INSTITUTE FOR DEVELOPMENT STUDIES

UNIVERSITY COLLEGE, NAIROBI

Discussion Paper No. 40

THE CALCULATION OF SMALLHOLDER TEA YIELDS IN KENYA BY MULTIPLE LINEAR REGRESSION ANALYSIS.

Dan M. Etherington

December 1966

The Calculation of Smallholder Tea Yields in Kenya by Multiple Linear Regression Analysis.

by Dan M. Etherington [1]

Accurate forecasts of tea yields of smallholder tea in Kenya are needed at both the national and local level because of the growing importance of tea, and increasingly smallholder tea, as an export crop and as a major tool in raising rural incomes in many parts of the country. The Kenya Tea Development Authority finances its operations through international loans which have to be repaid through the cesses on green leaf deliveries. The time period in which these loans are repaid will differ markedly depending on the annual quantities of green leaf delivered to the Authority. The phasing of the construction programme for the new factories serving the smallholders is based on the known planting programme and the estimates of yields. Similarly transport requirements are based on these same estimates.

K.T.D.A. yield calculations are currently based on data obtained from tea estates (as opposed to smallholdings) and the Authority's own tea training farm at Kagochi. However, these data are pretty sketchy, usually referring to mature tea, an amalgam of teas of different ages, or very small samples. Even if reasonable data were available from these sources it is questionable as to whether the results would be applicable beyond a fairly limited ecological area or to areas where tea is under a very different system of management. At the moment the K.T.D.A. makes its calculations on "guesstimates" of what might be reasonably expected in a'reasonable year' with 'reasonable management'. This has been done by a forward projection of the yields achieved to date, on the basis of a yield rating. The yield ratings currently in use (for 1966) are EH; VH; H; M; L; VL; EL; (for Extra High; Very High; High; Medium; Low; Very Low; Extra Low). The yield rating chosen for any area is that which appears to 'fit' best with the actual production obtained so far, and each rating has its own theoretical scale of yield according to age.

---

[1] The author is particularly grateful for the invaluable assistance of the Mrs. S. Levine.

As a means of obtaining yields the normal division of total output by acreage (or some such common input measure) is meaningless in the case of immature permanent crops which are grown over a number of consecutive years. It generally takes from two to three years after planting before any green leaf is plucked and the K.T.D.A. have based their Official Development Plans on tea maturing over a six year period from the time of planting (tea bushes are not, however, fully mature until their ninth or tenth year). Under the Authority's plan, it is quite possible, indeed common, for an individual grower to have planted tea in four or five different years (since 1959) so that his total acreage differs considerably in maturity between the first and last planting. One possible method of calculating the yields of tea of differing maturity is to obtain figures from farmers who have planted tea in only one year, however, for obvious reasons this is unlikely to be a representative sample. So far, then, the Authority, in its planning, has relied heavily on estimates based on the yield ratings, which in turn have been based on very scanty and inadequate evidence and opinion. The only statistical data has been that obtained from the K.T.D.A.'s Kagochi Tea Training Farm. The figures in Table 1 were extracted from the records by the Chief Technical Officer of K.T.D.A.

For the purpose of the annual production forecast made in March 1966, the K.T.D.A. has assumed maturity in the ninth year. This annual forecast is distinct from the Official Development Plan forecast made in 1964 which assumed maturity in six years. The Authority's interpolation of immature yields is based on the percentages in the third column, although for practical purposes these are rounded out to the nearest five percent. While this method has been the only one possible so far, it leaves much to be desired and any improvement on these 'rule of thumb' techniques would be welcomed by the K.T.D.A.

Table 1

Yield of Made Tea Per Acre by Year at Kagochi

Year after Planting	lbs. of Made Tea	% of Mature yield (1200 lbs.)
3rd	200	17
4th	450	38
5th	650	54
6th	850	71
7th	1000	83

Year after Planting	lbs. of Made Tea	% of Mature yield (1200 lbs.)
8th	1150	96
9th	1200	100

Source: K.T.D.A. Technical files.

The K.T.D.A. has in fact a mass of statistical data that can be used to give it the very information it seeks. Among the 20 thousand smallholder tea growers in Kenya, or even the hundred odd delivering to any one buying centre, there are considerable variations in the number of tea stumps planted and the years in which they were planted. It is basically because of these two variations that there are variations in the production of tea between farms within any one ecological area. It would seem reasonable that there should be a statistical relationship between planting time and yield. Where there is the variability that there is among these farms in the number of stumps planted each year, one can estimate an average yield for stumps of different ages.

This paper gives the results of a first attempt at estimating these yield coefficients by means of a multiple linear regression programme which was run for two samples of 104 and 99 farms respectively. The samples were taken as all those growers delivering tea to two buying centres, one in Kisii and one in Kericho District, which had been randomly selected in connection with another study.[2] Planting data were available for each farm from the first year of planting (1959 in these areas) until the 1965 planting. Only those farms which had stumps of three or more years maturity were considered (i.e. a three year time lag was used) since these were the only farms delivering green leaf. Output figures for green leaf deliveries were obtained from K.T.D.A. records for the financial years 1961/2 to 1964/5 (referred to as 1962 and 1965 below). The input and output data were fitted to a multiple linear equation of the form

$$Y_t + 3 = a_i + b_i X_i + b_{i+1} X_{i+1} + b_{i+2} X_{i+2} \dots b_t X_t$$

where

Y_{t+3} = Output of green leaf in the season of year t +3

a_i = intercept

2 The author's Economic Survey of Smallholder Tea in Kenya, an IDS project currently being analysed.

$b_i \dots b_t$ = yield of green leaf in pounds per bush planted in years i .... t

$X_i \dots X_t$ = No. of stumps planted in year i .... t

The data were sufficient for four equations. No attempt has yet been made to amalgamate these production function into a single function since each year is considered unique. However, this will be attempted later possibly using a dummy variable.

The four equations obtainable for each area were

1. $Y_7 = a + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4$

2. $Y_6 = a + b_1X_1 + b_2X_2 + b_3X_3$

3. $Y_5 = a + b_1X_1 + b_2X_2$

4. $Y_4 = a + b_1X_1$

Where subscript 1 = 1959, 2 = 1960, 3 = 1961, 4 = 1962 or 1961/2, 5 = 1962/3, 6 = 1963/4, 7 = 1964/5

The results, given below, are very encouraging. The R^2s in Kericho varied from 0.78 to 0.88 and all coefficients were highly significant at the 1% level. In the Kisii example the R^2s varied from 0.59 to 0.81 with all but two of the regression coefficients being highly significant. One of the other two (b_3 in year 6) was significant at the 5% level while b_4 (year 7) was not significantly different from zero.

The equations obtained are

A. <u>Kericho District</u>

$$Y_7 = -359.2366 + \underset{(0.1542)}{1.5461X_1} + \underset{(0.2182)}{0.9786X_2} + \underset{(0.1564)}{0.9313X_3} + \underset{(0.1405)}{0.8349X_4} \quad (R^2 = 0.7944)$$

$$Y_6 = 13.9380 + \underset{(0.0882)}{1.3946X_1} + \underset{(0.1219)}{0.3940X_2} + \underset{(0.0906)}{0.2745X_3} \quad (R^2 = 0.8426)$$

$$Y_5 = -68.6448 + \underset{(0.0856)}{1.4452X_1} + \underset{(0.1023)}{0.4119X_2} \quad (R^2 = 0.8799)$$

$$Y_4 = -70.6504 + 1.0772X_1 \quad (R^2 = 0.7957)$$

The Standard Deviation of each coefficient and the R^2 for each equation is given in parentheses. All coefficients are highly significant at the 1% level.

B. <u>Kisii District</u>

$$Y_7 = 634.044 + \underset{(0.2799)}{2.2613X_1} + \underset{(0.2037)}{0.8685X_2} + \underset{(0.2512)}{0.8858X_3} + \underset{(0.2918)}{0.2452X_4}^{*} \quad (R^2 = 0.6707)$$

$$Y_6 = 52.27 + \underset{(0.1524)}{1.8091X_1} + \underset{(0.1079)}{0.6046X_2} + \underset{(0.1241)}{0.2983X_3}^{**} \quad (R^2 = 0.7972)$$

$$Y_5 = 12.617 + \underset{(0.1174)}{1.4685X_1} + \underset{(0.0872)}{0.4895X_2} \quad (R^2 = 0.8111)$$

$$Y_4 = 12.386 + \underset{(0.0476)}{0.5751X_1} \quad (R^2 = 0.5888)$$

The above equations do show yields following the expected pattern of increasing output as the tea stumps mature. In the sixth year of maturity in the Kericho sample the tea bushes were yielding the "reasonably expected" amount exactly: 4500 lbs. of green leaf per acre (1.55 x 2904 stumps per acre). In Kisii the mature yield is very high: over 6,500 lbs. of green tea per acre.

The K.T.D.A.'s Kagochi records implied the following comparable equation -

$$Y_k = a + 1.3172X_1 + 1.0072X_2 + 0.6973X_3 + 0.3099X_4$$

where the subscripts 1, 2, 3, 4 now refer to tea planted 6, 5, 4 and 3 years ago. k refers to the present year.

If this is the basis of the K.T.D.A.'s average (medium) yield rating then the equations above do confirm that they are justified in rating Kericho as "High" and Kisii as "Very High". In these two sample areas at least, it seems reasonable to expect mature yields considerably in excess of those originally planned for, if cultural standards are maintained.

The equations show that there is a tendency for the regression coefficients for tea of any particular age to decline as more tea is planted (e.g. in the data above see how third year tea yields fall off with subsequent planting: b_1 in year Y_4, b_2: Y_5, b_3:Y_6, and Y_7 etc). This is perfectly reasonable: cultural attention gets spread to the new plantings and plucking intensity of first plantings is reduced. What are probably more disturbing are the very high coefficients in the early years of the 1959 plantings. Both these observations

---

** Significant at 5% level. * Not significant. All other coefficients are highly significant.

may be explained by the fact that in both these areas theft of tea stumps from the nearby estates was a real problem, particularly in the early phases of the planting programme. Illegal planting of tea would mean incorrect information on the input side and at the same time give an upward bias to yields.

Conclusions:

This first attempt at obtaining yield coefficients for smallholder tea in Kenya shows some interesting and significant results but the analysis needs to be taken further. There is a lack of consistency between some of the coefficients which cannot be convincingly explained away by K.T.D.A. pruning practice. Other sample areas need to be included in the analysis. Alternative functional forms could be fitted. A similar linear equation but with a zero intercept should be tried - particularly as this is the common-sense location of the intercept, however the more mature the tea the less sense there is in forcing the equation through the origin. If the results are to be of any use to the K.T.D.A. in its planning and forecasting, it is essential that some "average" coefficients be obtained for each area. As the analysis stands, the variation between years is such that the authority would still have the difficulty of deciding which coefficients actually to use.