FORECASTING U.S. COTTON PRODUCTION

Fred A. Vogel
National Agricultural Statistics Service
Washington, DC

Abstract

One of the primary responsibilities of the National Agricultural Statistics Service (NASS) in the U. S. Department of Agriculture (USDA) is to provide forecasts and estimates of current and future supplies of production from the nation's farms. Cotton is a major commodity in the U.S. and the world, therefore, a considerable effort goes into providing current and timely information.

The schedule of reports about cotton production is designed primarily to provide early season projections of the size of the current year's crop. The first such measure is the March Prospective Plantings report which provides an estimate of how many acres growers expect to plant to cotton and other crops in the upcoming crop season. This is followed in June by estimates of the number of acres actually planted to cotton in the current season. The first survey based forecast of production is made in August. These production forecasts are updated each month through January. Shortly after cotton harvest begins, biweekly reports of bales ginned are issued based on a complete census of all cotton gins. These reports continue until the ginning season is finished and are the basis for the final estimates of cotton production.

The cotton production forecasts are determined by separately estimating the acres to be harvested and then forecasting a yield per acre. State-of-the art sample surveys are conducted to first determine the acres planted to cotton. One survey is based on a sample of growers who are asked to report the acreage of cotton and other crops that they have either already planted or intend to plant. The other survey is based on a personal enumeration of producers who operate land identified in scientifically selected random sections of land.

The yield forecasts are also based on two independent surveys. First, a sample of producers is interviewed each month to obtain the pounds of lint they expect to harvest from their operation. In addition, a random sample of cotton fields is selected to be visited each month during the growing season. Within these sample fields, random plots are located and counts of plants, blooms, and bolls in their various growth stages are obtained. These fruit counts are used in statistical models to forecast the number of bolls and boll weight. The official forecasts of yield are based on a combination of the grower reported expectations and the field measurements. When harvest begins, ginning data are also used to project the final crop production.

The following sections will provide more detail about the methods used to provide forecasts and estimates of the nation's cotton production.

Estimates of Cotton Acres Planted and Harvested

Two different statistical methods have been developed by NASS to estimate crop acreage. Each has strengths and weaknesses, but together provide very accurate estimates of cotton and other crop acreages. All NASS estimates of acreage are based on sample surveys of cotton producers and also surveys of land areas. The following paragraphs describe the sampling plans used.

Area Sampling Frame

The entire land mass of the U.S. has been classified by type of land use using satellite imagery, aerial photography, geological survey maps and computer software packages. These land use classifications range from intensely cultivated areas to marginally cultivated areas, grazing land, waste land, and urban areas. Within each land use type, the land mass is further subdivided so that random samples of segments of land about one mile square in size can be randomly selected. The purpose of the land use classification is to allow cultivated areas to be selected with greater frequency than areas with less cropping or agricultural activity. Table 1 shows the land use classification for Texas and the number of sample segments selected for the June survey. A very powerful feature of the area frame sample design is that every acre of land in the U.S. has a known probability of being selected, which means that every acre of cotton also has a known probability of being selected. These probabilities are used to determine sampling weights which when multiplied by the cotton acres in the sample and summed across all sample segments, will provide an estimate of the State's cotton acres.

Two area frame samples are used in the NASS survey program. One is designed to provide estimates of acreage devoted to spring planted crops. It is surveyed in June with a U. S. sample size of about 10,000 segments. The other sample contains about 6500 segments for the entire U.S. It was designed to provide estimates of winter wheat seedings and cattle inventories and is surveyed in December each year.

The 10,000 sample segments are surveyed each year during the first two weeks in June. Trained enumerators are given aerial photographs and county maps showing the location of each assigned sample segment. They find the segment, locate farm operations that operate the land in the segment, and with the operator's help, draw the boundaries of each field on the aerial photograph. At that time the enumerator also obtains the crop and acres in the field. If the crop is

Reprinted from the *Proceedings of the Beltwide Cotton Conference*
Volume 1:217-219 (1999)
National Cotton Council, Memphis TN

not yet planted, the operator is asked to report what will be planted in each field.

The strengths of the area sampling frame is that the aerial photographs show clear field boundaries, so reporting errors are minimized. One source of error can occur when crops not yet planted do not get planted to the crop the farmer reported. When crop seasons are late, these operators are recontacted to learn whether or not their intentions were realized.

One disadvantage of this survey is that it is expensive compared to other survey methods because a personal interview is required. As a result, sample sizes are generally not sufficient to estimate the number of acres planted to minor crops for the State. The frame development activities are also time consuming and expensive. As a result, the same sample of segments is used for 5 years. However, one-fifth of the sample is rotated each year so that each year 80 percent of the segments are the same as those used the previous year.

List Sampling Frame

Each NASS State office maintains a list of farms for their State. This list contains names, addresses, and relative measures size based on historical survey data or other administrative data. These size measures can be used to classify each name by its crop size. The classification of the Texas list frame for crops/stocks surveys is shown in Table 2.

The sample depicted is surveyed in June for acres planted and to be planted, in September to obtain small grain acres harvested and production, in December for spring planted crops acres harvested and production, and in March to obtain acres farmers expect to plant in the next crop season. The producers in the sample are also asked to report the quantities of grain in storage on their operation.

Note that the sampling rates for the very largest operations are 1.0. Because of their size and effect on the cotton acreage, these farms are selected for all surveys. All sample farms are interviewed during the first two weeks in June and asked to report the number of acres they planted to each crop on their operation. This survey provides an important measure of minor crop acreage and also of cotton acres in States with small acreage because the cotton farms can be targeted. An advantage of this sample is that data collection costs are less because the data can be collected by mail or by telephone. A serious disadvantage is that the list is incomplete. Even with a complete census, operators enter and exit from agriculture faster than the list developers can keep up with the changes. While the list sample provides an efficient way to measure crop acres, its estimates will be incomplete. For that reason, NASS, using statistical theory from Iowa State University, developed a strategy called multiple frame sampling. This is a way to statistically combine the incomplete list survey results with the completeness of the area frame results to obtain a measure of total crop acres planted.

These two sampling frames also provide the basis for the yield forecast surveys that follow starting in August each year. The methodology used to prepare the yield forecasts follows below.

Yield Forecasts

Grower Survey

A random sample of growers selected in the list sample and who reported cotton acres planted during the June Survey is selected for the monthly yield forecasts. Each month starting in August they are interviewed. In August, they are asked whether all acres reported in June were actually planted and how many they expect to harvest. This forms a part of the basis for the acres harvested estimate first issued with the August Crop Production report and planted acreage adjustments. Then, the growers are asked to report the yield they expect from their acreage. The same growers are recontacted each month from September through December to obtain an update on yield prospects.

Objective Yield Survey

A random sample of fields is selected from those identified on aerial photographs during the land use survey conducted during the first two weeks in June. The operators of the sample fields are interviewed in late July or the first few days in August to verify the acres in each field and whether they were actually planted to cotton. Changes in acreage are used with the above sample to update the planted acres estimates. The operator is also asked to report the acres to be harvested. This, along with the information obtained from the above sample of growers, is the basis for the estimates of cotton acres for harvest. Finally, enumerators ask permission to enter the selected cotton fields to make counts and measurements.

Data are obtained from each sample field beginning in August and continues each month through harvest. In August, two random plot locations are derived for each field by assigning a random number of rows that are to be walked along the edge of the field and a random number of paces to walk into the field. These random locations are determined before the field is assigned to a field enumerator. Each sample field contains two sample plots. Each sample plot is two rows wide and 10 feet long. One row of each plot is extended another 3 feet where more detailed counts of blooms, squares, etc., are obtained. Each plot is visited every month until the field is harvested. Each month, open bolls in the sample plots are removed, sent to a lab, dried, and weighed. Boll weights are adjusted to reflect a standard 5 percent moisture. After the sample field has been harvested, the enumerator returns to the field, randomly locates two more plots, and picks up the cotton remaining in the field, which provides a measure of harvest loss.

From previous years, this survey process has provided a file of individual plot level information of early season fruit counts followed by end of season numbers of open bolls, boll weights, and harvest loss. The relationship between the early season counts and the final yield from individual sample plots is used to prepare forecast models used in the current year to project early season counts into yields per harvested acre.

The historical models, based on previous years' data, are used to forecast the number of bolls to be harvested from each sample plot, and their average weight. The plot level yields are averaged to a State level yield and the estimate of harvest loss subtracted. The information used to determine the plot level forecast depends on the unit's maturity. Table 3 shows the variables counted by level of maturity. These are the variables used in the regression models employing similar data from the previous 5 years.

Note that if a plot's maturity is such that no fruit are present, the 5-year average number of bolls per acre is used along with 5-year average fruit weights and harvest loss. As the crop progresses into the fruiting stages, numbers of squares and small and large bolls are used in the regression models to forecast the final number of bolls to be harvested. As large bolls open, they are removed and sent to a lab where they are dried and weighed. When 20-85 percent of the expected large bolls have been removed, current weights are used in a model to project final weights. Current weight and harvest loss data replace the 5-year average and model after 85 percent of the large bolls have been removed.

The relative accuracy of each month's forecast is very dependent on the crop maturity which determines how much reliance is placed on current vs. historical data. The main unknown is future weather and those effects.

The grower and Objective Yield surveys are conducted at the end of each month, with data summarized and analyzed for the forecast which usually is published around the tenth. The yield and production forecasts are based on conditions as of the first of the month along with the assumption that normal conditions will prevail the rest of the crop season. The concept of "normal" assumes that temperatures and precipitation will be at historical averages for the remainder of the season. When forecasting cotton yields, NASS does not attempt to predict future weather conditions. Long range weather forecasts are not used in any forecast models.

Accuracy of Monthly Crop Forecasts

Table 4 shows how much each forecast, on the average differs from the final estimate. The table also shows the number of years, out of the last 10, each month's forecast is above or below the final. The key point to see is the steadily improving accuracy as the season progresses.

There are several things that knowledgeable data users can do to make their own assessment of the accuracy of the NASS acreage and production reports. First, if cotton planting is about complete at the time of the June Survey, the resulting acreage estimates will be very precise. It is only if there is an unusual planting season that the estimate of acres planted, based on the June survey, will be subject to later revision. The next thing to assess is the historical relationship between planted and harvested acres. This will vary considerably depending on weather conditions after planting. In 1998 for example, a severe drought resulted in a considerable amount of cotton acreage being reseeded to sorghum.

The National Agricultural Statistics Service also publishes a weekly report of crop conditions and crop progress during the growing season. This report can be used during the season to assess the probable accuracy of the monthly crop reports. For example, if the cotton crop has progressed so that by August 1 a large proportion of the crop is setting bolls, then the forecast is likely to be more accurate than if no fruit are present. Data users can also make their own assessment of how the current year may be different than the previous 5 years which form the basis for the forecast models when the crop is immature.

Summary

Extraordinary efforts are used to ensure that the survey results and the official estimates are not made available to anyone prior to the release of the report on the scheduled date and time. First, all NASS employees are subject to severe penalties if they allow a report to be released prematurely or make use of the data prior to release. The survey data for States like Texas are so sensitive that they are encrypted when transferred via computer to the Headquarter's office. The Headquarter's office does not make the final data review and prepare the final forecasts or estimates until employees are sequestered in an area where phones have been disconnected, windows sealed, and an armed guard prevents anyone from leaving. People working on the cotton estimates arrive at the workplace around 10:00 the night before the report is to be released, which is at 8:30 a.m. the next morning. Around 8:00 a.m. the Secretary of Agriculture or a representative of that office comes into the "lockup" area to be briefed about the new information to be released to the public. At exactly 8:30 a.m., the report is released to the public by opening phone lines and data bases available to the public. News reporters, who are allowed in a media room in lockup, are allowed to open their phone lines and transmit their stories at 8:30 a.m. as well.

In summary, the forecasts and estimates are only as good as the data reported by cotton producers and cotton ginners. We want to acknowledge and thank the thousands of producers and ginners who voluntarily report for their operations several times a year. Their cooperation and support is essential and greatly appreciated.

Table 1. Description of Texas Land Use Classification and Resulting Sample Sizes

Land Use Classification	Segment Size	Number of Segments in Frame	Sample Size	Sample Weight
	(1) (Sq. Miles)	(2)	(1)÷(2)	Percent Cultivated
>50%	1.00	13,511	90	150
>50% - Cotton	1.00	23,785	243	98
>50% - Peanuts	1.00	1,768	12	147
>50% - Rice-East	1.00	2,838	18	157
15-50% - West	2.00	20,626	132	156
15-50% - Cotton	2.00	9,421	168	56
15-50% - Peanuts-West	2.00	2,682	6	447
15-50% - Rice	2.00	2,420	15	161
Agri-Urban:	0.25	16,300	6	2,717
Commercial:	0.10	7,744	3	2,581
Open Range-East	4.00	6,459	9	718
Open Range-West	8.00	5,278	42	125
Wooded Range-East	4.00	6,835	12	570
Arid Range	Variable	1,349	4	337
			1,165	

Table 2. Classification of Texas List of Farms for Crop Acreage and Grain Stocks Survey

Boundaries	Population	Sample Size	Interval
Cropland 100-899 Acres	28,212	1,050	26.9
Storage Capacity 1-14,999 Bushels	5,890	650	9.1
Cropland 900-14,999 Acres	7,132	1,225	5.8
Peanuts 100+ Acres	680	180	3.8
Capacity 15K-199,999 Bushels	1,181	450	2.6
Capacity 200K-999,999 Bushels	60	14	4.3
Soybeans 250+ Acres	245	49	5.0
Barley 50+ Acres	44	15	2.9
Sunflowers 50+ Acres	154	30	5.1
Rice 100-499 Acres	438	150	2.9
Rice 500+ Acres	152	75	2.0
Pima Cotton 100+ Acres	47	20	2.4
Cropland 15,000+ Acres	14	14	1.0
Capacity 1,000K+ Bushels	5	5	1.0
	44,254	**3,927**	

Table 3. Variables Used in Sample Level Forecasts of Cotton Yields

Maturity Stage	# of Bolls	Weight Per Boll	Harvest Loss
No Fruit Present	5-yr. Avg.	5-yr. Avg.	5-yr. Avg.
Squares Present	Number of Squares	5-yr. Avg.	5-yr. Avg.
Bolls Present	Cumulative # Large bolls & Small bolls	5-yr. Avg.	5-yr. Avg.
20-85% Large bolls open*	Cumulative # Large bolls	Model of Current Weight	5-yr. Avg.
Over 85% Large Bolls Open	Cumulative # Large Bolls	Current Net Weight	5-yr. Avg.
Field Harvested	Total bolls Picked in Unit	Final Net Weight	Current Har. Loss

* When open bolls are present, they are picked, sent to a lab, and weighed.

Table 4. Reliability of NASS Cotton Production Forecasts

Forecast	Average Difference from Final	Number of Years Below Final Percent	Above Final Percent
August	8.6	6	4
September	6.0	6	4
October	4.1	7	3
November	2.6	6	4
December	1.6	5	5