Summary Report on Texas Citrus Research

Research and Development Branch

Research Division

Statistical Reporting Service

United States Department of Agriculture

December 1973

Summary Report on Texas Citrus Research

by

Raymond R. Bosecker

Research and Development Branch

Research Division

Statistical Reporting Service

United States Department of Agriculture

December 1973

Contents

I. Background

Research into the application of objective yield procedures[1/] to estimate the Texas citrus crop was conducted during the 1968, 1969, and 1970 growing seasons in the lower Rio Grande Valley. The Research and Development Branch (R&DB) of the Statistical Reporting Service (SRS) was responsible for conducting these surveys and presenting the results. The project was sponsored and funded jointly by the Texas Citrus Mutual, Texas Dept. of Agriculture, and the Statistical Reporting Service.

The study conducted during the 1968-69 growing season demonstrated the feasibility of using limb counting techniques and ground photography of citrus trees to estimate the fruit per tree.[2/] This initial study was confined to a small number of non-randomly selected blocks of fruit.

Research during the 1969-70 growing season provided estimates of variances and work requirements for citrus estimation.[3/] A sample of 54 blocks with four trees per block provided a good measure of within and between block variance components. The study of ground photography as an auxiliary variable was continued. Forecasting of fruit size and drop was also studied.

1/ Estimation procedures based on actual fruit counts from randomly selected limbs.

2/ Richard D. Allen and Donald H. Von Steen, "Use of Photography and Other Objective Yield Procedures for Citrus Fruit," 1968 Texas Research, USDA, SRS, June 1969.

3/ Richard D. Allen, "Evaluation of Procedures for Estimating Citrus Fruit Yield," USDA, SRS, February 1972.

The 1970-71 research project was designed to test the procedures and techniques developed in the earlier research and provide recommendations for an operational survey. The objectives of this pilot study were to evaluate various estimating models, fruit counts on photographs, size and drop estimators, and the Texas Citrus Mutual's list as a sampling frame.

II. Sample Selection

A listing of trees compiled by Texas Citrus Mutual for the lower Rio Grande Valley was used as the sampling frame. Trees were designated by type of citrus (early oranges, Valencia oranges, or grapefruit) and age (0-3 years, 4-7 years, or 8+ years) at the time of listing (about 1967). A stratified, multi-stage systematic random sample was then drawn for each type of citrus.

The primary unit or first stage in the sample selection was the block or grove of trees within type and strata. The second stage was selection of trees within blocks. The third stage was primary limbs [4/] within trees and the fourth stage was the selection of terminal limbs [5/] on primaries.

III. Sample Size for Operational Survey

The 1969-70 study indicated a first stage sample size of 65 to 85 blocks is needed to provide a 10-percent coefficient of variation (C.V.) in the estimated average fruit per tree for each type of citrus. [6/]

4/ Primary limbs or scaffolds are major limb divisions emerging from the main trunk.

5/ Small limbs emerging from the primary limbs used as sample units for counting fruit (defined as a limb with a cross-sectional area (CSA) between 0.6 and 1.2 square inches).

6/ Allen (February 1972), page 19

This result provides the basis for calculating the expected costs for an operational survey with this degree of precision. The report on the 1969-70 project also showed that selection of two trees per block, two primaries per tree and two terminals per primary is near the optimum.[7/]

IV. Survey Procedures

Blocks were chosen by systematic selection from a random start. This procedure gave each block an equal chance of selection within each type and strata. Trunk measurements of a sample of 30 to 40 trees in each grove selected provided the basis for the systematic selection of two sample trees from a listing arranged by trunk cross-sectional area (CSA).

Primary limbs were identified, CSA measurements were made, and two sample limbs were selected from all the primaries with equal probability. Terminal limbs on the sample primaries were then identified using a limb selection gauge and an equal probability sample of two terminals was selected on each sample primary.

Fruit counts were made on the four sample terminal limbs by a team of two enumerators counting the fruit once together. Size measurements and drop counts were made on the first terminal limb selected on each of the sample primaries. All fruit on the selected limbs were tagged and numbered.

Photography of one side of the sample trees for later photo counting provided the estimates of correlation between fruit estimated from limb counts and fruit visible on color slides.

7/ Ibid., page 17

Analysis of variance from the 1968 study indicated one side of the tree was sufficient for estimation purposes and perhaps even diagonal quarters would be enough.[8/] The average photo counts from two counters in close agreement can be used for the photo count. For about two-thirds of the trees, two photographs per tree were sufficient while one-third of the trees were too large and had to be photographed in quarters.

V. Time Costs

Data available from the research projects on the time required for the survey procedures are presented in Table 1 in minutes per tree for each type of fruit. Times shown are an approximation of the averages across strata and over all three surveys.

Table 1--Minutes per tree for various citrus study field operations by type

Procedure	Early Oranges	Valencias	Grapefruit
First Stage mapping[9/]	13.0	10.0	11.0
Second Stage mapping[10/]	10.0	9.0	10.0
Counting fruit	11.0	8.0	7.5
Tagging fruit	7.5	6.5	7.5
Measuring fruit	5.5	8.0	4.0
Photographing trees	5.0	6.5	5.5
Counting on Photos	10.0	10.0	10.0
Misc. within block	9.5	10.0	9.5

8/ Allen and Von Steen (June 1969), page 16

9/ Measuring trunk and primary limbs, sample limbs selected and flagged. (1st visit only).

10/ Identification of terminals on sample primaries, selection and flagging sample terminals (1st visit only).

VI. Estimated Fruit Per Tree 11/

The total number of fruit on the sample tree was the first estimate to be considered. The average fruit per block (primary unit) was then expanded to stratum totals which were added together for total valley production.

Estimates of tree production are made from limb counts and measurements. Results from the 1969 citrus research project suggested the use of either an equal probability or a ratio-cluster estimator. The equal probability expansion simply multiplies the fruit count on the sample limbs by the number of terminals and then by the number of primaries. The ratio-cluster estimator utilizes the ratio of sample primary CSA to total CSA to expand the estimated fruit on the primary limbs to the tree level.

The equal probability sample (EPS) estimator may be written:

$$\hat{Y}_{tb} = \frac{N}{n} \sum_{i=1}^{n} \left[\frac{M_i}{m_i} \left(\sum_{j=1}^{m_i} X_{ij}\right) + X_{i\cdot}\right] + X_{\cdot\cdot}$$

where:

$\hat{Y}_{tb}$ is the estimate of fruit on the tth tree of the bth block

N is the number of primaries in the tree

n is the number of primaries sampled

M_i is the number of terminal limbs on the ith primary

m_i is the number of terminals sampled on the ith primary

X_{ij} is the number of fruit on the jth terminal of the ith primary

$X_{i\cdot}$ is the number of path fruit 12/ on the ith primary

$X_{\cdot\cdot}$ is the path fruit on the trunk of the tree

11/ Principle reference on statistical models was a text by William G. Cochran, Sampling Techniques (New York: John Wiley & Sons, Inc., 1963).

12/ Path fruit are those connected directly to primary limbs or the trunk by stems too small to qualify as terminal limbs.

The EPS estimator is unbiased. A terminal limb was defined to be between 0.6 and 1.2 inches CSA so it was reasonable to assume with this narrow range that each primary was a cluster of equal size terminal limbs.

The ratio-primary (R/P) estimator utilizes the CSA measurements of the primary limbs to expand to a tree total. The R/P estimate may be written:

$$\hat{Y}_{tb} = \frac{1}{n} \sum_{i=1}^{n} \frac{A}{A_i} [\frac{M_i}{m_i} \sum_{j=1}^{m_i} X_{ij} + X_{i} \cdot] + X \cdot \cdot$$

where only A and A_i are new notations with A_i being the CSA of the ith primary and $A = \sum_{i=1}^{n} A_i$, the total CSA of all primary limbs on the tree.

The ratio-primary expansion provides a weighted estimate which is biased in this case since the primaries were selected with equal probability. This bias could be removed by selecting the primary limbs with probability proportional to size. Another possibility is to correct for the bias as suggested by Hartely and Ross (1954).[13/] This makes the means and variances of the ratio-primary more comparable with those of the unbiased estimators.

Correlations between primary limb CSA and the estimated number of fruit on the primary from terminal fruit count expansions are presented in table 2. Only three correlations were significantly greater than zero at $\alpha = .05$ level. However, the standard deviations of the CSA variable were generally lower than those of the primary limb counts so the auxiliary measurements may improve the tree estimate.

13/ H.O. Hartley and A. Ross, "Unbiased Ratio Estimates," Nature, 174 (1954), 270-271.

Table 2--Correlation between primary limb CSA and expanded fruit count for each primary limb by type and strata, 1970.

Type	Primary CSA vs. Fruit Expansion		
	Correlations by Strata		
	0 - 3	4 - 7	8 +
Early Oranges	.368	.277	.403
Valencia	.428	.843*	.615*
Grapefruit	.719*	.398	.217

* Indicate that correlation is greater than zero p = 0.95

Correlations between tree fruit count expansions and measures of tree sizes were also computed to check this relationship since this determined the array of the trees for systematic sampling. Results of the 1970 study are presented in Table 3 by type and strata for both the EPS and R/P estimators. It would have been desirable to correlate the actual fruit counts for each sample tree with tree size but these counts were not available. However, judging by the correlations of tree estimates to tree size, there is not a strong relationship in most strata.

Table 3--Correlations between fruit expansions and measures of tree size by strata, 1970 crop year.

Type Stratum	Trunk CSA		Total Primary CSA	
	EPS	R/P	EPS	R/P
Early Oranges				
I	-.623	-.597	-.494	-.464
II	.411	.378	.465	.433
III	.165	.173	.408	.422
Valencia Oranges				
I	.207	.198	.806*	.764*
II	.736*	.577*	.775*	.655*
III	.522	.766*	.431	.802*
Grapefruit				
I	-.054	.160	.095	.283
II	.644*	.546	.653*	.595
III	-.155	-.064	.217	.052

* Indicates that correlation is greater than zero with p = .90

VII. Estimated Total Fruit Produced

The estimated production per tree for each sample block (p.s.u.) is a simple average of the two sample trees in each block. This estimate, whether from EPS or R/P tree expansion, is then expanded to stratum totals. Two possible methods of estimating stratum means and totals are by using an unbiased expansion estimator or a ratio-to-block-size estimator.

The unbiased estimate[14/] of average fruit per tree in the stratum is denoted by:

$$\bar{Y}_u = \frac{1}{k\bar{H}} \sum_{b=1}^{k} H_b \bar{Y}_b$$

where: H_b is the number of trees in the bth block of the stratum

$\bar{H} = \frac{1}{K} \sum_{b=1}^{K} H_b$ is the average number of trees per block in the stratum

K is total number of blocks in the stratum

k is the number of sample blocks in the stratum

Y_b is the average fruit per tree in the bth block

The ratio-to-block-size[15/] is a biased estimate, but the bias becomes negligible as the number of blocks sampled becomes large. This bias can be eliminated by selecting the sample blocks with probability proportional to size. The estimate of average fruit per tree in the stratum is:

$$\bar{Y}_r = \frac{1}{\Sigma k \; Hb} \Sigma k H_b \bar{Y}_b$$

Estimated mean fruit per tree for each stratum with lowest variance relative to cost is the goal when considering procedures and a model for an operational survey.

14/ Cochran, page 304

15/ Ibid., page 300

The effects of improved tree estimates through a ratio estimator are greatly reduced at the strata level by large divisors and finite correction factors associated with the primary limb variance component. The formula used for calculating the variance of the estimated fruit per tree is:

$$S_{\bar{y}}^2 = \frac{S_b^2}{n} + \frac{S_t^2}{na} + (1-\bar{f}_3)\frac{S_p^2}{nab} + (1-\bar{f}_4)\frac{S_w^2}{nabc}$$

where: $S_{\bar{y}}^2$ is the calculated variance of the estimated mean

S_b^2 is the block variance with n blocks sampled

S_t^2 is the tree variance with a the number of trees sampled

S^2p is primary limb variance with b primary limbs sampled per tree

S^2w is the terminal limb variance with c terminals per primary sampled

$(1-\bar{f}_3)$ is the finite correction factor at the primary limb level

$(1-\bar{f}_4)$ is finite correction factor at terminal level

Only the variance associated with the primary limb is affected by using the auxiliary CSA variable since all other expansions within the strata are identical. A considerable reduction of the within tree variance (40-50 percent) will reduce the coefficient of variation for the stratum mean by less than 1 percent. Therefore, the ratio estimator, even corrected for bias, is not recommended at the tree level since additional cost is required for measuring primary limbs. Perhaps a device similar to the terminal selection gauge could be used to define the primary limbs. Equal probability selection and expansion is suggested for estimating fruit per tree.

The ratio to block size expansion of block means to derive the strata average fruit per tree resulted in significant reductions in the coefficients of variation compared to the unbiased estimator. To eliminate the bias from this ratio estimator, it is recommended that the sample blocks be selected with probability proportional to the size of block.

A viable model has now been developed using only objective yield counts on sample limbs. This multistage model expands terminal limb fruit counts to a tree estimate with the EPS estimator and to strata estimates with the ratio-to-block-size estimator.

The estimated total fruit in the valley from terminal limb counts on sample trees is derived through the formula:

$$\tilde{Y} = \sum_{l=1}^{L} Hl \left[\frac{1}{\Sigma\, Hb} \sum_{b=1}^{k} Hb \left(\frac{1}{hb} \sum_{t=1}^{hb} \frac{Nt}{nt} \sum_{i=1}^{nt} \left[\left(\frac{Mi}{mi} \sum_{j=1}^{mi} X_{ij} \right) + X_{i\cdot} \right] + X_{\cdot\cdot} \right) \right]$$

where: $\tilde{Y}$ = total number of fruit in the valley for a given type

Hl = total trees in the lth stratum

Hb = total trees in the bth block

hb = number of sample trees in the bth block

k = number of sample blocks in lth stratum

Nt = number of primaries in the tth tree

nt = number of primaries sampled in the tth tree

Mi = number of terminal limbs on the ith primary

mi = number of terminals sampled on the ith primary

X_{ij} = fruit count on the jth terminal of the ith primary

$X_{i\cdot}$ = number of path fruit on the ith primary

$X_{\cdot\cdot}$ = number of path fruit on the trunk of the tree.

VIII. Fruit Counts From Photography

The next question is whether a better model is possible through the use of tree photography in a double sampling scheme. A larger number of trees can be included in the overall sample using photography at less cost than increasing the size of the limb counting sample by a similar number. It must now be determined whether this auxiliary variable is effective in reducing the variance and whether the reduction is worth the cost.

Correlations between photo counts and limb count data are shown in Table 4 by type of citrus for the three growing seasons studied. Further analysis shows correlations are generally better for smaller trees than large trees. Grapefruit correlations differ greatly between years due to changing fruit sets on the tree from far out on the limbs to nearer the trunk.

Table 4--Photo count and limb count expansion correlations by type, 1968, 1969, 1970 growing seasons16/

Type of Citrus	CORRELATION COEFFICIENT			
	1968	1969	1970	1970(corrected)
Early Oranges	.798**	.684**	.757**	.836**
Valencias	.546	.715**	.534	.592
Grapefruit	.915**	.617**	.288	.427

** Indicates that correlation is greater than zero with p = 0.99

The effect of incorporating the photo counts into two estimating models was examined using 1970 data with correlations correct for error.

16/ 1968 data are for 8 research blocks and are based on actual tree counts. 1969 and 1970 data are for randomly selected blocks using EPS expanded tree totals. The 1969 correlations were not corrected for the sampling errors of the tree estimates but 1970 correlations are presented both corrected and uncorrected for comparison.

The double sampling estimators which were suggested in the 1969 research report were the <u>photo count (p.c.) regression estimator</u> and the <u>p.c. ratio estimator</u>. The p.c. regression estimator for the <u>ith</u> stratum can be written:

$$\bar{\hat{W}}_i = \bar{\hat{Y}}_i + b\,(\bar{X}_{ia} - \bar{X}_i)$$

where:

$\bar{\hat{W}}_i$ is the regression estimated number of fruit per tree in the <u>ith</u> stratum,

$\bar{\hat{Y}}_i$ is the average estimate from limb counts for the <u>ith</u> stratum,

b is the slope of the regression line of photo counts on limb count estimates,

$\bar{X}_{ia}$ is the average photo count for all trees photographed in the stratum,

$\bar{X}_i$ is the average photo count for trees with both photographs and limb counts.

The p.c. ratio estimate for the <u>ith</u> stratum is written as:

$$\bar{\hat{z}}_i = \frac{\bar{\hat{Y}}i}{\bar{X}} \cdot \bar{X}_{ia}$$

where: $\bar{\hat{z}}_i$ is the ratio estimated number of fruit per tree in the <u>ith</u> stratum. $\bar{\hat{Y}}_i$, $\bar{X}_i$ and $\bar{X}_{ia}$ are the same as defined for the regression double sampling estimator.

The variances associated with these double sampling models are as follows:

P.C. Regression: $V(\bar{w}) = \frac{S_y^2\,(1-r^2)}{n} + \frac{r^2 S_y^2}{n'}$ 17/

P.C. Ratio : $V(\bar{z}) = \frac{S_y^2 - S^2 RrS_yS_x + R^2 S_x^2}{n} + \frac{2RrS_yS_x - R^2 S_x^2}{n'}$ 18/.

Where: S_y^2 = variance of estimated tree total from limb count expansion.

17/ Cochran, page 339

18/ Ibid, page 340

r = correlation coefficient between tree estimates and photo counts

Sx^2 = between tree variance of the photo counts

R = ratio of estimated tree total to photo count

n = number of trees in smaller sample having both limb and photo count

n´ = all trees with photo counts (large sample).

The first consideration was the question of how best to combine the data for a double sampling model. A sequential test procedure was used to determine an appropriate regression model for each type of fruit. Trees were grouped by strata to determine whether individual strata models were necessary. Only Valencia oranges gain significantly in explaining additional variation by working with the strata separately. Correlations and variances were, therefore, computed by strata for Valencia and by type for early oranges and grapefruit.

The means and variances for each citrus type under the double sampling models were then computed. The results are presented in Table 5. In order to compare the variance of the limb count sampling model on an equivalent cost basis with the larger samples using photo counts, a cost ratio of 2 photographs to 1 limb count was adopted. The original variance was then divided by the square root of the ratio of adjusted sample size [1/2 (n´+ n)] to the original sample size [n]. This adjusts the variance downward to reflect an increase in sample size.

Table 5--Comparison of estimating models with single and double sampling, Texas Citrus, selected types and strata, 1970 growing season.

Type, Strata	Limb Count Estimates				Regression Model			Ratio Model		
	Mean $\bar{Y}$	Orig. C.V.	Adj. C. V.	Adj.S.E. ($S\bar{y}$)	Mean ($\bar{W}$)	C. V.	($S\bar{w}$) S. E.	Mean ($\bar{z}$)	C. V.	($S\bar{z}$) S. E.
Early Oranges	470.8	21.1	19.5	91.8	430.7	19.3	83.3	427.7	19.7	84.3
Valencias Stratum I	101.88	38.1	35.5	36.2	95.80	31.4	30.1	95.72	31.7	30.3
Stratum II	350.64	12.1	11.3	39.6	352.32	10.6	37.2	353.91	13.0	45.9
Stratum III	260.52	15.3	14.2	37.0	237.19	15.8	37.5	198.15	22.6	44.8
Combined	238.7	10.0	9.4	22.4	226.8	9.5	21.5	209.2	12.1	25.3
Grapefruit	300.6	7.1	6.6.	19.8	298.7	6.8	20.4	291.06	34.2	99.5

Double sampling through the use of the regression estimator provided a 9 percent decrease in the standard error for early oranges. The standard error of the Valencia regression estimate combining all strata was reduced by 4 percent. The correlation for grapefruit was too low to make any gain with double sampling. The estimated average number of fruit per tree was smaller with the regression model than with the limb count model for each type. The standard errors divided by the smaller regression mean results in virtually no difference in the C.V.'s of the limb count model and the regression model.

The standard error of the ratio estimate will always exceed that of the regression estimate and in several instances in Table 5 it exceeds the standard error of the estimate from limb counts only. For early oranges and Valencia stratum I the ratio estimate standard error is nearly as good at that for the regression model and may be easier to use.

IX. Summary

1. Total production for early oranges, Valencia oranges, and grapefruit, can be estimated using fruit counts on sample limbs with a coefficient of variation of 10 percent from a sample of 65 to 85 blocks based on two trees per block, two primary limbs per tree, and two terminal limbs per primary.
2. Correlation between fruit counts on photographs and the limb count expansions indicate that photography can be effectively incorporated into a double sampling regression model for early oranges and for Valencia oranges. The fruit set of grapefruit on the tree is often inside the branches leading to poor correlations between photography and limb count expansions for grapefruit.
3. The recommended model (see page 10) for expanding limb counts to a valley production estimate for each type of citrus begins with an equal probability (EPS) expansion of terminal limb counts to the primary limb and then to the tree total. The average of the two trees in the block is then expanded to a strata total weighted by size of block. Alternatively, sample blocks could be selected with probability proportional to size (PPS) and a constant number of trees per block so the sample would be self-weighing and unbiased. A biased estimate which may be slightly more efficient would use the ratio-to-block-size (R/S) expansion to the strata total.

X. Recommendation for further study

Additional work needs to be done on size and drop estimation. This may be either new research or further investigation of existing data available within SRS or the Weslaco Research Station.

Many discrepancies in the Texas Citrus Mutual list were discovered in the research projects. This list needs further work to get maximum benefits from probability proportional to size (PPS) sample selection and ratio-to-block-size expansions to strata totals.

Further study of the use of aerial photography to count and stratify trees by type and size would be desirable. The 1968 project indicated this may be feasible.[19/] With high and low altitude photography it may be possible to improve or replace the listing of citrus blocks. Stratification of producing citrus trees into only two groups based on canopy area may provide improved estimates. Random sample selection of blocks might be made from aerial photography for each strata across all groves with probability proportional to the size of the block in each citrus grove.

[19/] Allen and Von Steen, page 25