt cotton contributed significantly to cotton yield growth, ranging from a 0.29 percent to 0.39 percent annual increase in yield for each percentage adoption in each state, or a total increase contribution of 19 percent over time between 1975 and 2010. But the results show that other key factors were consistently significant, especially the use of fertilizers and of hybrid seeds. Human labor, pesticides, and especially the use of irrigation are also found to have had significant effects in several of the regressions. Second, our findings suggest that Bt did contribute to the second increase in cotton productivity (after 2005) but remain inconsistent regarding the possible impact of unofficial Bt cotton adoption in the early years.

Several studies have reported the prevalent use of unofficial Bt cotton long before its official approval in 2002, especially in Gujarat, a state that has led the whole country in cotton production increase during the past decade. But due to incomplete data and lack of information on adoption rates of these unofficial varieties, we still cannot be sure of the actual contribution during this particular period.

While it is clear that Bt cotton was an engine of productivity growth, more research is needed to further explore the apparent contradiction between the early jump in yields (2002-2005) that cannot be explained by other factors and the reported low adoption of Bt cotton in many states at that time.

APPENDIX

Procedure to Derive the Absolute Rainfall Deviation

- Step 1: We computed the total rainfall amount during the key months of the growing season (a different months for each region):
  - Northern zone (Punjab, Rajasthan, Haryana)—Kharif: June, July, August, and September;
  - Central zone (Gujarat, Maharashtra, Madhya Pradesh)—Kharif: July, August, September, and October; and
  - Southern zone (Andhra Pradesh, Karnataka, Tamil Nadu)—Kharif and Rabbi: March, April, May, July, August, September, and October.
- Step 2: We calculated the average rainfall during the growing season for each state, using data from 1950 to 2008.
- Step 3: We generated the rainfall deviation for each state, defined as the rainfall amount in the growing season (step 1) minus the long-term average (step 2).
- Step 4: We took the absolute value of the deviation derived in step 3.

REFERENCES

Allison, P. 2000. “Multiple Imputation for Missing Data: A Cautionary Tale.” *Sociological Methods & Research* 28: 301–309.

Basu, A. K.,. and R. S. Paroda. 1995. *Hybrid Cotton in India. A Success Story*. Bangkok, Thailand: Asia Pacific Association of Agricultural Research Institutions.

Blaise, D.,. and K. R. Kranthi. 2011. “Cry1Ac Expression in Transgenic Bt Cotton Hybrids Is Influenced by Soil Moisture and Depth.” *Current Science* 101 (6): 783–786.

Cameron, A. C., and T. K. Trivedi. 2005. *Microeconometrics: Methods and Applications*. New York: Cambridge University Press.

Choudhary, B., and K. Gaur. 2010. *Bt Cotton in India: A Country Profile*. ISAAA Series of Biotech Crop Profiles. Ithaca, NY, US: International Service for the Acquisition of Agri-biotech Applications.

Chow, G. C. 1960. “Tests of Equality between Sets of Coefficients in Two Linear Regressions.” *Econometrica* 28 (3): 591–605.

Directorate of Statistics. 1991. Cost of Cultivation /Production and Related Data. New Delhi, India: Department of Agriculture and Cooperation, Ministry of Agriculture, Government of India.

Directorate of Statistics 1996. Cost of Cultivation /Production and Related Data. New Delhi, India: Department of Agriculture and Cooperation, Ministry of Agriculture, Government of India.

Directorate of Statistics 2011. Cost of Cultivation /Production and Related Data. New Delhi, India: Department of Agriculture and Cooperation, Ministry of Agriculture, Government of India. http://eands.dacnet.nic.in/Cost_of_Cultivation.htm.

Finger, R., N. El Benni, T. Kaphengst, C. Evans, S. Herbert, B. Lehmann, S. Morse, S. and N. Stupak. 2011. “A Meta-analysis on Farm-level Costs and Benefits of GM Crops.” *Sustainability* 3:743–762.

Glover, D. 2010. Undying Promise: Agricultural Biotechnology’s Pro-poor Narrative, Ten Years On. STEPS Working Paper No. 15. Brighton, UK: STEPS Centre.

Gruere, G.P., P. Mehta-Bhatt, and D. Sengupta. 2008. “Bt Cotton and Farmer Suicides: Reviewing the Evidence”. IFPRI Discussion Paper 00808. Washington, DC: International Food Policy Research Institute.

Gruere, G. P., and D. Sengupta. 2011. “Bt Cotton and Farmer Suicides: An Evidence-based Assessment.” *Journal of Development Studies*. 47 (2): 316–-337.

Hausman, J. A. 1978. “Specification Tests in Econometrics.”*Econometrica* 46 (6): 1251–1271.

Pray, C.E., L. Nagarajan, J. Huang, R. Hu, and B. Ramaswami. 2011. “The Impact of Bt Cotton and the Potential Impacts of Biotechnology on Other Crops in China and India.” In *Global Welfare of Genetically Modified Crops*, vol. 10, Frontiers of Economics and Globalization, edited by C. A. Carter, G. Moschini, and I. Sheldon, 83-114. vol. 10. Bingley, UK: Emerald.

IndiaStat.com.2010. Official Statistics of India. Consulted in Fall 2010.

International Cotton Advisory Committee. (ICAC). 2010. Factors Influencing the Use of Pesticides in Cotton in India. Report from the Expert Panel on Social, Environmental and Economic Performance of Cotton Production. (SEEP). Washington, DC: ICAC.

Karihaloo, J. L., and P. A. Kumar. 2009. *Bt Cotton in India—A Status Report*. 2d ed. New Delhi, India: Asia-Pacific Consortium on Agricultural Biotechnology and Asia Pacific Association of Agricultural Research Institutions

Kouser, S., and M. Qaim. 2011. “Impact of Bt Cotton on Pesticide Poisoning in Smallholder Agriculture: A Panel Data Analysis.” *Ecological Economics* 70:2105–2113.

Lalitha, N., C. E. Pray, and B. Ramaswami. 2008. *The Limits of Intellectual Property Rights: Lessons from the Spread of Illegal Transgenic Seeds in India*. Discussion Paper 08-06. New Delhi, India: Indian Statistical Institute. www.isid.ac.in/~pu/dispapers/dp08-06.pdf.

Little, R. J. A.,. and D. B. Rubin. 1987. *Statistical Analysis with Missing Data*. J. Wiley & Sons, New York.

Marchenko, Y. 2009. “Multiple-imputation Analysis Using Stata’s mi Command.” Presentation given to the 2009 UK Stata Users Group Meeting, London, UK, on September 10., 2009. StataCorp.

Pray, C. E., P. Bengali, and B. Ramaswami. 2005. “The Cost of Biosafety Regulations: The Indian Experience.” *Quarterly Journal of International Agriculture*, 44:267–289.

Press Trust of India. 2011. *Government Moving Cautiously on GM Crops Commercialisation*: Pawar. October 19, 2011. www.moneycontrol.com/news/wire-news/govt-moving-cautiouslygm-crops-commercialisation-pawar_602388.html.

Ramaswami, B., C. Pray, and T. Kelley. 2002. “Dissemination of Private Hybrids and Crop Yields in the Semi-arid Tropics of India.” *Indian Journal of Agricultural Economics* 57 (1): 39–51.

Raney, T., and I. Matuschke. 2011. “Current and Potential Farm-level Impact of Genetically Modified Crops.” In *Global Welfare of Genetically Modified Crops*, vol. 10, Frontiers of Economics and Globalization, edited by C. A. Carter, G. Moschini, and I. Sheldon, 55-82. Bingley, UK: Emerald.

Rani, S. U. 2007. “Extension Strategies for Cotton Outreach Programme.” In *Training Manual on “Cultivation of Long Staple Cotton (ELS)*,” edited by N. Gopalakrishnan, S. Manickam, and S. M. Sabesh. http://cicr.org.in/pdf/ELS/so1.pdf.

Rao, N. C., and S. M. Dev. 2010. *Biotechnology in Indian Agriculture: Potential, Performance and Concerns*. New Delhi, India: Academic Foundation.

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

Rubin, D.B. 1996. “Multiple Imputation after 18+ Years (with Discussion).” *Journal of the American Statistical Association* 91: 473–489.

Schafer, J. L. 1997. *Analysis of Incomplete Multivariate Data*. London: Chapman & Hall.

Subramani, M.R. 2011. “Growing Cotton in New Areas Dragging Yield.” Hindu Business Line,. December 7. www.thehindubusinessline.com/industry-and-economy/agri-biz/article2695314.ece#.TvnmuBr6s7M.email.

Subramanian, A., and M. Qaim. 2009. “Village-wide Effect of Agricultural Biotechnology: The Case of Bt Cotton in India.” *World Development* 37:256–267.

________. 2010. “The Impact of Bt Cotton on Poor Households in Rural India.” *Journal of Development Studies* 46:295–311.

RECENT IFPRI DISCUSSION PAPERS

For earlier discussion papers, please go towww.ifpri.org/pubs/pubs.htm#dp.
All discussion papers can be downloaded free of charge.

1169. *Including women in the policy responses to high oil prices: A case study of South Africa*. Ismael Fofana, 2012.

1168. *Economic statecraft in China's new overseas special economic zones: Soft power, business, or resource security?* Deborah Bräutigam and Tang Xiaoyang, 2012.

1167. *Revisiting the palm oil boom in Southeast Asia: The role of fuel versus food demand drivers*. Daniel J. Sanders, Joseph V. Balagtas, and Guillaume Gruere, 2012.

1166. *The food security system: A new conceptual framework*. Olivier Ecker and Clemens Breisinger, 2012.

1165. *Farmers' information needs and search behaviors: Case study in Tamil Nadu, India*. Suresh Chandra Babu, Claire J. Glendenning, Kwadwo Asenso-Okyere, and Senthil Kumar Govindarajan, 2012.

1164. *Rural demography, public services, and land rights in Africa: A village-level analysis in Burkina Faso.* Margaret McMillan, William A. Masters, and Harounan Kazianga, 2012.

1164. *Reforming the public administration for food security and agricultural development: Insights from an empirical study in Karnataka.* Regina Birner, Madhushree Sekher, and Katharina Raabe, 2012.

1163. *Economic development, external shocks, and food security in Tajikistan*. Kamiljon T. Akramov and Ganga Shreedhar, 2012.

1162. *Infectious disease detection with private information*. Alexander E. Saak, 2012.

1161. *Economic transformation in Ghana: Where will the path lead?* Shashi Kolavalli, Elizabeth Robinson, Xinshen Diao, Vida Alpuerto, Renato Folledo, Mira Slavova, Guyslain Ngeleza, and Felix Asante, 2012.

1160. *Globalization, structural change, and productivity growth*. Margaret McMillan and Dani Rodrik, 2012.

1159. *A review of input and output policies for cereals production in India*. Ganga Shreedhar, Neelmani Gupta, Hemant Pullabhotla, A. Ganesh-Kumar, and Ashok Gulati, 2012.

1158. *Demand and supply of cereals in India: 2010-2025.* A. Ganesh-Kumar, Rajesh Mehta, Hemant Pullabhotla, Sanjay K. Prasad, Kavery Ganguly, and Ashok Gulati, 2012.

1157. *Close eye or closed eye: The Case of export misinvoicing in Bangladesh*. Pranav Kumar Gupta, Devesh Roy, and Kaikaus Ahmad, 2012.

1156. *The sophistication and diversification of the African Agricultural sector: A Product Space Approach*. John Ulimwengu and Thaddée Badibanga, 2012.

1155. *Why women are progressive in education?: Gender disparities in human capital, labor markets, and family arrangement in the Philippines*. Futoshi Yamauchi and Marites Tiongco, 2012.

1154. *Resource-rich yet malnourished: Analysis of the demand for food nutrients in the Democratic Republic of Congo*. John Ulimwengu, Cleo Roberts, and Josee Randriamamonjy, 2012.

1153. *Putting gender on the map: Methods for mapping gendered farm management systems in Sub-Saharan* Africa. Ruth Meinzen-Dick, Barbara van Koppen, Julia Behrman, Zhenya Karelina, Vincent Akamandisa, Lesley Hope, and Ben Wielgosz, 2012.

1152. *Household preferences and governance of water services: A Hedonic analysis from rural Guatemala*. William F. Vásquez, 2011.

1151. *Peer effects, risk pooling, and status seeking: What explains gift spending escalation in rural China*? Xi Chen, Ravi Kanbur, and Xiaobo Zhang, 2011.

1150. *Agricultural productivity and policies in Sub-Saharan Africa*. Bingxin Yu and Alejandro Nin-Pratt, 2011.

1149. *Common-pool resources, livelihoods, and resilience: Critical challenges for governance in Cambodia*. Blake D. Ratner, 2011.

1148. *The impact of global climate change on the Indonesian economy*. Rina Oktaviani, Syarifah Amaliah, Claudia Ringler, Mark W. Rosegrant, and Timothy B. Sulser, 2011.

INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE

www.ifpri.org

IFPRI HEADQUARTERS

2033 K Street, NW
Washington, DC 20006-1002 USA
Tel.: +1-202-862-5600
Fax: +1-202-467-4439
Email: ifpri@cgiar.org