Does Poverty Scoring Work in Ag Value Chains? An Example of Analysis with Kenyan Tea Farmers

Mark Schreiner

20 December 2015

This document is available at `SimplePovertyScorecard.com`.

Abstract

Poverty scoring is a low-cost, transparent way to estimate the consumption-based poverty rates of participants in pro-poor programs. Poverty scorecards are constructed and tested with nationally representative data, so it is an open question whether they also work with sub-national groups such as farmers in agricultural value chains. This paper finds that scoring does indeed work in that Kenyan tea farmers with higher poverty likelihoods also tend to be poorer in terms of other intuitive, well-measured indicators of poverty. The analysis process is explained step-by-step.

Acknowledgements

This paper was commissioned by Frank DiGiovanni of the Ford Foundation. Thanks go as well to Don Seville and Emily Shipman of Sustainable Food Lab (SFL).

Author

Mark Schreiner directs Microfinance Risk Management, L.L.C. He is also a Senior Scholar at the Center for Social Development at Washington University in Saint Louis.

Does Poverty Scoring Work in Ag Value Chains? An Example of Analysis with Kenyan Tea Farmers

1. Introduction

Poverty scoring is a low-cost, transparent way to estimate the consumption-based poverty rates of participants in pro-poor programs. In the case of Kenya, the poverty scorecard (Table 1) is constructed and tested with data from the nationally representative sample of households covered by the 2005/6 Kenya Integrated Household Budget Survey (KIHBS, Schreiner, 2011).

The Kenya scorecard's accuracy is known for nationally representative samples but not for sub-national groups. The general question of the size of errors in poverty-rate estimates for sub-national groups is asked in the context of poverty scorecards by Diamond *et al.* (2015 and 2014) and González Flores (2014) and in the context of poverty maps by Demombynes *et al.* (2008), Elbers, Lanjouw, and Leite (2008), Tarozzi (2008), and Tarozzi and Deaton (2007). While these papers report the size of sub-national errors, they do not set a benchmark for judging when errors are "too big".

This paper asks a more basic question: Does poverty scoring even work at all for sub-national groups? Here, *work* means "tends to assign higher poverty likelihoods to households that truly are poorer", and the sub-national group is Kenyan tea farmers.

This basic question has been answered (Yes) when there is data on consumption (Schreiner, 2011). But without data on consumption, it is difficult to answer.

A draft analysis[1] with data that lacks a measure of consumption suggests that scoring's poverty likelihoods for farmers in agricultural value chains are often weakly—or even backwardly—related with other intuitive poverty indicators (such as land or income). If correct, this would mean that poverty scoring does not work, at least in this type of sub-national group.

Using the same data as that draft study, this paper benchmarks poverty likelihoods against other single-indicator definitions of *poverty*. If the relationships make sense (that is, poverty likelihoods decrease as poverty by other indicators decrease, and the decrease fits theory/logic and intuition), then it is inferred—perhaps somewhat circularly, but nevertheless reasonably—that scoring (and the other indicators) "work".

[1] Shipman (2013). The revised, final version of that draft analysis (SFL, 2014) includes most of the improvements and conclusions discussed here.

If the poverty-likelihood/indicator relationship does not make sense (say, if households with higher likelihoods are more likely to own a television), then one or more of the following is inferred:

- Poverty scoring does not work (that is, it is not sensibly related with other indicators that are intuitively related with poverty)
- The other indicator of poverty does not work
- The other indicator is not measured well
- The data is unreliable
- The true relationship between poverty likelihoods, the other indicator, and poverty is so complex that it does not fit intuition, preventing:
    - Determining whether they work or whether they are measured well
    - Convincing a lay person that the non-intuitive relationship is real

To sum up, if the scoring/indicator relationship in the data fits intuition, then it is inferred that poverty scoring (and the other indicator) "work". If not, then the source of the issue could be with scoring, the other indicator, the data, or intuition.

To eliminate the effects of unreliable measurement, this paper focuses on indicators that are likely to be accurately measured (because they are simple, concrete, and central in households' lives) and that are also commonly used as poverty indicators (due to their intuitive links with poverty and their low cost of collection).

The main result is that scoring works in agricultural value chains (at least for Kenyan tea farmers). That is, there are intuitive relationships between scoring estimates and other well-measured indicators or poverty.

This conclusion differs from the draft with the Kenya tea farmer data because it:

- Corrects errors in the calculation of scores
- Looks at poverty likelihoods (rather than scores)
- Focuses on cross-tabs (rather than correlation coefficients)
- Adjusts income and production to account for household size
- Draws conclusions chiefly based on simple, low-cost indicators that probably are collected reliably and that have intuitive relationships with poverty

Along the way, this paper demonstrates and explains sound analysis approaches.

2. Improving a draft analysis

This paper is a response to a draft analysis in Table 2 (Shipman, 2013): in 12 of 13 cases, poverty scores have weak or backwards relationships with indicators for Kenyan tea farmers. Because these relationships are expected to be strong and intuitive if the scorecard works, Table 2 implies one or more of the following:

- Poverty scoring does not work with Kenyan tea farmers (and perhaps by extension with agricultural value chains in general)
- Intuition about the other indicators is wrong
- The data are not measured reliably
- The calculations have mistakes
- The analysis approach is inappropriate

This paper argues that poverty scoring does work with Kenya tea farmers, intuition is right, and the results in Table 2 are due to mistaken calculations, unreliable data, and inappropriate analysis. Fixing these things reverses the results (except for outcomes that may be measured unreliably).

2.1 Mistaken calculations

Quality analysis starts with quality data and calculations. Here, the data is taken as given; once collected, it cannot be "fixed" (nor is it possible to detect whether it needs to be fixed). Mistakes in calculations, however, can be detected and fixed, and the calculations in the draft analysis behind Table 2 have some big mistakes.

While these errors fixable and while everyone makes mistakes, the point is that the conclusions of an analysis can hinge on prosaic details. Of course, errors can creep in in limitless ways, so it is best to use experienced analysts who have developed processes and habits to prevent and detect errors. For example, the unexpected results in Table 2 are a red flag to check for calculation errors.

The poverty score was computed wrong. The *score* is the sum of the points linked with the ten responses to the ten scorecard indicators (Table 1). But the draft analysis added up responses' *codes*, not their *points*. For example, the four responses to the indicator for the number of rooms are coded 0, 1, 2, and 3, and the corresponding points are 0, 2, 5, and 8. The contribution to the score of a response of "Four or more" should be 8 (the response's points), but the draft calculation used 3 (the response's code).

In addition, the original score calculation also adds in the values of the codes of two survey items (electrification and source of water) that are not even in the scorecard.

With errors, scores range from 4 to 55; when corrected, they range from 6 to 96. The fix changes estimated poverty rates a lot.

Poverty line[2]	**Poverty rate with errors (%)**	**Poverty rate after fix (%)**	**All-Kenya reference poverty rate in 2005/6 KIHBS (%)**
100% of national	60.0	18.1	37.9
150% of national	81.9	39.8	60.7

$n = 564$ for the Kenya tea farmer survey, and $n = 12{,}644$ for the 2005/6 KIHBS

Why harp on this? Not to berate the draft analysis; all analyses have mistakes. The point is that even the most basic analysis steps demand a lot of care. And later calculations in the analytical process are far more complex and error-prone.

So analysts should be careful, go slow, and check their work, especially if the results seem odd. Here, fixing errors matters a lot, reducing the estimated poverty rate of tea farmers from 60.0 to 18.2 percent (national line) and revealing that tea farmers are half as likely to be poor—not much more likely—than the average Kenyan.

Other calculations behind Table 2 are also wrong. This reinforces the point that managing data well, writing correct spreadsheet formulas, and reporting results accurately cannot be taken for granted.

- "Access to electricity" has its codes reversed (see below)
- "Net income from tea" is incorrectly computed as the amount sold multiplied by the amount sold (rather than the amount sold multipled by the price of tea)
- The number of communication devices is computed wrong
- The reported correlation of +0.05 for scores and "Net income from tea" is actually the correlation for scores and "Amount of tea sold".

So four of the 13 outcomes in Table 2 are computed wrong.

The draft analysis does not explain its labels. Table 2 labels the correlations between scores and outcomes as "None or negligible", "Weak negative", or "Strong positive". But what do these labels mean?

Higher scores imply lower poverty likelihoods, and higher values of outcomes imply less poverty, so positive correlations suggest that scoring has an intuitive relationship with the outcome indicators and thus works.

[2] This paper focuses on Kenya's national poverty line and 150% of the national line. These lines are more relevant in Kenya (and easier to explain) than, for example, the $125/day 2005 PPP line.

But what do "weak" and "strong" mean? The draft analysis does not say. Perhaps the labels are based on the "statistical significance" of the estimated correlation coefficients for some unstated confidence level; if so, it is a mistake, albeit a common one in academic work (McCloskey, 1985; Cowger, 1984).[3] Or perhaps the labels come from eyeballing the results, in which case an explicit benchmark is needed. Why is a correlation as large as –0.28 or –0.27 "weak"? Why is +0.54 "strong"? Why is +0.04 (let alone +0.19) "none or negligible"? The draft's analyst has some standard in mind, but because it is left unstated, its relevance cannot be discussed or assessed. Analysis is meaningful—and thus potentially useful—not because it stands on numbers but because it can be productively discussed. The essence of both good science and good management is being clear about the assumptions and judgments that lead to conclusions. Rigor is not certainty, but rather transparency about uncertainty.

2.2 Improvements in the analysis approach

Even if the draft's calculations and analysis were error-free, the analysis approach can still be improved.

Analyze poverty likelihoods, not scores. The draft analysis uses scores, but it should use poverty likelihoods.[4] Scores are ordinal symbols, like letters in the alphabet or colors in the spectrum. As ordinal symbols, 10 is less than 20, and 17 is less than 27, but the "distance in poverty units" between a score of 10 and a score of 20 is not the same as the "distance in poverty units" between a score of 17 and a score of 27. Thus, scores cannot be meaningfully added up or averaged.[5]

In contrast, poverty likelihoods are cardinal numbers; they can be added up and averaged. The "distance in poverty units" between poverty likelihoods of 10 and 20 percent is the same (10 percentage points) as between likelihoods of 17 and 27 percent.

This matters, for example, because the estimation of Pearson correlation coefficients assumes that its inputs are cardinal. If an input is ordinal (as are scores), then Spearman correlation coefficients are appropriate.

---

[3] In analysis that aims to inform decisions (which should be the only type of analysis), the standard for the usefulness of an estimate is not the statistical risk that the sample size is too small to take the estimate at face value but rather whether the result is "actionable" in that knowing it—accounting for the known risks related with its accuracy—has material effects on a real-world decision (Schreiner, 2015).

[4] Using scores instead of poverty likelihoods is the most-common mistake in scoring analysis. While scorecard points are added up to get scores, it does not follow that the scores themselves can be added or averaged.

[5] It is easy to forget that scores are ordinal because the symbols used to represent scores are not letters or colors (which clearly cannot be added or averaged) but rather numbers. Of course, when numbers are counts of something, then they can be added up or averaged. Scores represented by numbers, however, are not counts but rather markers of the relative order of a household in the distribution of poverty likelihood.

The best solution? Convert scores to poverty likelihoods (using the look-up tables in each scorecard's documentation), and do all analysis in terms of poverty likelihoods.

Use cross-tabs, not correlation coefficients. Correlation coefficients boil relationships down to a single number, but that often obscures important features. Cross-tabs reveal more while still being simple to understand.

More important, non-specialists can understand cross-tabs better than correlation coefficients (or regression coefficients, or other more complex statistics).

When relevant, adjust other indicators for household size. Divisible indicators (such as land used or net income, but not television ownership nor the source of water) should be adjusted to account for the different "needs" of households of different sizes. For example, two Ha of land may be enough for two people but insufficient for ten.

Kenya's official consumption-based definition of *poverty* divides a household's consumption by the number of its adult equivalents, recognizing that a household's poverty depends in part on the number of mouths it has to feed.[6] This matters because Kenya's poverty scorecard is based on this definition of per-adult-equivalent consumption. Comparing (say) total household income to poverty likelihoods is inappropriate because it incorrectly assumes that a given income has the same effect on poverty for a household of two as for a household of ten.

Focus analysis on outcomes that are likely to be well-measured. An explicit factor in the selection of indicators for the poverty scorecard is that they be well-measured. For example, a household knows whether it has a television.[7] In contrast, net income is notoriously ill-measured. Households (and especially farmers) often do not know their net income, as it is the difference between many sources of gross income (some of which do not involve cash changing hands) and many types of expenses, with both grosss income and expenses accumulated over many events in a long period. It is difficult to measure income (or worse, consumption) well, which is why the gold-standard measurements from national household surveys take a day or more to complete.

Focusing on well-measured outcomes also reduces the risk that:

- Ill-measured outcomes will distort estimates
- Scoring's ability to work will be confused with the quality of measurement

[6] Kenya's formula for the number of adult equivalents is not documented, so the analysis here divides by the number of household members.

[7] A household can lie about having a television, but it also knows that an enumerator can easily check or that on-lookers might snicker at the lie, giving them away.

3. New analysis of scoring's estimates and other poverty indicators for Kenyan tea farmers

Using corrected calculations, this section presents an improved analysis. Its estimates are more understandable and informative, and they also lead to a different conclusion (that scoring does indeed work with Kenyan tea farmers).

3.1 Cross-tabs of poverty likelihoods with well-measured categorical outcomes for Kenyan tea farmers

This sub-section uses cross-tabs to juxtapose scoring-based poverty rates (that is, average poverty likelihoods) with other well-measured indicators of poverty:

- Whether the farmer is female
- Whether the household has electricity
- The household's source of water
- Whether the household has a telephone
- Whether the household has a television
- Whether the household has a radio
- Number of cows owned by the household
- Whether all boys ages 5 to 12 attend school at least 80 percent of the time
- Whether all girls ages 5 to 12 attend school at least 80 percent of the time

Among these, the best non-scoring poverty indicators for Kenya are:

- Television ownership
- Source of water

Whether the farmer is female

Female tea farmers have about the same scoring-based estimated poverty rate as male farmers. This may be because the "farmer" is not necessarily the "household head." (The tea farmer survey does not record the sex of the household head.)

Sex of farmer	***n***	**Natl. line**	**150% natl. line**
Male	404	18.5	39.6
Female	160	17.3	40.4

Source: Tea farmer survey

Contrary to intuition, female-headed households in Kenya's 2005/6 KIHBS—as in many countries—are less likely to be poor than male-headed households:

Sex of head	***n***	**Natl. line**	**150% natl. line**
Male	11,156	40.6	63.9
Female	1,488	16.0	35.0

Source: 2005/6 KIHBS

Given this unexpected (but real) relationship, and given that the sex of the tea farmer is not linked with scoring's estimate of poverty, female headship does not work as an indicator of poverty.

Whether the household has electricity
Whether the household has a television

Tea farmers with electricity are much more likely to be poor than tea farmers without electricity.

Electrified?	***n***	**Natl. line**	**150% natl. line**
Missing data	21	29.9	52.3
No	213	8.1	23.4
Yes	330	23.9	49.7

This strange result is a strike against poverty scoring. But before concluding that scoring does not work, it makes sense to double-check the data and calculations.

The electrification/poverty relationship in the tea-farmer survey can be triangulated with that same relationship in the 2005/6 KIHBS using its consumption-based definition of *poverty*. In the KIHBS, electrified households are much less likely to be poor (11.6 percent) than non-electrified households (43.2 percent). So either tea farmers are exceedingly unusual, or something is wrong with the data.

Could the codes for electrification in the tea farmer data be reversed? Televisions require electricity, and a cross-tab of electrification with television ownership in the tea farmer surveys gives non-intuitive results:[8]

	Owns a television?	
Electrified?	No	Yes
No	25	212
Yes	295	48

It is not possible that most television owners lack electricity.

So either television ownership or electrification is miscoded. Which one is it? The cross-tab of television ownership with scoring's poverty-rate estimates makes sense, and that suggests that it is electrification that is miscoded.

Owns a TV?	***n***	**Natl. line**	**150% natl. line**
No	314	26.2	52.6
Yes	248	7.7	23.6

Fixing the codes results in strong, intuitive relationships between scoring's estimates and two non-scoring indicators that are probably collected reliably.

[8] The 2005/6 KIHBS asks about electrification but not about television ownership.

In particular, non-electrified tea farmers are three times more likely to be below the national poverty line than electrified farmers, and tea farmers without a television are four times more likely to be below the national line than tea farmers who own a television.

These strong, intuitive links between scoring and other outcome indicators suggest that both scoring and the other indicators work.

The draft correlation for scores (incorrectly computed) and electrification (incorrectly coded) is –0.25 and called "weak". With corrected data, the correlations are:

	Natl.	150%
Between scoring's poverty likelihoods and electrification:	–0.36	–0.49
Between scoring's poverty likelihoods and television ownership:	–0.46	–0.57

These make sense. Tea farmers with "more" electricity or "more" television are less poor. Still, the cross-tabs show both direction and magnitude more clearly.

Source of water

Water source[9]	***n***	**Natl. line**	**150% natl. line**
Pond/river	194	30.3	56.8
Unprotected well	27	30.2	58.5
Protected well	91	16.6	39.1
Piped water to house	251	8.4	25.1

Scoring and the water source work as expected; better water sources are linked with lower estimated poverty rates.

The cross-tabs show that the magnitudes are large; for the national line, having an unsafe source (pond/river, or unprotected well) is linked with at least twice the poverty risk as safe sources (protected wells, or piped water to the house).

The draft analysis—using incorrect scores incorrectly—produces a "strong and positive" correlation of +0.54. But the calculation incorrectly assumes, for example, that the difference between "pond/river" and "unprotected well" is the same as the difference between "protected well" and "piped water to house".

[9] The response options in the tea farmer survey do not cover all common sources of water in Kenya. In particular, they omit bottled/sachet water and public standpipes.

Telephone ownership

Telephone ownership has an intuitive relationship with scoring's poverty-rate estimates, suggesting that these two indicators work.

Owns telephone	***n***	**Natl. line**	**150% natl. line**
No	41	29.1	48.9
Yes	521	17.2	39.1

About 12 in 13 tea farmers live in households that own a telephone. Compared with owners, non-owners are about 33 percent more likely to be poor.

Almost all tea farmers own telephones, so it is not a useful indicator of poverty. It only rarely helps to distinguish among households' poverty.

Radio ownership

Radio ownership is like telephone ownership; non-owners are poorer (in line with intuition), but the indicator is not useful because almost all tea farmers own radios.

Owns radio	***n***	**Natl. line**	**150% natl. line**
No	26	37.2	60.3
Yes	536	17.1	38.8

Radio ownership is also a weak indicator because as development progresses, its relationship with poverty reverses. As incomes rise, only the poorest households in the most rural areas still use radios, to the point where having a radio (instead of a stereo or an MP3 player) is a signal of greater poverty.

Number of communication devices

This indicator is problematic because it adds together radios, telephones, and televisions, forcing each of these different items to have the same link with poverty.

The draft analysis reports a "none or negligible" correlation of –0.15 between "number of communication devices" and the (incorrect) score. Without a benchmark or context, however, it is difficult to say whether –0.15 means "no relationship".

# devices	***n***	**Natl. line**	**150% natl. line**
None	10	49.5	68.7
One	37	29.8	54.0
Two	277	24.2	50.6
Three	238	7.8	23.7

In stark contrast to the draft label of "none or negligible", the rare tea farmer with no devices is about twice as likely to be under the national poverty line than farmers with two devices and three times more likely to be under the national line than farmers with three devices.

Most tea farmers have two or three devices. Those with two devices are about three times as likely to be poor than those with three devices. This is a strong, intuitive relationship; the indicator's link with poverty is far from "negligible".

Number of cows

Most tea farmers have zero, one, or two cows, but a few have many more. In general, additional numbers of a given asset have a progressively weaker relationship with poverty, so the change in poverty risk going from (say) no cows to one cow is a lot larger than going from (say) four to five cows.

The Pearson correlation coefficient assumes a linear relationship. Cross-tabs do not, so they can reveal non-linear relationships. Indeed, the linear correlation may be weak even if the true relationship is strong.

# cows	*n*	Natl. line	150% natl. line
None	129	13.0	31.1
One	143	21.5	45.5
Two	111	16.2	37.6
Three	59	16.3	39.3
Four	48	23.8	45.1
Five or more	70	20.7	43.9

The draft analysis with incorrect scores shows a "weak and negative" correlation of –0.28; more cows are intuitively linked with higher scores (lower poverty likelihoods). The relationship here, however, is less clear-cut and more "complex". So either:

- Scoring (or cow ownership) does not work as a poverty indicator, or
- Multiple forces linked with cow ownership are at play, pushing in different directions

The relationship between cow ownership and poverty is non-intuitive in many national consumption surveys, suggesting that multiple factors are at work. While this comes up too often to be an ex-post rationalization, its drivers are still not clear.

Perhaps tea farmers without cows are less likely to be fully agricultural. For example, they may live in peri-urban areas, or have household members with non-farm jobs. In most countries, owning cows signals more poverty risk—vis-a-vis the average household—because it signals greater dependence on agriculture, and agriculture is linked with greater poverty. So while owning a cow signals less poverty for farmers, it may not fully compensate for the disadvantage of being a farmer in the first place.

In the tea farmer data, going from none to one and then one to more shows no clear pattern. While intuition would suggest that more cows signal less poverty for farmers, the scoring estimates do not show this. (In the last three ownership categories, the small samples may allow sampling variation to obscure patterns.)

So either scoring does not work, or the number of cows (which is likely well-measured) has a more complex relationship with poverty than expected. Because scoring works with other well-measured indicators, it seems likely that cows have a non-intuitive—albeit real—relationship with poverty.

School attendance by boys and girls ages 5 to 12

This is an odd question in the tea farmer survey. First, it is asked separately by the sex of the child. In poverty scorecards, an "all children" indicator of school attendance always does better than sex-segregated indicators.

Second, it asks about attendance "at least 80 percent of the time". No national surveys ask about school attendance in this way.

Third, it is complex. Looking at the data, it seems some households get confused and report the share of days attended by members, not the number of members who attend 80 percent of the time. And some tea farmers may not grasp percentages.

Fourth, the question asks about "Of all your children . . .", but it should ask "Of all household members . . .".

Fifth, the use of sex-specific indicators it makes it more likely that there are no members in the age range. This makes the indicator relevant for fewer households.

The sixth (and most banal) issue is that it is not clear how households with no school-age children are coded in the data.

All this decreases the usefulness that can be expected for this indicator.[10]

Do all boys 5 to 12 go to school?	***n***	**Natl. line**	**150% natl. line**
No	143	25.5	50.7
Yes	214	20.6	44.7
No boys 5 to 12	207	10.5	27.3

Do all girls 5 to 12 go to school?	***n***	**Natl. line**	**150% natl. line**
No	117	26.6	52.1
Yes	199	21.9	47.4
No girls 5 to 12	248	11.1	28.0

The data can be combined into an "all-children" indicator.

Do all boys and girls ages 5 to 12 go to school	***n***	**Natl. line**	**150% natl. line**
No	171	23.9	48.6
Yes	222	20.2	43.8
No one 5 to 12	171	9.7	26.1

[10] In general, questions in bespoke surveys should mimic similar questions in national surveys. Most questions in national surveys are designed and defined by survey professionals. Matching them also allows for later apples-to-apples comparisons between results from the national and bespoke surveys.

The cross-tabs show that:

- Poverty is greater in households where some school-age children do not go to school. This intuitive pattern holds for boys, for girls, and for boys and girls together
- Poverty rates are lowest in households without school-age children

Households without school-age children are less likely to be poor both because there are fewer "unproductive" mouths to feed and because poverty generally decreases as households age, and older households have fewer school-age children.

As for all indicators so far except "number of cows", scoring's estimates fit intuition.

3.2 Cross-tabs of poverty likelihoods with less-well-measured continuous production outcomes for Kenyan tea farmers

Continuous indicators of production or income are measured less well than those discussed so far. They require recalling and adding up flows in the past that are irregular in both time and quantity, so answers tend to be rough guesses.

Still, these indicators are common because, if measured well, they are direct drivers poverty and they sum up a household's production/control/use of resources.

The question is whether their value as direct indicators compensates for their lower reliability.

The Pearson correlation coefficient assumes continuous indicators, so that assumption is no longer violated. Still, the analysis here uses cross-tabs because they do not assume linearity and because they are easier to understand. For cross-tabs, the values of continuous indicators are ordered and then grouped into three, four, or five bins, each with roughly the same number of households.

As before, if the link between scoring's estimates and other indicators fits intuition, then it implies that scoring and the other indicators "work". If not:

- Scoring does not work (contrary to evidence above for well-measured indicators)
- The continuous indicators are not reliably measured
- The continuous indicators are reliable, but their true relationship with poverty is more complex than intuition, rendering them unhelpful as indicators of poverty or as indicators of the usefulness of scoring

The draft analysis looks at these continuous indicators:

- How much land do you farm?
- How much land is in green-leaf tea?
- How much green-leaf tea did you harvest last year?
- Net income from tea
- Gross income from non-tea cash crops?
- Gross off-farm income?
- Total net household income?
- In the past year, did the household always have enough food?

Land farmed

Other than household labor, land is a farmer's main productive asset. Of course, the poverty-reducing capacity of a hectare of land depends strongly in its soil quality, slope, irrigation/precipitation, age of plants, etc.

Land area is probably measured reliably, as land is a concrete thing central to farming that tends to change slowly (if at all) year to year. On the other hand, farmers may not know the quantity of their land in units of hectares or even in local units.

Ha. ag. land	***n***	**Natl. line**	**150% natl. line**
0.00 to 0.50	102	11.2	29.8
0.51 to 1.00	147	17.1	37.5
1.01 to 2.00	124	21.2	44.4
2.01 to 4.00	120	20.3	44.7
4.01 or more	71	21.0	42.0

Unexpectedly, poverty rates increase with more land (up to 1 Ha), then level off. The draft analysis sees this as a sign that scoring does not work. But it could be that:

- Land is not measured reliably
- Land has a complex, non-intuitive relationship with poverty

One way to distinguish among explanations is to compare land with other poverty indicators for which data quality and the relationship with poverty is more certain.

Ha.	***n***	**Electricity?**	**TV?**	**All in school?**	**Protected/piped H_2O?**
0–0.50	102	41	45	60	78
0.51–1	147	41	41	59	58
1.01–2	124	41	39	60	49
2.01–4	120	48	51	61	63
≥4.01	71	45	52	27	63

Table note: Figures are *not* average poverty likelihoods, but rather the percentage of households in a given land-area bin who have electricity, a TV, all children ages 5-to-12 in school, or protected/piped water.

Electrification increases with ag land (once above 2 Ha), but it is constant in the three bins at 2 Ha or below. In contrast, electrification increases as scoring's poverty likelihoods decrease. These three patterns (scoring/electrification, scoring/land, and land/electrification) are not consistent. Given that the scoring/electrification pattern makes sense, either the land data is not well-measured or the true pattern is more complex than intuition would suggest. Perhaps larger farms are farther from urban centers and thus are less likely to be electrified, regardless of the household's poverty.

Television ownership decreases as agricultural land goes from zero to 2, then increases. This is difficult to explain, if more land is linked with less poverty.

School attendance is constant until it drops sharply for the largest farms. This is non-intuitive, but perhaps land-rich households pull children out of school to farm.

The quality of the water source decreases as ag land goes from none to 2 Ha, and increases with more ag land above 2 Ha. This is the same pattern as for television ownership, and the opposite of the scoring/land pattern. This is more evidence in favor of a complex relationship between land and poverty.

Given that poverty scoring works with well-measured indicators, and that land sometimes has non-intuitive relationships with other indicators of poverty, it seems likely that land has a non-intuitive relationship with poverty.

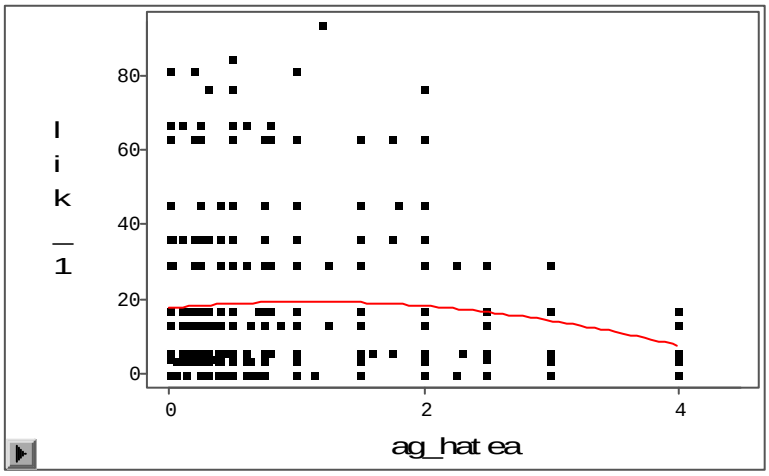
Land in tea

Land planted to green-leaf tea is a subset of all ag land. Thus, it is both more likely to be reported accurately, and—for a given level of reliability—less closely linked with poverty (because it ignores other land).

Ha. in tea	*n*	Natl.	150% natl.	Electricity?	TV?	All in school?	Good H_2O?
0.00–0.20	117	20	41	42	42	39	56
0.21–0.49	114	15	36	32	34	62	67
0.50–0.99	145	19	40	42	44	56	61
1.00–1.99	111	21	44	44	46	59	56
≥2.00	76	16	39	58	63	67	78

Scoring's poverty rate across tea-land bins can be described either as roughly constant or as down/up/down. Both patterns are non-intuitive.

A quadratic regression line superimposed on a scatterplot of poverty likelihoods versus Ha in tea (dropping an outlier with 1,150 Ha and seven with 5 Ha or more) is:



This suggests that poverty is roughly constant before falling in the upper decile or so of Ha in tea.

Electrification follows the same down/up/down pattern as scoring's estimates, except that the largest tea farmers are much more likely to have electricity.

Television ownership follows the pattern as electrification.

School attendance differs from everything else, going up/down/up.

Having protected/piped water roughly follows the pattern of scoring, electrification, and television.

All in all, scoring and the other indicators consistently have no consistent relationship with Ha in tea, other than that the largest decile of tea farmers (≥2.0 Ha) are the best-off. So if scoring is off, then so are other intuitive, well-measured indicators.

How can this be? Surely land in tea is linked with tea farmers' production and income. Perhaps land in tea is not the only—or even the main—factor driving poverty, so the land/poverty link is obscured by the effects of factors—correlated with both tea land and poverty—that are not accounted for here.

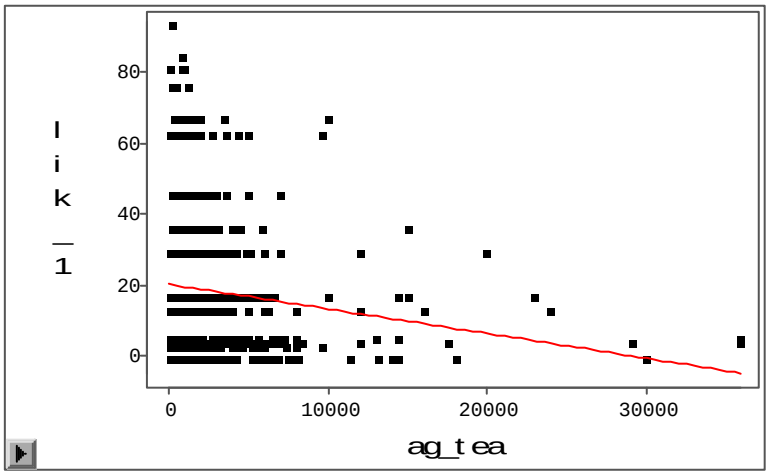
As noted above and as discussed more below, it may make sense to normalize for household "needs" when analyzing income-producing assets (such as ag land or tea land) or components of income (such as total net income or income from tea).

Tea harvested

Tea harvested in the past year may be a better indicator of poverty than land in tea because it implicitly controls for land quality. Also, production outputs should be more closely linked with poverty than production inputs.

Kg. tea	*n*	Natl.	150% natl.	Electricity?	TV?	All in school?	Good H_2O?
0–600	119	24	47	31	25	55	44
601–1,199	109	20	44	25	26	60	54
1,200–1,999	107	18	40	43	45	57	62
2,000–3,999	110	17	38	45	53	44	74
≥4,000	116	12	31	67	71	63	75

This is the most reliable measure so far of (partial) income, and it has sensible relationships with scoring and with other well-measured indicators (except school attendance). That is, larger tea harvests are consistently linked with lower poverty likelihoods. The scatterplot (omitting a 72,000 kg outlier) also shows this.



Kg of tea harvested is a better poverty indicator—because it is more directly is related with income—than is hectares of ag land or tea land.

This is again consistent with the idea that the lack of an intuitive relationship between farm land and poverty indicators is not a signal that the indictaors do not work but rather than land has a complex, non-intuitive relationship with poverty.

What if the tea harvest is adjusted for household "needs" by dividing by the number of household members? Not only does Kenya's consumption-based poverty line adjust for household size, but it also makes sense to expect that the poverty effects of a given harvest to differ by household size.

Kg. tea per person	***n***	**Natl.**	**150% natl.**	**Electricity?**	**TV?**	**All in school?**	**Good H_2O?**
0–124	113	28	54	26	21	59	37
125–249	105	25	51	27	28	58	43
250–499	117	18	40	35	39	47	64
500–999	110	13	33	51	54	56	75
≥1,000	116	8	23	66	75	61	81

These intuitive relationships are stronger than before. Normalizing income (or production) for household "needs" strengthens indicator/poverty links. This make sense, because a confounding factor (household size) is now accounted for.

The relationships are "stronger" because scoring's estimates (and the values of other poverty indicators) change more across bins of tea harvested per person than they do across bins of tea harvested per household.

For example, the estimated poverty rate for the national line with per-household tea harvest ranges from 24 percent (first bin) to 12 percent (fifth bin). For per-capita tea harvest, the range is 28 to 8 percent.

The gradient is also sharper for electrification, television ownership, and source of water. (The pattern fails for school attendance, but this is itself a consistent pattern, suggesting that school attendance does not work as a poverty indicator.)

From now on, this paper will adjust for household size.

Net income from tea (annual, calculated)

This indicator may be less useful than kg of tea harvested per person because:

- It combines several financial values (units of tea sold multiplied by price per unit, less expenses for labor and inputs), none of which are particularly well-measured
- It depends on the price of tea. In the data, prices vary from 37 to 55 KES/kg
- Its poverty importance depends on tea's share in a household's total income

On the other hand, if prices are accurate and reflect tea quality, then net income from tea might be useful.

Net tea income/ capita	***n***	**Natl.**	**150% natl.**	**Electricity?**	**TV?**	**All in school?**	**Good H_2O?**
Missing	91	13	31	61	68	63	77
Negative	29	20	37	36	52	24	48
1 to 2,999	65	24	49	33	33	53	42
3,000–8,999	101	23	47	33	35	58	52
9,000–17,999	92	19	42	45	44	51	62
18,000–34,999	97	17	39	36	38	62	71
≥35,000	89	14	33	44	40	55	67

The indicators are less consistent with each other than there were for tea harvested. While there is a sensible gradient for scoring's estimates (ignoring bins for missing data and negative net income), the gradient is flatter than for the harvest.

The gradients for the non-scoring indicators are also less sensible than for the tea harvest. Households with 18,000–34,999 KES per person are worse off than those in the bins just above and below. As usual, school attendance has a non-intuitive pattern. More concerning, electrification and television ownership show only slight gradients. The second-best measure—after scoring—is the source of water, but even there the gradient is not steep and flattens in the top two quintiles.

On net, accounting for prices and expenses seems to add noise. In principle, net income from tea is a better indicator than the gross harvest. In practice, however, measurement issues get in the way. Non-scoring indicators show weaker and less-uniform gradients than the easier-to-measure (but less directly relevant) tea harvest.

Gross income (annual) from non-tea cash crops

This indicator matters to the extent that the household has non-tea income and to the extent that the gross measure is more reliable than the net measure.

Gross non-tea income/ capita	***n***	**Natl.**	**150% natl.**	**Electricity?**	**TV?**	**All in school?**	**Good H_2O?**
Missing	117	16	35	55	62	62	70
Zero	245	16	38	40	40	56	64
1 to 19	73	20	43	38	39	57	53
20 to 50	62	28	50	40	37	45	43
≥60	67	18	40	38	42	48	66

Three patterns stand out. First, households with missing data or no non-tea crop income are the least-poor by all indicators.

Second, gross non-tea crop income is essentially zero for the vast majority of tea farmers. At KES87 per USD1, the highest decile has gross non-tea crop income per year per person of less than $1.00. This also may signal some error in recording/processing.[11]

Finally, the gradients across categories are roughly flat and non-intuitive.

All this suggests that this indicator is unreliable and not strongly and intuitively linked with poverty. As such, it is not capable of indicating whether scoring works.

Given that gross income from non-tea cash crops is negligible, the analisis of net income from non-tea cash crops is omitted.

[11] It also bodes ill for data quality that about one-fifth of households have missing values and that these households are better-off.

Gross off-farm income (annual)

Off-farm income could signal less poverty (agriculture typically "pays" less than non-agriculture) or more poverty (if the off-farm work is day labor).

Off-farm income/ capita	***n***	**Natl.**	**150% natl.**	**Electricity?**	**TV ?**	**All in school?**	**Good H_2O?**
Missing	91	13	31	61	68	63	77
Zero	237	19	41	41	41	58	58
1 to 6,000	69	26	50	30	40	45	51
6,001 to 19,999	75	21	43	37	38	47	61
≥20,000	92	13	35	38	35	55	67

Ignoring households with missing values, the pattern is that low (but non-zero) off-farm income is associated with higher scoring-estimated poverty than having no off-farm income (possibly due to poorer farmers hiring themselves out as day laborers). At the same time, having more off-farm income—given that there is some off-farm income—is linked with less poverty (perhaps because these households do not depend solely on tea because a member has a non-agricultural job other than day labor).

This case shows the danger of assuming a linear relationship. The Pearson correlation coefficient forces linearity and is +0.19, but cross-tabs reveal non-linearity.

The pattern of electrification fits the pattern of scoring.

Unexpectedly, the pattern of television ownership runs counter to that of scoring and of electrification. Part of this may be due sampling variation in the last three bins, where samples are small.

But something else is afoot; for households with missing data, 68 percent report having a television, but only 61 percent report having electricity. For per-capita off-farm income of 1–6,000, 40 percent have televisions, but only 30 percent have electricity. This is especially puzzling because television ownership in other cross-tabs increases as poverty decreases.[12] (Or maybe I made an error, although I triple-checked.)

The relationship between gross off-farm income and school attendance fits with scoring, with more off-farm income—if there is any—linked with better attendance.

The source of water also fits with scoring.

In sum, electricity, schooling, and water fit scoring, but television does not. No off-farm income—or more off-farm income (given that there is off-farm income)—is associated with less poverty risk. This is plausible, but it is non-intuitive and takes some explanation, so off-farm income does not work as a poverty indicator.

[12] About one in seven households who own a television report not having electricity. Some of these are surely cases of bad data, but others may be correct. After all, television is a status item, and some households use car batteries or pirated electricity.

Total net household income (annual)

This single question tries to capture a basic quantity of interest: net income from all sources. Of course, the data may be unreliable. The respondent may not know, and he/she has a host of social/psychological reasons to dissemble. This is why national income surveys measure total net income with a lengthy battery of questions. Furthermore, consumption is a better indicator of general well-being, especially for farmers who consume some of their own produce.[13]

Still, a single question is better than nothing, it is low-cost, and it may in fact do well, so it makes sense to test it.

Net income/ capita	***n***	**Natl.**	**150% natl.**	**Electricity?**	**TV ?**	**All in school?**	**Good H_2O?**
Missing	56	19	42	49	56	60	73
0–5,999	93	27	53	21	20	66	40
6,000–11,999	99	27	54	22	23	53	45
12,000–19,999	91	23	48	47	36	47	55
20,000–39,999	114	11	30	52	57	55	70
≥40,000	117	6	21	62	71	51	88

Scoring's estimates consistently decrease as per-capita net income increases. This makes sense, so both scoring and the single-question measure of income work.

The gradient is not steep in the lowest three non-missing bins, perhaps because some of the inflow of resources (income) for the poorest farmers is in-kind and so may not be counted. Still, the pattern is clear and intuitive.

Likewise, the patterns for electrification, television, and source of water fit; higher income is linked with less poverty. (As usual, school attendance does not fit.)

Scoring's poverty likelihoods have a strong, intuitive relationship with total net income, the indicator that most people associate first with poverty. This strongly recommends scoring.

[13] The Kenya poverty scorecard estimates consumption-based poverty, not income-based poverty. The two are highly and positively correlated, but not perfectly correlated.

Hunger

In principle, this indicator is promising; less poverty should mean less hunger. It also allows households to define their own "needs". Of course, this may also be a weakness, as is the fact that the response is based on recall.

About 80 percent of tea farmers report never being hungry in the past year.

Ever hungry?	***n***	**Natl.**	**150% natl.**	**Electricity?**	**TV?**	**All in school?**	**Good H_2O?**
Yes	113	33	57	25	22	29	29
No	451	14	35	47	51	64	71

All the relationships are strong and intuitive. Scoring's estimated poverty rate for the ever-hungry is more than twice that of the never-hungry. The same pattern holds for the four other indicators.

Ever/never hungry works, with intuitive results in line with other poverty indicators that also work.

4. Conclusion

Does poverty scoring work with sub-national groups such as farmers in ag value chains? Yes; Kenyan tea farmers with higher estimated poverty likelihoods also tend to be poorer in terms of other intuitive, well-measured indicators of poverty.

This result differs from that of the draft analysis with this data that prompted this paper. The differences are due to:

- Fixing calculation errors
- Taking a better analysis approach
- Adjusting for household size

Of course, poverty scoring is not perfect, and it is not "good enough" in all contexts nor for all purposes. After all, Kenyan tea farmers are just one sub-national group, and the standard for "working" here is only that scoring's results go in the same direction as those of other intuitive, well-measured poverty indicators.

This analysis is a single data point, and its results may not extrapolate beyond Kenyan tea farmers. Still, it is consistent with the hope that scoring can at least sometimes contribute to social-performance management in ag value chains.

The stated purpose of the draft analysis that prompted this paper was to test whether poverty scoring adds value to a package of indicators of the well-being for participants in ag value chains. The re-analysis here shows that only some of the other non-scoring indicators in this package work. In particular, good indicators are:

- Related with poverty in an intuitive way. Some measures of farming inputs and outputs turn out to have complex, non-intuitive—albeit real—relationships with poverty that make them less useful. Examples include:
  - Area of land farmed
  - Livestock ownership
- Reliably measured, which usually means "simple to ask and answer". Examples are:
  - Electrification
  - Television ownership
  - Source of water
  - Whether ever/never hungry in the past year
- Indicators of outputs (such as production or income), rather than indicators of inputs (such as land or cows)
- In per-capita units (when divisible), to adjust for differences in household "needs":
  - Per-capita net household income (from a single net-income question)
  - Per-capita crop production

This list may not apply beyond Kenyan tea farmers. Nevertheless, these indicators—along with the poverty scorecard—do work for Kenyan tea farmers, and this is a vote in favor of their being among the indicators considered in other contexts.

References

Cowger, Charles D. (1984) "Statistical Significance Tests: Scientific Ritualism or Scientific Method?", *Social Service Review*, Vol. 58, No. 3, pp. 358–372.

Demombynes, Gabriel; Elbers, Chris; and Peter Lanjouw. (2008) "How Good a Map? Putting Small-Area Estimation to the Test", *Rivista Internazionale di Scienze Sociali*, Vol. 116, No. 4, pp. 465–494.

Diamond, Alexis; Rebolledo Dellepiane, Miguel Angel; Skoufias, Emmanuel; Vinha, Katja; Xu, Yiqing; and Nobuo Yoshida. (2014) "An Evaluation of the Simple Poverty Scorecard™ for Estimating Poverty Rates", International Finance Corporation: Washington, D.C.

_____; Gill, Michael; Rebolledo Dellepiane, Miguel Angel; Skoufias, Emmanuel; Vinha, Katja; and Yiqing Xu. (2015) "Estimating Poverty Rates in Target Populations: An Assessment of the Simple Poverty Scorecard™ and Alternative Approaches", World Bank, `dl.dropboxusercontent.com/u/2042671/SPS_Evaluation_Paper_FINAL.pdf`, retrieved 16 December 2015.

González Flores, Mario. (2014) "Evaluation of the Progress Out of Poverty Index®: The Case of Peru", chapter in *Three Essays on Evaluation and Measurement in Developing Countries*, PhD dissertation, American University.

Elbers, Chris; Lanjouw, Peter; and Phillippe George Leite. (2008) "Brazil within Brazil: Testing the Poverty-Map Methodology in Minas Gerais", World Bank Policy Research Working Paper No. 4513, Washington, D.C, `papers.ssrn.com/sol3/papers.cfm?abstract_id=1092691`, retrieved 16 December 2015.

International Finance Corporation. (2014) "Using Market Insights to Improve Local Supply Chains: The Case of Cocoa Smallholders in Côte d'Ivoire", Inclusive Business Insights and Analytics Case No. 3, `ifc.org/wps/wcm/connect/0aac6c0044900aab84728cc66d9c728b/MarketInsights_Case+Study+3_May2015.pdf?MOD=AJPERES`, retrieved 17 December 2015.

McCloskey, Donald N. (1985) "The Loss Function Has Been Mislaid: The Rhetoric of Significance Tests", *American Economic Review*, Vol. 75, No. 2, pp. 201–205.

Schreiner, Mark. (2015) "The Process of Poverty-Scoring Analysis", `microfinance.com/English/Papers/Scoring_Poverty_Outline_Analysis.pdf`, retrieved 17 December 2015.

_____. (2011) "Simple Poverty Scorecard™: Kenya", simplepovertyscorecard.com/KEN_2005_ENG.pdf, retrieved 16 December 2015.

Shipman, Emily. (2013) "Initial Review of Kenya Tea PPI data", 26 November, electronic mail to Mark Schreiner and Don Seville.

Sustainable Food Lab. (2014) "Using the Progress Out of Poverty Index® in Agricultural Value Chains: A Case Study in Kenyan Tea", sustainablefoodlab.org/images/PDF/performance_measurement/Sustainable_Food_Lab-PPI_in_Ag_Value_Chains_Study-Kenya_Tea_1.pdf, retrieved 17 December 2015.

Tarozzi, Alesandro. (2008) "Can Census Data Alone Signal Heterogeneity in the Estimation of Poverty Maps?", *Journal of Development Economics*, Vol. 95, No. 2, pp. 170–185.

_____; and Angus Deaton. (2007) "Using Census and Survey Data to Estimate Poverty and Inequality for Small Areas", *Review of Economics and Statistics*, Vol. 91, No. 4, pp. 773–792.

Table 1: Simple Poverty Scorecard™, Kenya

			Name	Identifier
Interview ID:				
Interview date:		Participant:		
Country:	KEN	Field agent:		
Scorecard:	001	Service point:		
Sampling wgt.:		Number of household members:		

Indicator	Value	Points	Score
1. How many household members are aged 25 or younger?	A. Three or more	0	
	B. None, one, or two	8	
2. How many household members aged 6 to 17 are currently attending school?	A. Not all	0	
	B. All	8	
	C. No children aged 6 to 17	21	
3. What is the material of the walls of the house?	A. Mud/cow dung, grass/ sticks/makuti, or no data	0	
	B. Other	5	
4. What kind of toilet facility does your household use?	A. Other	0	
	B. Flush to sewer, flush to septic tank, pan/bucket, covered pit latrine, or ventilation improved pit latrine	2	
5. Does the household own a TV?	A. No	0	
	B. Yes	16	
6. Does the household own a couch or sofa?	A. No	0	
	B. Yes	14	
7. Does the household own a gas or electric stove?	A. No	0	
	B. Yes	12	
8. Does the household own a radio?	A. No	0	
	B. Yes	8	
9. Does the household own a bicycle?	A. No	0	
	B. Yes	5	
10. How many head of cattle are owned by the household currently?	A. None or unknown	0	
	B. One or more	9	
		Score:	

Table 2: Draft analysis of correlations between scores and outcomes for Kenyan tea farmers

Outcome	Correlation Coefficient (Pearson)	Relationship
Hectares of land farmed	+0.01	None or negligible
Net household income	+0.19	None or negligible
Net income from tea	+0.05	None or negligible
Gross income from other crops	–0.09	None or negligible
Net income from other crops	–0.05	None or negligible
Gross income from off-farm sources	+0.19	None or negligible
Gross income from tea	–0.09	None or negligible
Number of cows owned	–0.28	Weak negative
Number of months of food insecurity	–0.25	Weak negative
Access to electricity	–0.27	Weak negative
Quality of water source	+0.54	Strong positive
Number of communication devices	–0.15	None or negligible
Amount of tea sold	–0.04	None or negligible