The Living Income
Community of Practice

IMPACT INSTITUTE

Estimating farmer household income

How to use secondary data to estimate farmer household income illustrated by two specific use case scenarios

V1.2

Prepared for the Living Income Community of Practice

Authored by Impact Institute

November 2020

Supported by the

Federal Ministry for Economic Cooperation and Development

Implemented by

giz Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH

iseal alliance

Sustainable Food Lab

Acknowledgements

Principal Investigators:

Andrea Rusman | *Senior Associate, Impact Institute*

Anne Mesguich | *Fellow, Impact Institute*

Commisioned by:

ISEAL Alliance

Funded by:

GIZ

Technical Advisory Committee:

We are grateful to the Living Income Technical Advisory Committee (TAC) that both continuously reviewed and supported the creation of this document and the ISEAL team, in particular Adam Romo, Sheila Senathirajah and Kristin Komives. The TAC is composed of representatives from GIZ (Dr. Tim Loos, Dr. Eberhard Krain), COSA (Carlos de los Rios), KIT (Marcelo Tyszler), Impact Institute (Andrea Rusman), Heartwood LLC (Jessica Grillo), University of Wageningen (Yuca Waarts), University of Malawi (Dr. Levison Chiwaula), IDH (Honore Johnson) and Sustainable Food Lab (Kealy Sloan).

Contents

Executive summary

Estimating farmer income traditionally requires extensive primary data collection. However, collecting the primary data required can be challenging. The goal of this guidance piece is to present a framework that can help a wide variety of users in estimating farmer household income and the gap to a living income when the availability of primary data is limited.

Using secondary data to estimate income will likely not yield the same precision, accuracy or even scope when compared to calculation conducted solely using primary data. However, this document provides guidance on how to best use secondary data (e.g. from sector studies, academic studies, or international databases) in order to estimate farmer income. Additionally, data extrapolation is possible and makes for relatively quickly readily available figures that can inform strategy definition.

To address the different goals and starting points of users, this guidance is tailored to select use cases for measuring the living income gap, chosen in consultation with key stakeholders. The focus of this document is on two separate use cases, (1) a one-time estimate for farmer income and the gap, and (2) monitoring progress towards closing the gap.

This guidance starts by introducing a framework design, defining the scope of the income estimation (e.g. country, region, crop, year of analysis) and the key variables, such as yield and prices, that are required in order to assess farmer income.

After the design, the sourcing of data is considered. Insights into designing a legible database and types of data sources are provided. In addition, the concept of a data hierarchy is introduced. By making use of a data hierarchy, the right data source for the right data point can be chosen.

Once secondary data has been sourced, the calculation framework, which guides the user from key variables towards the total farmer income, can be filled and results can be produced and interpreted. Finally, limitations of the framework and ways forward are discussed.

The result is an estimated farmer income figure, based on secondary data, that can be compared to a living income benchmark.

Glossary

Cost of production	All costs associated with the production of a crop. This includes variable and labour costs.
Farm area	A measurement of the size of land area used for agricultural production
Fixed costs	The relatively stable costs involved in crop production. This includes things like maintenance of equipment and land taxes.
Household size	The (average) size of the household producing the crop in question. This includes adults and children.
Labour costs	The costs of labour required to produce a crop.
Labour intensity	The amount of labour required to produce one unit of the crop in question.
Living income	The net annual income required for a household in a particular place to afford a decent standard of living for all members of that household.
Living income gap	The difference between the actual income of a farmer household and the living income for a household in that same region.
Farm income net of variable costs	The income from all economic activities other than the production of primary agricultural commodities. This includes self-employment activities and wage employment activities.
Framework	The structure for the development of an analysis, including the goal, scope and calculations of the analysis.
Net off-farm labour income	The income from all economic activities other than the production of primary agricultural commodities. This includes self-employment activities and wage employment activities.
Other income (non-farm, non-labour)	Non-farm non-labour income mainly comes from public and private transfers, gifts, and remittances from non-household sources, as well as income from land rentals and from sharecropping.
Price	The market price for a given crop. Price here should be the price as received by the farmer in the context at hand.
Primary data	Data collected from the subject group by the researchers involved in the project.
Secondary data	Data collected from secondary sources. This includes data collected from the subject group by a third-party that is not involved in the current project.
Total net household income	The income of the household from all sources (e.g. from commodities and off-farm income) after deducting all costs of the household. These include fixed and variable costs from all crops.
Variable costs	The costs of crop production prone to vary from year to year. This includes things like the costs of seed.
Yield	A measurement of the amount of agricultural production harvested per unit of land area

1 Introduction

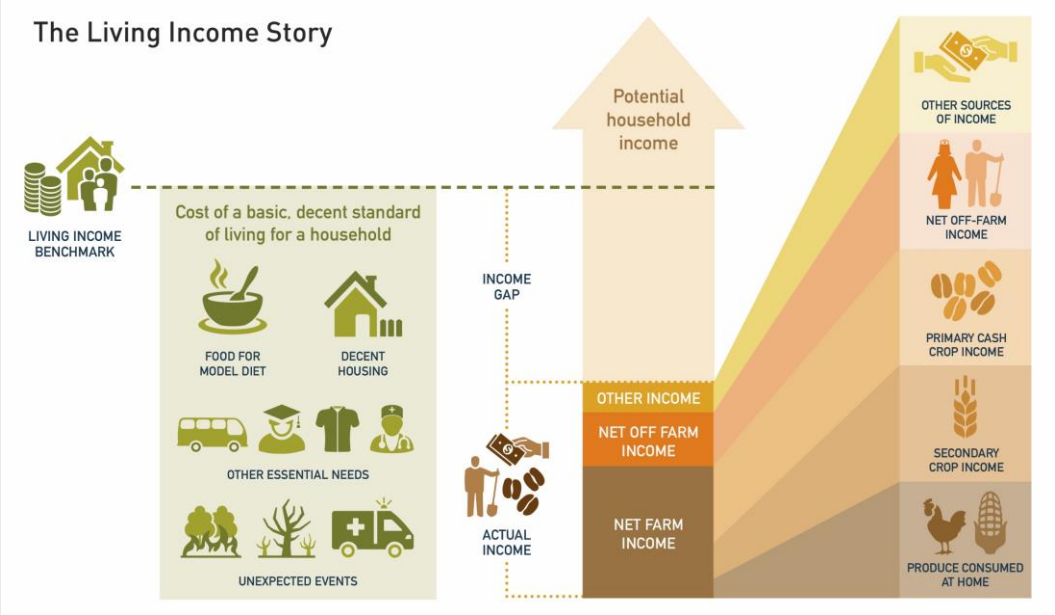
Estimating farmer income traditionally requires primary data collection. However, collecting the primary data required can be challenging. Farmers may not always have the necessary information prepared, and it can be time-intensive to collect and scale for the organisations seeking to understand the living income gap and integrate it into their strategies. For this reason, using secondary data to fill primary data gaps in estimating farmer household income can prove useful.

In this document guidance is provided on sourcing secondary data for the estimation of farmer household income to fill primary data gaps or substitute primary data where it is absent. In addition, recommendations are provided on how to best use this secondary data to come to a farmer household income figure. This guidance assumes that a Living Income benchmark is available for the context and geography that an organisation is operating in or interested in.

Box 1. Farmer income and living income concepts

This document does presuppose some background knowledge of farmer and living income concepts. The Living Income Community of Practice has developed a variety of resources that provide further guidance and justification for the advice given in this document. The most notable for reference are:

- Guidance on calculating household income provides information on how to measure and calculate actual incomes for smallholder households in agricultural contexts. This document is important background information to understand the various components of farmer income and how to relate the components between one another.
- Guidance manual on calculating and visualizing income gap to a living income benchmark provides information on how to calculate and interpret key indicators of the gap between actual income and a Living Income Benchmark and on how to visualize key indicators of the gap. This document is useful in the final steps of this guidance, where the estimated farmer household income figure can be compared to a living income figure.



This document is appropriate for a wide variety of users, ranging from users with little to no familiarity with living income concepts to users with high levels of familiarity. See Box 1 for further information on farmer and living income concepts. Example users include research specialists, analysts and M&E practitioners from NGOs, standards organisations, governments, and commercial organisations.

There are a variety of reasons why stakeholders will be interested in estimating farmer household income and the income gap. They will also approach income estimation from different starting points. Different needs and different starting points will have implications for how to source secondary data and the resulting precision of the estimate. It is thus important for the user to clearly define the desired level of precision, as a more precise estimate will often result in a higher cost and time investment, while a lower precision estimate can be performed quicker and at lower cost. For this reason, this document focuses on the estimation of farmer income illustrated by two specific use case scenarios.

The two use cases were selected in consultation with key stakeholders ranging from programme managers to managing directors working at development agencies, NGOs and corporations. An overview of the full list of use cases identified and considered in this consultation can be found in Table 1. The use cases selected for this document are: 1. 'A one-time estimate' and 2. 'Monitoring progress towards closing the gap'.

A one-time estimate of farmer income contains the assessment of farmer income over a set period of time, typically one year. The one-time estimate is the assessment of farmer income and living income gap for one region and one primary crop. The goal is to understand the magnitude of a problem in the chosen region and crop. An assessment of a one-time estimate of farmer income can be useful in shaping possible necessary interventions and development strategies.

Monitoring the progress towards closing the living income gap contains the assessment of farmer income and living income gap estimates, and their evolution over time. The purpose of this use case is therefore a comparison of living income gap estimates over multiple years. Possibly, drivers of improvement in the evolution of the living income gap can be used in determining next steps for that specific region and crop.

Table 1. Overview of use cases proposed in stakeholder consultation

Use case	Description
1. A one-time estimate	**A one-time estimate allows for the assessment of living income, farmer income and living income gap for one region and one crop. This allows for understanding the magnitude of the problem in a given region for a given crop and shapes possible interventions and development strategies.**
2. Living income gap hotspot analysis	A living income gap hotspot analysis is used to get estimates for different commodities and regions in order to compare them. This allows to assess in which region, and for which crop the biggest gaps lie and thereby prioritize decision-making for which regions and crops to focus on.
3. Progress towards closing the gap	**This situation is for the assessment of living income gap estimates and their evolution over time. The goal of this use case is to monitor progress for a given region and crop, and thereby shape possible interventions and next steps for that specific region and crop.**
4. Progress of specific farmers	This use case is for the assessment of living income gap estimates of specific farmers and their evolution. The goal of this use case is to assess progress for a given region and crop in a specific value chain.

5. Effect of programs	This use case concerns the assessment of living income gap estimates before and after a specific intervention. Interventions can be financial (e.g. price increases) and non-financial (e.g. training provision on yield improvement).
6. Profit/production costs	The assessment of the relationship between profit and production costs enables the assessment of measures to be taken in order to improve farmer productivity. The goal of this use case is to identify productivity improvement measures.
7. Reference price estimate	An estimate of what price farmers need to get for a decent livelihood in a specific region for a specific crop, farm size and productivity level. The goal of this use case is to inform price discussions.

Since the use cases differ in purpose there will be differences in the estimation framework that each requires; for instance, on which variables will be more relevant for the user. When estimating a one-time farmer income figure, the focus will be on getting the correct granularity level for the variables with the most influence on the final farmer income figure. Granularity is further defined in Box 2. To assess progress towards closing the gap, the emphasis will be on year-dependent variables, such as yield and prices.

Box 2 Granularity level

Granularity can refer to different data characteristics. It can either refer to how a larger variable is divided into multiple data points, or to how close to the scope a certain variable is. For the former, it concerns the aggregation level of the data used in a calculation: for instance, is the cost of each fertilizer known or is only the total input costs available? On the latter, granularity is about whether the data available fits the scope of the project, such as a regional versus national data.

It is important to emphasise that using secondary data to estimate farmer household income comes with several potential limitations. Using secondary data produces relatively quick estimates, which can be used in high level discussions on closing the living income gap. Secondary data can also be a good supplement to primary data where it is lacking. However, a significant drawback of using secondary data is that, in some cases, it is difficult to obtain a holistic understanding of household dynamics. Both use cases have a strong focus on the primary crop of the farmer, and therefore do not allow for the assessment of key drivers of a number of strategies, such as diversification. Hence the user's strategy is key in defining the goals of the farmer income estimation and whether secondary data serves that purpose.

The document is outlined as follows:

1. the design of the framework for estimating farmer household income is presented (Chapter 2).
2. secondary data is considered, including a suggested structure for a database that compiles secondary data, an overview of secondary data sources and the construction of a data hierarchy method (Chapter 3).
3. a calculation framework is developed for the use cases at hand, outlining the steps that should be followed. (Chapter 4).

Figure 1. Overview of the farmer household income estimation process using secondary data

To give applicable guidance on how to approach the different scenarios, each chapter contains a general section and a deep-dive section for each of the use cases considered. Producing results is in the scope of this document, while interpretation of results is addressed in a different guidance document. [1]

[1] See Guidance manual on calculating and visualizing income gap to a living income for further information.

2 Framework design

In this chapter, the first steps for estimating farmer household income using secondary data are outlined. It starts out by describing the different data needs of the user. To reiterate, this guidance is appropriate for users with starting points ranging from instances where primary data is partially available, to instances where primary data is not available or possible to collect. For instances where partial primary data is available, the framework described can be used to fill gaps with secondary data where necessary. Where primary data is not available at all, the full framework can be applied. Further description of primary and secondary data is provided in Box 3. Then, based on the specific use cases, scope and key variables for estimation are defined.

2.1 Developing a framework based on secondary data

Before defining key variables, it is essential to define the scope of the farmer income assessment. This includes country, region, crop, year of analysis, whether the farmer income assessment relates to certified or non-certified farmers, and so on. Then, based on the scoping, we define which variables are likely to be found directly in secondary sources, which may need to be extrapolated from other data points, and which may not be available. The assumption is that a living income benchmark is available for the context being considered.

Approaches to estimated income can range from a very granular estimate, where each data point is known and validated, to lower granularity, where some input data points are derived/estimated. Required levels of granularity will depend on the scope and goal of the project.

The most granular approach will entail collecting secondary data on all farmer income variables, ranging from yield to remittances. This is expected to be a tedious exercise, because of the lack of readily available data. The data will likely not be from the exact year that is needed, or there is a chance that it will not fit the exact scope of the project given the limited number of household surveys and sector reports that are generally conducted. While this approach will likely give the most detailed and precise results, it quickly becomes time-intensive for the user.

On the other end of the spectrum, the least granular approach entails collecting secondary data for some key variables and extrapolating others. What the key variables will be will depend on the use case at hand. If the use case involves historical analysis, then variables that are highly dependent on year will be the key variables. If the application is focused on a specific aspect of farmer income, e.g. production costs, the key variables will relate to that specific aspect of farmer income.

One issue in estimating average farmer household income is that the underlying distribution of farmer income is not known. Box 4 provides more information on this. Results might show that average farmer income is above the living income, but it is not clear whether all farmers earn above the living income or if this only driven by a few farmer households. Making use of ranges instead of using only one figure for average farmer income can be helpful. Additionally, statistical distributions and modelling techniques can provide guidance on visualising the underlying income distribution and make clear what the distribution of farmer income compared to living income entails.

Box 3. Primary vs. secondary data

Primary and secondary data have been defined for the purpose of this guidance document as follows:

- Primary data Data collected from the subject group by the researchers involved in the project.
- Secondary data Data collected from secondary sources. This includes data collected from the subject group by a third-party that is not involved in the current project.

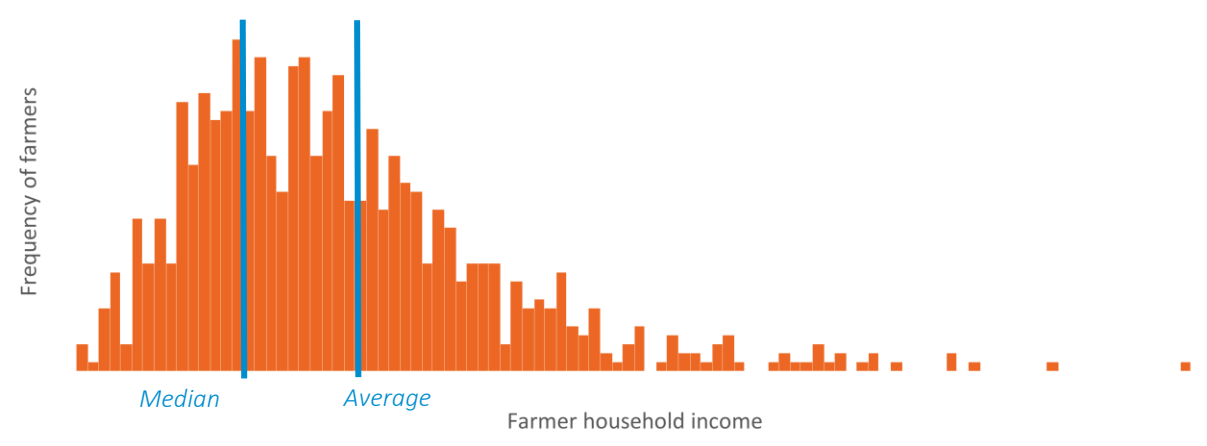
This guidance document focuses on secondary data to fill gaps in primary data where it is lacking. Secondary data can refer to all sorts of data available from secondary sources. This includes primary data that has been collected from third parties and has been made available to the public. This data can be specific to the subject group in scope, or to a subject group close to the one in scope. Where available, this type of data can be very valuable when estimating farmer household income using secondary data. It is recommended to use this data where available, after assessing its validity and applicability to the use case at hand.

Box 4. Average farmer income

Any estimation provided in this document refers to the estimation of average farmer income. This implies that the goal is to look for the average household, and asses what the farmer in that household would earn. However, average farmer income might not be representative of the variability between all farmer incomes in scope of the analysis.

Income tends to follow a lognormal distribution, because a large number of farmers earn below the average income, and a small number of farmers earn (much) more than the average income. In Figure 3, an example of a lognormal distribution graph is shown. There are instances where average farmer income will be above the living income, leading users to think that the living income gap is closed. This is a misrepresentation of reality, as most farmers will likely still earn below the living income. The median income might be a more representative figure of farmer income since it represents the central location of the data. Therefore, a better representation of the reality would be the use of median farmer income. However, data available is largely average data, e.g. average yield, average farm size, etc. Assessing the median farmer income or the farmer income distribution is possible using econometric modelling techniques.

Figure 2. Example lognormal distribution



2.2 Use Case 1 - One-time estimate of farmer income

In this section, the framework design of use case 1 – a one time estimate of farmer income and the income gap, is described.

Relevant variables for this use case that are likely to be found in, or can be extrapolated from, secondary sources are listed below. Some variables are more complex than others to find, further explanation is provided in Box 5. Orange variables are likely to be found directly in secondary sources and black variables can be extrapolated from others.

- Farm area (main crop)
- Yield (main crop)
- Price (main crop)
- Variable costs (main crop)
  - Input costs
  - Labour costs
  - Transportation costs
- Farm income net of variable costs (main crop)
- Fixed costs (total)
- Net off-farm labour income
- Other income (non-farm, non-labour)
- Household size

The variables that are likely to be found directly in secondary sources are farm area, yields, prices and (some) costs for the main crop of the farmer. These are highly likely to be reported on in various sources, such as international databases, sector studies and independent studies. Some of these variables, like yields and prices, can differ quite drastically from year to year. It is important, when researching them, to use sources as close to the defined scope as possible, especially the year of analysis.

However, if data on variables subject to change is entirely unavailable for the year of analysis, it may be possible to identify trends of change in the years prior to the analysis year. Identified trends can be used to scale the last known value up to the analysis year, offering a best estimate of the variable. In contrast, variables that are more or less stable over the years, like fixed costs and farm area, do not necessarily need scaling. Values further from the year of analysis can be used directly in this case unless there is evidence of major change.

Other variables may need to be extrapolated if they cannot be found directly in secondary sources, like farm income net of variable costs (from the main crop). The farm income net of variable costs may be derived from other, more readily available variables, like yield and price, if the identified yield and price of the main crop are robust.

Box 5. Production costs and labour requirements

Using secondary data to estimate farmer income comes with several drawbacks. A particularly salient drawback is that it is often difficult to find input figures, such as costs of production, while output figures, such as production quantity sold, are more easily accessible. An overview of the different possible granularity levels is provided here, along with a sense of what can and cannot be achieved with secondary data.

There are different levels of granularity possible, depending on what data is available. Some sources will report total production costs, other sources will have more granular data, e.g. the cost of fertilizer, the cost of hired labour, the cost of transportation.

An even more granular approach would be to, in the case of fertilizer, have the amount of fertilizer used, the type of fertilizer used, and the cost of each type of fertilizer. Similarly, for a granular approach to labour costs, one would need to know the labour availability in the household, the labour requirements of the crop, and the cost of hired labour. Labour availability in the household is crucial because smallholder farmers tend to rely heavily on labour from family members living in the household. Finally, for transportation costs, one would need to know the cost of fuel, the type of vehicle used and the kilometers to be travelled to get to a market where the crops can be sold.

It is recognized that these very granular approaches can often not be achieved using secondary data. While they would yield more precise results, this guidance document takes a pragmatic approach to what secondary data is available and provides suggestions on where to be most granular and where it would be fine to be less granular.

2.3 Use Case 2 - Monitor progress towards closing the gap

The use case 'Monitor progress towards closing the gap' implies that the user is looking to assess farmer income in multiple, consecutive years. It differs from the 'One-time estimate' in that the output of interest is not a single farmer household income figure, but the evolution of that single figure.

Given the application at hand, some variables need to be collected with higher accuracy and highest match with the scope of the farmer income assessment project. Other farmer income variables can then be approximated using modelling techniques, which can range from quite complex to relatively straightforward. Complex modelling techniques, such as the use of statistical distributions of variables, will yield more precise results, but will require some background knowledge and effort. Straightforward modelling techniques, such as the use of ratios, will yield more readily available results.

Data points to be collected for this second use case are listed below. Orange data points are the non-negotiable data points that need to be collected. The others can be extrapolated.

- Farm area (main crop)
- Yield (main crop)
- Price (main crop)
- Variables costs (main crop)
- Farm income net of variable costs (main crop)
- Fixed costs (total)
- Net off-farm labour income
- Other income (non-farm, non-labour)

The variables that must be collected for all years in the assessment are yields and prices for the main crop of the farmer. This is because yield and prices are highly dependent on the year that they are collected in. One of the most important factors to influence crop yield is climatic conditions, such as annual precipitation, humidity, temperature, and weather patterns. These are likely to change over the years. Further, market forces, such as supply and demand, drive prices up and down, forcing prices to change on a yearly basis. Therefore, secondary data on yields and prices will have to be collected with a high level of granularity, for all years that the evolution of farmer income is being analysed.

The other variables can be extrapolated. Amongst these variables, some are likely to change over time, but will depend on yield and/or prices. These include, for instance, variable costs or how much of the farm income net of variable costs comes from other crops, livestock, and livestock byproducts. Variable costs, as the name indicates, are variable to the amount produced. Therefore, it is possible to extrapolate a ratio of variable costs to production value over the years.

On the contrary, some variables will follow an inverse relationship with production value for the main crop. If the farmer produces more of the main crop, all else equal, there is less capacity for other crops and livestock as well as less time for wage employment.

Other variables can be assumed to vary less throughout the years. These variables include, for example, fixed costs and remittances from non-household family members. These values can be collected and used directly, where the only required adjustment is an inflation adjustment to correct for currency movements (reliable inflation figures can be found at World Bank data).

In the next chapter, we focus on collecting these variables from secondary sources. In the chapter 'Modelling, calculations, and results', we then define secondary data frameworks for estimating incomes and the gap for both use cases described in this document

3 Secondary data

Secondary data is useful when primary data collection is not feasible, or when only partial primary data is available and secondary data is used to fill gaps. Using secondary data overcomes the drawbacks of primary data collection, which is often time-intensive to collect and scale. However, secondary data sourcing can become overwhelming when there are multiple sources that can answer the same question. Finding the right data sources that are appropriate for the specific scope, use case and contextual granularity is also challenging. It is therefore important to consider sources of secondary data in a structured manner, and to classify data according to a data hierarchy.

In this chapter, general background on structuring secondary data, an overview of secondary data sources and a data hierarchy method are presented. They are then applied to the specific use-cases of this guidance.

3.1 Structuring data collection

The suggested structure for secondary data sourcing for estimating actual income is outlined in Table 2. An example on how to fill this table with actual data is provided in Appendix A. Example of data characteristics to be compiled in a database: Yield.

Table 2. Data characteristics to be compiled during data collection

Variable	Description
Variable name	Name of the variable as found in the original source
Variable category	Actual income component the variable refers to
Region	Region to which the data point refers to; can be a region, a country, a district…
Crop/sector	Crop/sector to which the data point refers to
Year	Year to which the data point refers to
Unit	Measurement unit in which the data point is expressed
Value	Raw data point as found in the original source
Comments	Use this column for any reasoning and information that is not in other columns, e.g. calculation or measurement approach, scope of the original study, limitations of the study
Citation	Direct copy of the sentence in the report, page number and/or table number
Source name	Author and title of source
Source type	Type of source, e.g. peer-reviewed study, international database, national statistics
Source link	Link to web page where the source can be found

Making a structured database and collecting more data than strictly necessary is important because it will allow the user to choose a calculation framework based on the available data and the most relevant data point based on a data hierarchy method.

3.2 Overview of secondary data sources

Sourcing secondary data for estimating farmer income is challenging, because not all data points come from the same data source. Furthermore, some source types are more relevant than others, depending on which variable is being investigated. Different source types include:

- International databases Database compiling large amounts of data for different countries and regions and commodities. Examples include FAOstat, ILOstat and World Bank.[2]
- Sector reports Reports conducted by research institutions on a specific matter in a given sector. Examples include the Cocoa Barometer.
- National statistics Statistics compiled by national governments. Examples include data published by the Uganda Bureau of Statistics or the Institut National de la Statistique in Ivory Coast.
- Household surveys Statistics compiled by national governments and research institutes and the underlying data of sector reports on household characteristics. Household survey examples include the Enquête sur le Niveau de Vie des Ménages in Ivory Coast or the Ghana Living Standards Survey.
- Academic studies Research conducted by academia on specific issues in a given sector, as well as the underlying data, generally published in renowned journals. Examples journals include the Journal of Agricultural & Applied Economics or the Journal of Rural Studies.
- Company data Statistics compiled at company level. A number of companies collect data in-house on the farmer income variables of their key producers.
- Certification schemes Statistics compiled by certification schemes on their key producers. A number of certification organizations collect data on a regular basis on the farmer income variables of their key producers. Example certification schemes include Fairtrade International or the Rainforest Alliance.
- Other studies Any source that does not fall in the categories defined above. This includes, amongst others, the data collected by first buyers and program implementers.

It is important to assess the relevance of these data sources considering the data hierarchy outlined in the next section and to bear in mind the quality and reliability of each data source. Validated studies have preference over non-validated, single studies. The source should have a similar scope to the project at hand.

Source types have been ranked on how relevant they are for each farmer income variable in Table 3. The ranking has been determined based on how applicable each source is to each variable. For this guidance piece, a secondary data search was conducted, and a corresponding database was constructed. Based on the outcome of the database, source types were ranked based on how often a variable was found in each source type. This is complemented by a ranking of which variables would most benefit from being primary data, where high (H) are those that would most benefit from being collected directly and low (L) would benefit the least.

[2] Data from international databases should be used with caution. These databases often have different calculation rules than farmer-level studies. For instance, FAOstat will report yield as total production divided by total area of land cultivated, whereas farmer-level study would report yield as average farmer production divided by average farmer land cultivated. These will provide different results and depending on the purpose of the study, the user might have a preference for one over the other.

Table 3. Source types ranked per relevance for each variable (1 - most relevant; 5 - least relevant)

Variable	International database	Sector reports	National statistics	Household surveys	Academic studies	Company data	Certification schemes	Primary
Average wage in the region	2	4	1	7	6	3	5	L
Farm area	4	2	5	1	6	7	3	M
Farm income net of variable costs	6	2	5	1	3	7	4	M
Fixed costs	6	2	5	1	3	7	4	H
Labour available in the household	6	2	5	1	3	7	4	M
Labour intensity rate	2	4	1	5	3	6	7	L
Net off-farm labour income	6	2	5	1	3	7	4	H
Non-farm non-labour income	6	2	5	1	3	7	4	H
Total costs of production	5	1	4	2	6	7	3	H
Total value of production	6	1	4	2	5	7	3	M
Price	3	2	1	6	7	5	4	L
Yield	1	2	5	3	6	7	4	H

Generally, sources where the most granular farmer income data is available are sector reports and household surveys, followed by academic studies. Sector reports will have very granular data available, but very often this data will be applicable to a very specific scope, which means it will not always be possible to extrapolate. Household surveys will also have granular data available, but not all farmer household income elements will be covered. Academic studies will likely have granular data on some elements. International databases have the advantage of having lots of data available over multiple years but might relate to other producer types than smallholder producers. Performing validation of the data can prove key in understanding the data collected and farmer income results. More information is provided in Box 6.

Even though source types are more or less relevant depending on the farmer income variable at hand, each type of source comes with certain benefits and limitations. These are outlined in Table 4.

Table 4. Source types along with their benefits and limitations

Source type	Benefits	Limitations
International database	The available data will be validated and often repeated over the years.	The data points reported on are broad and do not often provide an overall picture. They likely will not be directly applicable to more specific contexts. Often, it is also unclear where the data originated from or if it was validated.
Sector reports	The available data will be validated and has a high chance of being reported over multiple years. It is also more applicable to studies related to the sector.	The data may not be very applicable to contexts other than the sector being studied. The data being reported on is also not guaranteed to give an overall view of farmers' income and lives.
National statistics	The available data will be validated and often repeated over the years.	The data points reported on are broad and do not often provide an overall picture. They also likely will not be directly applicable to more specific contexts.
Household surveys	The data will be quite granular and fairly accurate, especially on data points that are not reported in other statistics or are very variable.	Household surveys are expensive to commission and often difficult to find, especially with more specific scopes.
Academic studies	The data will have a very specific scope and often provide a good overall picture of the subject. They are also likely to be peer-reviewed or otherwise validated.	The study is not guaranteed to be repeated over the years and the data is also unlikely to apply to subjects outside of the scope of the study.

Every source type has specific benefits and limitations, but, in general, similar types share similar characteristics. International databases, sector reports and national statistics will be validated and likely repeated, but often will provide very general information that may not necessarily fully apply. Conversely, household surveys, academic studies and other studies will likely offer very granular and applicable data (provided the subject covers the same or a similar scope), but household surveys and other studies are not guaranteed to be repeated or validated (peer-reviewed or otherwise confirmed for their accuracy) and it can be difficult to locate or commission these studies.

Box 6. Qualitative validation of secondary data

Where possible, it can be very valuable to perform qualitative validation. Some degree of qualitative local data can be helpful in this.

Interviews can be conducted with key stakeholders early in the project, which may help to understand better the study area and the focus of the living income assessment. At the outset of the project, it can be useful to, for instance, understand the key drivers of farmer income: is farmer income driven largely by the primary crop? How important is hired labour? Do farmers rely on off-farm income? This can prove helpful in refining the goal of the user and prioritize focus in data collection.

Once data has been collected, it can also be helpful to validate the key drivers of farmer income and perform sensitivity analyses. Again, interviews with key stakeholders who have a good understanding of the primary crop and the local area's situation can prove useful. Validation can be performed on farmer income variables and data collected that appears to be a key driver of these farmer income variables.

3.3 Data hierarchy method

A data hierarchy is a useful tool for understanding the data needs for the framework of a specific estimation use case. Box 7 provides some insights in the definition and uses of a data hierarchy. There are two levels to the proposed data hierarchy, as shown in Figure 3:

- Validated official statistics and sources vs. Non-validated single studies
- Same scope vs. Different scope

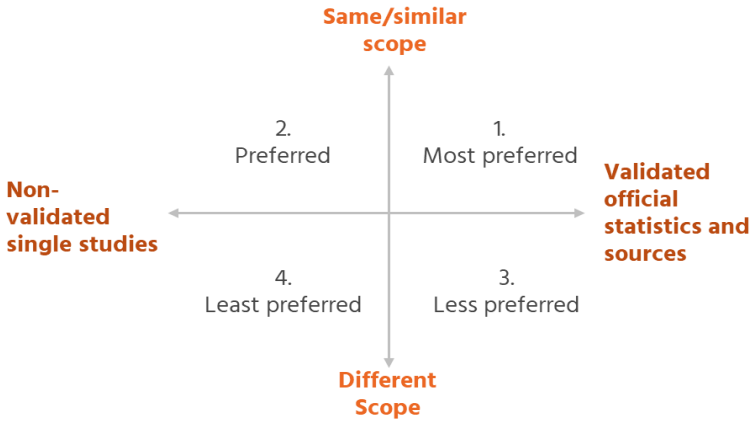


Figure 3. Preferred data sources

Validated studies or official statistics have preference over non-validated single studies. This is because results will be more robust and therefore more appropriate to be included in the farmer income assessment. It is important to acknowledge that there can be a trade-off between how robust some results are vs. how applicable they are in a certain setting.

Box 7. What is a data hierarchy?

When sourcing secondary data, it can happen that a user needs to choose among different sources for one data point. In general, the goal is to use sources with a similar scope to the research being conducted. It can happen that, for a given data point, there is no available data with a similar scope to the research being conducted.

This is when a data hierarchy becomes useful. A data hierarchy provides the user with a ranking of data characteristics in order to support them in deciding which data point is best suited for the variable at hand. The hierarchy ranks data characteristics from most to least relevant based on how dependent on this data characteristic a data point is.

Then there is the question of scope. Scope can relate to different elements, including year, country or region and focus crop. Depending on the framework needed for a specific use case, some elements will be more relevant than others. For some variables, year will be more or less important. Yield for instance is highly dependent on weather and market conditions. These conditions highly vary across years and cause yield to be significantly different in different years. Conversely farm area is far less likely to be different every year.

Finally, if the preferred data point from the data hierarchy is not available, the user should always use the most conservative estimate. The most conservative estimate is the estimate that will make the farmer income estimate smallest, an thereby the living income gap largest.

The following data characteristics can be included in a data hierarchy:

- Year — Is the secondary data from a different year?
- Country — Is the secondary data about a different country?
- Producer type — Are producers in the source the same? (e.g. plantations vs. smallholder)

- Agricultural sector Is the sector of the study an agricultural sector? (e.g. national average agricultural sector instead of coffee?)
- Certification status Is the data about certified farmers? (e.g. Fairtrade instead of sector)

Table 5 shows for each farmer income variable what type of secondary data characteristic is preferred in terms of data quality. Each variable is ranked on a scale of 1-5 based on what type of secondary data characteristic is preferred in terms of data quality and the application at hand. 1 Indicates most preferred in terms of data quality, while 5 indicates least preferred. The ranking is determined based on how sensitive a variable is to a specific data characteristic and is based on expectations of experts over what sources would provide the most reliable estimate for the reference population. For instance, if average farm area is not available for a different year (3), then the researcher should look for average farm area in a different certification status (2); if farm area is not available for a different certification status, then the research should look for farm area in a different country (1). Entries with n/a imply that the data characteristics is not applicable to the variable at hand.

Table 5. Data hierarchy for farmer income variables (1 – most preferred; 5 – lesat preferred; n/a: not applicable)

Variable	Year	Country	Producer type	Agricultural sector	Certification status
Farm income net of variable costs	4	3	2	1	5
Total value of production	3	2	1	n/a	4
Yield	2	1	3	n/a	4
Farm area	3	1	n/a	n/a	2
Total costs of production	2	1	3	4	5
Net off-farm labour income	2	3	n/a	1	4
Non-farm, non-labour income	2	3	n/a	1	4

3.4 Use Case 1 - One-time estimate of farmer income

The one-time estimate use case will require a set of secondary sources that are comparable across various scoping choices, such as time frame, geography, primary crop, type of agriculture, etc. Once these scoping choices have been made for the one-time farmer income estimate, secondary sources that utilise the same scoping choices can be researched, with the sources sharing the highest amount being the most preferred.

Using secondary data allows for filling gaps at farmer level, the estimation of average farmer income and, in the best-case scenario, the estimation of median farmer income. The requirement for an exercise that will use secondary data to make a one-time estimate of farmer income, is to find available sources of the following variables:

- Farm area (main crop)
- Yield (main crop)
- Price (main crop)
- Variable costs (main crop)
  - Input costs
  - Labour costs
  - Transportation costs
- Farm income net of variable costs (main crop)
- Fixed costs (total)
- Net off-farm labour income
- Other income (non-farm, non-labour)
- Household size

When determining which sources to use, the choice between the comparability of different scoping choices will arise. For example, it may be necessary to choose between a study that is closer to the year of analysis and one that focuses on the same crop and region. It is important, here, to consider comparability in relation to use and extrapolation. The comparability of certain scoping choices will be more relevant for different variables. For extrapolating the off-farm labour income and other income, for example, it may be more relevant to compare sources from similar regions with a different primary crop than it is to compare sources from different regions with the same primary crop. This is because labour and other income are more likely to be influenced by average wage and the availability of work in the region than primary crop. However, if fixed costs are being extrapolated, it may be more relevant to compare sources with the same primary crop, as they are more likely to have similar fixed costs.

Some data points may be largely unavailable in relevant secondary sources. An example of this is percentage of production that is consumed in the home. This is a variable not often reported on, except for select individual studies, and may prove elusive in research. In cases such as these, it may be best to construct a model that uses a predicted percentage from a different region or crop, while acknowledging the limitations and caveats of using this data point, rather than build a model that does not consider these data points at all.

3.5 Use Case 2 - Monitor progress towards closing the gap

Given the application at hand, the requirement for an exercise that will use secondary data to monitor progress towards closing the gap is to have a baseline sector report or household survey with the following variables available:

- Farm area (main crop)
- Yield (main crop)
- Price (main crop)
- Variable costs (main crop)
- Farm income net of variable costs (main crop)
- Fixed costs (total)
- Net off-farm labour income
- Other income (non-farm, non-labour)

Because monitoring progress towards closing the gap implies yearly estimation of farmer income, it is important to be able to collect data on a yearly basis for the key variables defined in 2.3. Thereby the recommendation is to follow the data hierarchy more closely for these key variables: yield and price. For other variables, compromises on scope can be made. The year of the baseline sector report or household survey will be the base year. For the base year, data can be collected directly from the report or survey and used to derive farmer income. In the assessment of farmer income in the other years, we will make use of ratios to extrapolate some key variables described in the next chapter.

4 Modelling, calculations, and results

Once you have a good overview of the potential data sources, the next step is building the calculation framework. This chapter describes approaches for a user to model calculations and lays out how to process results to estimate incomes and the income gap of each of the described use cases.

4.1 Use Case 1 - One-time estimate of farmer income

A summary of the framework for use case 1 is provided in Table 6, from the first variables (1, 2 and 3) to the total farmer income in step 16. The table should be read from the top to the bottom. Calculations are boldened. Above each calculation are the variables needed as input for this calculation, along with their required granularity and assessed availability. Granularity and availability are given a ranking of either high (H), medium (M) or low (L), where a low ranking indicates the variable does not need to be very granular for this use case and it is unlikely to be found in secondary sources respectively. Conversely, a high ranking indicates the use case requires the variable to very granular and the variable is likely to be found in secondary sources. A detailed overview of the approach is provided below.

Table 6. Granularity level and summary framework for a one-time estimate of farmer income

ID	Variable	Calculation	Granularity	Availability	Suggested approach
1	Farm area (main crop)		M	H	Collect directly
2	Yield (main crop)		H	H	Collect directly
3	Price (main crop)		H	H	Collect directly
4	**Production value (main crop)**	**1 * 2 * 3**		**L**	**Farm area (main crop) * Yield (main crop) * Price (main crop)**
5	Variable costs (main crop)		M	M/L	Collect directly (if available)
6	Labour available in household		M	H/M	Collect directly
7	**Required labour**			**L**	**Farm area * Labour intensity * (Yield / Farm area)**
8	**Labour costs**	**(7 – 6) * 13**		**L**	**(Required labour – labour available in household) * Average wage**
9	**Gross profit (main crop)**	**4 - 5**		**L**	**Production value (main crop) – Variable costs (main crop)**
10	% farm income net of variable costs (other sources)		L	M/L	Collect directly (if available)
11	Fixed costs (total)		L	L	Collect directly (if available)
12	**Net farm-income (total)**	**(9 / 10) - 11**		**L**	**Gross profit (main crop) / % farm income net of variable costs (other sources) – Fixed costs (total)**
13	Average wage		M	H	Collect directly
14	Labour participation rate		L	M	Collect directly
15	**Net off-farm labour income**	**(7 – 6) * 13 * 14**	L	**L**	**Net off-farm labour income = (Required labour – labour available in household) * Average wage * Labour participation rate**
16	Other income (non-farm, non-labour)		L	M/L	Collect directly, if available
17	**Total farmer income**	**12 + 14 + 16**		**L**	**Net farm-income (total) + Net off-farm labour income + Other income (non-farm, non-labour)**

The first step for generating a one-time estimate of income is to bring all data to be used to a similar timeframe. Where the data is in monetary units and is a relatively stable variable, like fixed costs, it only needs to be inflated to the analysis year. Where the data is monetary, but is more subject to change, trends need to be identified where they exist. The data will then need to be multiplied by the rate of change necessary to bring it to the analysis year. This same process holds for all other changing, non-monetary variables that have identifiable trends. Lastly, those non-monetary variables that are stable, but outside of the time period, will be used as is, if they are the most applicable.

The total farmer income can be calculated by following the following steps:

1. The first step is the calculation of **gross profit from the main crop**. If gross profit from the main crop is unavailable, it can be calculated through **production value**, **variable costs** and **fixed costs**. Production value is calculated as the multiplication of farm area, yield, and price.

   (4) *Production value = Farm area * Yield * Price*

   The same process holds for secondary crops, if applicable. Ideally, this data is collected from international databases such as FAOstat, but it can also be collected from validated studies, sector studies or household surveys. Farm area, yield and prices can usually easily be collected from secondary data sources, but not always for the scope at hand. For instance, it might be difficult to collect yield specifically for smallholder farmers. Prices can be farm-gate prices, FOB prices or EXW prices depending on the source used. It is important to understand what price the farmers get, which is often dependent on the crop at hand, and collect secondary data accordingly.

2. The next step in calculating gross profit is the calculation of **variable costs**, which typically includes labour costs and production costs, which may be found in secondary literature. However, if they are missing, they can be broken down and extrapolated from different sources.

   (5) *Variable costs = Costs of hired labour + Input costs + Other variable costs*

   If the **labour costs** necessary for the main crop are missing (a key aspect of variable costs), they can be extrapolated from other sources, though this might prove difficult. Labour costs could be extrapolated from the required labour to produce the crop, the available labour in the household and the average wage for the region. **Required labour** can be calculated as follows:

   (7) *Required labour = Farm area * Labour intensity * Yield*

   The required labour then needs to be compared to labour available in the household.[3] If the required labour is equal to the labour available per household, the labour costs will be zero, as the crop will be produced wholly by the household. If the required labour is lower than available household labour, the remaining labour (excluding household management labour, such as childcare) can be multiplied by the average wage and the labour participation rate for the region to provide the potential off-farm labour income (assuming there is no secondary crop). If the required labour is higher than the household labour, conversely, the remaining labour will have to be multiplied by the average wage in the region to provide the labour costs for the crop. See Box 8 for further description of the approach. For comparability, it is best for the source providing labour

[3] This can be estimated per region based on Anker & Anker https://www.elgaronline.com/view/9781786431455/chapter13.xhtml

to have a comparable crop, while the average wage should have, at least, a comparable region. It is also important to note the seasonality of the crop, as this will influence labour intensity throughout the year.

Box 8. Labour requirements and availability in the household

Using secondary data sometimes means simplifying assumptions have to be made. The way labour requirements and therefore labour costs are approached in this use case is as follows:

1. Labour availability. Calculate the number of FTEs in the household per Anker & Anker[3]:

Number of full-time equivalent workers per family = 1 + [LFPR × (1 −U) × (1 −PT ÷ 2)]

Where LFPR is the labour force participation rate, U is the unemployment rate and PT is the proportion of full-time work in a country.

2. Labour requirements. Calculate the labour requirements of the primary crop:

*Labour requirements = Farm area * Yield * Labour intensity*

Where farm area and yield can be collected easily, labour intensity is often times more challenging to find and can be collected from sector reports and academics studies.

3. Compare labour availability with labour requirements.

If labour requirements are higher than labour availability, this means that the farmer needs to hire labourers. The number of labourers to be hired, is exactly that difference between labour availability and labour requirements.

If labour requirements are lower than labour availability, this means that the farmer can go work off-farm and earn off-farm labour income. Alternatively, this can mean that the farmer can earn on-farm income from secondary crops. The limiting assumption is therefore that the former can proxy for the latter.

Where **input costs** are missing from formula (5), it may be necessary to research the costs of common fertilisers, seeds and pesticides and the amount applied per acre or hectare. The prices should be found in the same region as the study while the ratio of application should be found in relation to the crop being studied. These costs can be combined with labour costs to extrapolate the total variable costs. This process will have to be repeated for all crops involved. Gross profit from the main crop is calculated as production value minus variable costs. If gross profit from the main crop is known, but the variable costs unknown, this calculation can be re-engineered to discover variable costs.

3. Comparatively, **fixed costs**, necessary in equation (11) below, might be more difficult to estimate. They can include things like taxes on land and depreciation of farm equipment. While some studies may report on these, it is important that the variables are sufficiently comparable for either region or crop respectively. One way to account for fixed costs is to look for a source with percentages or ratios, such as the ratio of fixed costs to total costs.

(11) *Fixed costs = Cooperative membership fees + Security + Taxes + Operation and maintenance + Depreciation + Cost of establishment + Cost of land + Any other fixed costs*

It is possible that fixed costs include quite large, one-time investments. These investments would technically be expensed over multiple years. It is important to collect the context in which the data

on fixed costs is collected in the first place, so as to know how to use the data point in the way that provides the most accurate picture of yearly fixed costs.

It may be the case that sources indicate total costs for the analysis period, but not a breakdown of costs. If it is certain that these costs consider both variable and fixed costs, the calculation for farmer income can be adjusted accordingly.

4. Finally, farmer income is calculated as follows:

(16) *Total farmer income = Production value - Total costs + Other income + Off-farm labour income*

Other income may be available to be collected directly. Otherwise, income metrics like remittances can be researched. It is important here to make sure the variables have a comparable household size. If not, a ratio can be made of other income per household member and scaled to the household size being measured. In-kind income is also a key part of farmer income and dependent on crop diversification of the farmers. It is possible to extrapolate from sector reports or household surveys how much the farmer consumes of his own produce.

All variables need to be brought within the same time frame, whether this be through inflation or identified trends. However, as noted previously, when deciding between secondary sources for different variables, the comparability of certain scoping factors may matter more than others. Table 7 provides an overview of the most important scoping factors per farmer income variable. *Yearly variability* refers to how much a variable can change from year to year; those with high variability are very likely to change while those with low variability will remain stable over time. *Sensitivity to the model* refers to how much a variable would alter the resulting farmer income if it were to be changed. Variables with a high sensitivity will change the farmer income significantly when changed and those with low variability will not affect the farmer income much if changed. The *most necessary scoping factor* is the factor which is most relevant for comparability between the subject of the study and the secondary source. For each farmer income variable at least one scoping factor is given, but a secondary is also offered when necessary.

Table 7. The most necessary scoping factors for all actual income variables

ID	Variable	Yearly variability	Sensitivity to model	Most necessary scoping factor
1	Farm area (main crop)	L	H	Main crop
2	Yield (main crop)	H	H	Main crop (type of farming is also a highly important factor here)
3	Price (main crop)	H	H	Main crop (region is also a highly important factor here)
4	Variable costs (main crop)	H/M	M	Main crop (type of farming is also important here)
5	Labour available in household	L	L	Region
5	Gross profit (main crop)	H/M	H	Main crop (if found directly)
6	Net farm-income (total)	M	H	Main crop (if found directly)
7	Fixed costs (total)	L	M	Main crop (region is also important here)
8	Net off-farm labour income	L	M	Region
9	Average wage	L	L	Region (year is also important here)
10	Other income (non-farm, non-labour)	M/L	M	Main crop, region

4.2 Use case 2 – Monitoring progress towards closing the gap

As previously discussed, the use case 'Monitoring progress towards closing the gap' will require the estimation of farmer household income in multiple, consecutive years. The calculation framework below refers to an individual year and can be repeated multiple times for different years to assess and monitor progress towards closing the living income gap.

Table 8 provides an overview of variables and the suggested approach for the calculation, from the first variables (1, 2 and 3) to the total farmer income in step 12. The table should be read from the top to the bottom. Calculations are in bold. Above each calculation are the variables needed as input for this calculation, along with their required granularity and assessed availability. A detailed explanation of the approach is provided below.

Table 8. Granularity level and summary framework for monitoring progress towards closing the gap

ID	Variable	Calculation	Granularity	Suggested approach
1	Farm area (main crop)		M	Collect once, from a household survey or a sector report and assume it stays the same over the years
2	Yield (main crop)		H	Collect from international databases, year-dependent
3	Price (main crop)		H	Collect from international databases, year-dependent. Important to make sure prices are in the right currency-year; otherwise inflate.
4	**Production value (main crop)**	**1 * 2 * 3**		**Farm area (main crop) * Yield (main crop) * Price (main crop)**
5	Variable costs (main crop)		M	Collect once, from a household survey or a sector report and assume it stays the same over the years Calculate the ratio of Variable costs (main crop) to Production value (main crop) in original study, and multiply with **Production value (main crop)** calculated in row **4**.
6	**Gross profit (main crop)**	**4 – 5**		**Production value (main crop) – Variable costs (main crop)**
7	% farm income net of variable costs (other sources)		L	Collect once, from a household survey or a sector report and assume it stays the same over the years. Calculate the ratio of Farm income net of variable costs from other sources (other crops, livestock) over Farm income net of variable costs (total) in original study; then divide **Gross profit (main crop)** in row **6** with that ratio % farm income net of variable costs (other sources)
8	**Gross profit (total)**	**6 / 7**		**Gross profit (main crop) / % farm income net of variable costs (other sources)**
9	Fixed costs (total)		L	Collect once, from a household survey or a sector report and assume it stays the same over the years. Inflate to year of analysis
10	Net off-farm labour income		L	Collect once, from a household survey or a sector report and assume it stays the same over the years. Inflate to year of analysis

11	Other income (non-farm, non-labour)		L	Collect once, from a household survey or a sector report and assume it stays the same over the years. Inflate to year of analysis
12	**Total farmer income**	**8 – 9 + 10 + 11**		**Net farm-income (total) – Fixed costs (total) + Net off-farm labour income + Other income (non-farm, non-labour)**

The total famer income can be calculated in five steps.

1. The first step is the calculation of **production value** for the main crop ($_{mc}$) in year t ($_{t}$) (*Production value*$_{mc,\,t}$). Production value is calculated as the multiplication of farm area, yield and price. The assumption is that farm area will remain the same during the scope of the analysis, while yield and price of the main crop depend highly on the year of analysis. Therefore, farm area can be collected once from a household survey or sector report for the base year (year 0 leading to *farm area*$_{mc,\,0}$)) and be used in subsequent years. On the other hand, yield and prices will have to be collected for each year of analysis (year t leading to *yield*$_{mc,\,t}$ *and price*$_{mc,\,t}$). Ideally, this data is collected directly from farmers. In the case that primary data collection is not feasible, it is possible to collect yield and prices from international databases such as FAOstat.

$$(4)\ Production\ value_{mc,\,t} = farm\ area_{mc,\,0} * yield_{mc,\,t} * price_{mc,\,t}$$

2. The second step is the calculation of **gross profit from the main crop** in year t (*Gross profit* $_{mc,\,t}$). Variable costs for the main crop can be collected from the baseline household survey or sector report for the base year (*Variable costs* $_{mc,\,0}$). We then calculate the ratio of variable costs to production value in the base year (*Production value* $_{mc,\,0}$) and assume this ratio remains constant over the years. This is justified by the fact that variable costs are proportional to quantity produced. This ratio of variable costs to production value is then multiplied with production value for the year of analysis to get variable costs in monetary unit for the year of analysis. Farm income net of variable costs from the main crop in year t is then calculated as production value minus variable costs corrected for year t. A real-life example is presented in Box 9.

$$(6)\ Gross\ profit_{mc,\,t} = Production\ value_{mc,\,t} - \frac{Variable\ costs_{mc,\,0}}{Production\ value_{mc,\,0}} * Production\ value_{mc,\,t}$$

$$\leftrightarrow Gross\ profit_{mc,\,t} = Production\ value_{mc,\,t} * \left(1 - \frac{Variable\ costs_{mc,0}}{Production\ value_{mc,0}}\right)$$

3. The third step is the calculation of **gross profit from other sources**, including other crops and livestock raised on the farm. This involves collecting farm income net of variable costs from other sources from the baseline study or report (*Farm income net of variable costs* $_{other,\,0}$) and dividing it by farm income net of variable costs from all sources (*Farm income net of variable costs*$_{total,\,0}$). This yields the percentage of farm income net of variable costs that comes from other sources. We then divide the gross profit from the main crop in year t (*Farm income net of variable costs* $_{mc,\,t}$) with this percentage to arrive to total farm income net of variable costs.

$$(7)\ \%\ \text{farm income net of variable costs}_{other,\ 0} = \frac{\textit{Farm income net of variable costs}_{other,\ 0}}{\textit{Farm income net of variable costs}_{total,\ 0}}$$

$$(8)\ \textit{Gross profit}_{total,\ t} = \frac{\textit{Farm income net of variable costs}_{mc,\ t}}{\%\ \textit{farm income net of variable costs}_{other,\ 0}}$$

4. The fourth step is the subtraction of **fixed costs**, and the addition of **other sources of income** (e.g. remittances, off-farm labour income). Fixed costs are collected from the baseline study and inflated to the year of analysis. Off-farm labour income and other income (non-farm, non-labour) can also be collected from the baseline study and inflated to the year of analysis.

$$(9)\ \textit{Fixed costs}_{t} = \textit{Fixed costs}_{0} * \textit{Inflation factor}_{t}$$

$$(10)\ \textit{Off-farm labour income}_{t} = \textit{Off-farm labour income}_{0} * \textit{Inflation factor}_{t}$$

$$(11)\ \textit{Other income (non-farm, non-labour)}_{t} = \textit{Other income (non-farm, non-labour)}_{0} * \textit{Inflation factor}_{t}$$

5. Finally, **total farmer income** is calculated as follows:

$$(12)\ \textit{Total farmer income}_{t} = \textit{Farm income net of variable costs}_{total,\ t} - \textit{Fixed costs}_{t} + \textit{Other income (non-farm, non-labour)}_{t} + \textit{Off-farm labour income}_{t}$$

This farmer income figure can then be compared with a living income benchmark, in line with the *Guidance manual on calculating and visualizing the income gap to a Living Income Benchmark*. This exercise can be repeated for the years in scope of the users' project, using the relevant extrapolation methods outlined above. Once a farmer income figure is available for the different years and compared with a living income benchmark, it becomes possible to monitor progress towards closing the gap.

Box 9. Example: Input costs in Tony's Chocolonely

To contextualize the various extrapolation methods presented in this section, we provide an example of project work Impact Institute has conducted with Tony's Chocolonely. During their true price assessment, Impact Institute has calculated farmer income and living income benchmarks in Ivory Coast and Ghana for Tony's Chocolonely in order to assess the progress they had made towards closing the gap in comparison to a benchmark from 2013 to 2017.

For input costs specifically, the following data was available:

- Input costs in 2013 for the benchmark
- Input costs in 2013 for Tony's
- Input costs in 2017 for the benchmark

Input costs in 2017 for Tony's were missing. They were therefore extrapolated as follows:

1. The ratio of input costs for Tony's to input costs for the benchmark in 2013 was calculated:

$$\text{Ratio Tony's to benchmark (2013)} = \frac{\text{Input costs for Tony's (2013)}}{\text{Input costs for benchmark (2013)}}$$

2. Second, we applied this ratio to input costs in 2017 for the benchmark in order to derive input costs for Tony's in 2017, as follows:

$$\text{Input costs for Tony's (2017)} = \text{Ratio Tony's to benchmark (2013)} * \text{Input costs for benchmark (2017)}$$

The explicit assumption that had to be made was the assumption that the ratio of input costs Tony's to benchmark would remain constant over the years, which allowed us to derive input costs for Tony's in 2017.

5 Limitations and other considerations

In this chapter, limitations of the frameworks developed in the previous chapter as well as considerations collected during stakeholder consultations are discussed. Ways forward are also considered.

5.1 Limitations of estimating farmer income with secondary data

Acknowledging the limitations of estimating farmer income with secondary data is necessary to lay the foundation for ways forward. The most material limitations are listed in this section.

The accuracy of a farmer household income estimated using secondary data will never match that of a farmer household income estimated using primary data. Collecting primary data ensures the exact scope of the household income is followed, while in collecting secondary data, there will be some deviations from the exact scope, mostly due to low data availability. Therefore, as a general rule, using primary data where it is available is recommended. However, extrapolation is possible and makes for relatively quickly readily available figures that can inform strategy definition. By following a data hierarchy that matches the application at hand, as well as using renowned, validated studies, the risk of inaccuracy is reduced, and results will likely be valid and reliable.

5.2 Ways forward

A complete list of potential use cases considered for this document is provided in the introduction (table 1). Additional use cases were suggested during the stakeholder consultation. These include use cases focused on specific components of farmer income (e.g. subsidies) and specific household characteristics (e.g. gender gap in reaching the living income).

One possible way forward is the exploration of these potential use cases and the development of frameworks for these. The more frameworks are developed, the wider the audience that is reached and the more farmer household income are estimated. Similarly, it might be worthwhile to explore different frameworks for the use cases considered in this document. More frameworks can tailor for different levels of farmer income knowledge, but also for different types of data available, for instance a framework could focus on how to assess progress towards closing the gap when there is not a baseline study or household survey available. Additionally, another way forward would be to confirm the applicability of these frameworks by conducting case studies that are able to apply and test them.

Appendix A. Example of data characteristics to be compiled in a database: Yield

Table 9. Data characteristics to be compiled in a database: the example of yield

Variable	Description
Variable name	Yield (tonnes per ha)
Variable category	Yield
Region	Ghana
Crop/sector	Cocoa
Year	2012
Unit	Tonnes per hectare
Value	0.5
Comments	Data also available for Ivory Coast
Citation	p. 27
Source name	Cocoa Barometer (2012)
Source type	Sector report
Source link	https://www.voicenetwork.eu/wp-content/uploads/2019/07/Cocoa-Barometer-Full-2012.pdf

The Living Income
Community of Practice

Site: www.living-income.com
LinkedIn: www.linkedin.com/groups/13784101/
Mail: livingincome@isealalliance.org

IMPACT™ INSTITUTE

Address: Haarlemmerplein 2, 1013 HS, Amsterdam
Site: www.impactinstitute.com
Facebook: /impactinstitutecom
Twitter: impact_inst
Tel.: +31 202 403 440
Mail: info@impactinstitute.com

Supported by the

Federal Ministry for Economic Cooperation and Development

Implemented by

giz Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH

iseal alliance

Sustainable Food Lab