THE LAND INEQUALITY INITIATIVE

DATA PAPER

METHODOLOGICAL CONSIDERATIONS ON LAND INEQUALIY

BY DR. DELFINO VARGAS AND DR. CASSIO LUISELLI

INTERNATIONAL LAND COALITION | UNITED FOR LAND RIGHTS

WORLD INEQUALITY LAB

METHODOLOGICAL CONSIDERATIONS ON LAND INEQUALIY

BY DR. DELFINO VARGAS AND DR. CASSIO LUISELLI

STRATEGIC PARTNERS & CORE DONORS

Federal Ministry for Economic Cooperation and Development

IFAD Investing in rural people

Government of the Netherlands

Sverige

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Agency for Development and Cooperation SDC

ISBN: 978-92-95105-68-3

 Published: March 2021.
Editing: *David Wilson*. *Design*: *Federico Pinci*.

CONTENTS

INTRODUCTION

This study presents an original methodological proposal for measuring land inequality. It is based on two related observations. First, we emphasise that it is not enough to use a single measure, such as land ownership (as is often the case in existing literature and global databases), to capture the complexity of land inequality. For a more accurate measure, a number of indicators need to be incorporated to enable an assessment of the multidimensional nature of this phenomenon. Second, we underscore that land inequality cannot be measured by the Gini coefficient methodology alone, as this allows only for single variable assessments, besides other challenges. It is for these reasons that we propose the development of a multidimensional land inequality index, based on a series of indicators, through which the relevance of the indicators selected is analysed and countries are grouped using confirmatory factor analysis (CFA) and latent class analysis (LCA).

This report consists of three sections, following on from this introduction. *Section 1* discusses the multidimensional nature of land inequality, by combining – besides the standard quantitative indicator of size of land plots – tenure, quality of land, asset endowment, and other indicators. *Section 2* describes the different methodologies traditionally used to measure land inequality, in particular the Gini and Theil coefficients, and discusses their advantages and shortcomings, especially in terms of assessing land inequality from a multidimensional perspective. *Section 3*, in conclusion, proposes an original methodology, aiming at building a land inequality index that would take into account the multidimensional nature of the phenomenon.

GLOBAL LAND INEQUALITY METHODOLOGICAL CONSIDERATIONS ON LAND INEQUALIY

1

THE MULTIDIMENSIONAL NATURE OF LAND INEQUALITY

A clear analysis of the main drivers and dimensions of land inequality, across the rural world, is necessary as a first step towards understanding this phenomenon. This is a challenging task, given the different ways in which land inequality manifests itself in different countries and regions. Nevertheless, it seems to be characterised by certain common features, which allows us to believe that a conceptualisation of land inequality and a common methodology for its measurement are possible. To do this, we need a set of clear-cut, robust, direct, operational, and yet simple indicators. There are four sets of indicators, including different aspects of land inequality, which appear to be sufficient to encompass its multidimensional nature.

Size of holdings is relevant but has major constraints

The distribution of plot sizes and land holdings is most commonly used to assess land inequality. Although this measure is certainly relevant in terms of data availability, it entails a number of challenges.

First of all, the size of plots does not allow an understanding of, for example, associative or multiple ownership of plots. Also, it may be argued that land size is a relative concept, since it is not the same to possess a large piece of desert or mountainous land as it is to have a fertile, flat, and well irrigated piece of land. Nevertheless, comparisons are still valid, taking into account the specific conditions of the country or region where the land holding might be located.

The need to recognise diversity of tenure and land rights

In many if not most emerging countries, private property rights to land are not the only way to access and control land and its assets. In many cases, collective and community-based land ownership rights hold sway, with communities themselves being owners or (local or national) authorities acting as guarantors to protect or restrict rights to possess, acquire, and use land. Other tenure systems such as leasehold and sharecropping can also exist, in parallel or in co-existence with these.

Although not an exhaustive list, these different tenure systems highlight the combination of property and formal or informal contracts needed to access and use land. They also highlight the potentially varied implications of these for land inequality (see *Box 1*) – which makes it important for land tenure and land rights to be considered in any assessment of land inequality.

Box 1: Combining ownership rights with formal and informal contractual arrangements

Property rights/ownership

Full ownership, fully accepted and certified by legal means, is one of the main ways to secure land rights and also to provide the right incentives for investment and production.[1] This creates a land market that in most countries is legally constrained in order to avoid excessive concentrations of land. At least formally, there are provisions in place to protect small farmers' ownership of their land, mostly against abusive credit or other economic agents who might put their possession at risk.

Formal/informal contractual arrangements to secure land access

Agrarian history shows numerous examples of what in general terms is known as "sharecropping" (aparceria or mediería in Spanish), where a landowner allows tenants to use their land in return for a share of the crops produced. From this basic concept, many diverse forms of contract or arrangement can be derived. This type of arrangement might also offer a way to accumulate access to productive land.

This is relevant with regards to more modern types of ownership, such as shareholdings by publicly listed entities (see Merlet, 2020) or private equity ownership of financialised land assets (see Wegerif and Anseeuw, 2020).

Examples of such tenure systems that need to be considered include:

- Sharecropping: Sharecropping (as described in *Box 1*) is a very common arrangement that has been widely analysed in terms of both its efficiency and equity terms. Even if it seldom entails an optimal distributive arrangement – the so-called "puzzle of sharecropping", as noted by Alfred Marshall (1920) – it is nevertheless very common. Sharecropping addresses market failures in credit or it can serve as a risk-sharing contract (Stiglitz, 1974).
- Land renting or leasing contracts: Leasing is a very common way to gain access to land.[2] Although leasing may not formally deprive the owner of his right to the land, in practical terms contractual arrangements of this kind give the lessee control over production and resources on the land, contributing indirectly to inequality. Both the enforcement and regulation of contracts to avoid excesses and undue accumulation are key variables to consider in this regard. Arrangements such as "contract agriculture" (commissioning certain crops, volumes, and qualities of produce) are also common.

1 The degree of security and certainty pertaining to these rights of tenure is very relevant, as is somewhat reflected in the u_1–u_8 set of variables or indicators. The decomposing of the general index is an additional task that can be achieved once we have formally defined these basic variables.

2 Leasing can include, for instance, the process known as "land grabbing", when an actor from a rich foreign country acquires control of land in the territory of another country, usually poorer and with weaker law-enforcing mechanisms; such arrangements are generally made on the basis of a lease or concession.

- Association as a distinctive way to control or acquire land: Lastly, there are arrangements of different forms that entail the association of a number of owners to hold ownership of a single piece of land, or the merging of lands owned by two or more parties. This is often connected to size and economies of scale. This mode of land accumulation is becoming very relevant in understanding inequality with regards to land and water and new ownership models such as shareholding.

Quality of land (or land asset value) as a key variable

Soil quality and fertility are major factors in differentiating land quality, irrespective of the ecosystem or geographical occurrence of the land holdings. Here, the main distinction to take into account is whether the land is irrigated or rain-fed. Availability of water is often a major factor determining inequality when it comes to land and resources. The advantages of irrigated land are obvious: it is not only more productive per unit of land area, capital, or labour, but it is also less risky. Often the value and price of such land are higher, depending on the investment made in technology.

There are at least two other key attributes of land that must be considered, besides water, to understand its value and the potential return on its assets and natural fertility. One is asset or capital endowment within the farm holding or productive unit. Proper proxy variables to this are tractor ownership or availability (the degree of "tractorisation"), as well as ownership of or access to equipment such as combine harvesters and irrigation pumps. The second key attribute is location and accessibility, including proximity to towns and cities (input or output markets), which lowers transaction costs.

Other dimensions of land inequality

Gender Inequality is a relevant dimension of inequality and exclusion. This is another key variable for understanding inequality in land access, via property or contractual – mostly informal – rights. It is quite common that within a household or family it is the male (husband, eldest son) who holds property or contractual rights.

Additionally, the distribution of work and domestic chores within a rural family is often biased, putting an enormous burden on women (wives, daughters, sisters, and others) (Scalise, 2020).

Foreign control of land or "land grabbing" in emerging countries as a disturbing trend

A recent phenomenon is the acquisition of vast tracts of land by an actor from one country (usually a government or a private company) in the agricultural area of another. This is often referred to as "land grabbing"[3] or "large-scale land acquisition".

Most of the time, this involves not a market purchase but a long-term leasing contract. These long-term leases are often arranged via government authorities, with no consultation with local stakeholders, and they frequently create in effect a foreign enclave within the country's territory. Very often, this implies the exclusion of local people from any type of benefit or participation, such as labour, supply of inputs, etc. In Madagascar and Mozambique, for instance, vast tracts of land have been acquired by foreign companies and in some cases governments in order to secure supplies of cash crops and other commodities for themselves. This not only potentially displaces poor populations from agricultural land, but also precludes them from access to food and other natural resources in their own country or region. This trend has been extensively documented in recent years (for example, see Vellvé and Rakototondrainibe, 2018).

Data availability – although limited, some useful sources do exist

Data are not homogeneous, and many sources do not include all measures needed to quantify inequality when using traditional measures. Administrative data are not available for all countries, and household survey data do not sufficiently cover rural areas, indigenous populations, or peasant women working in agriculture.

The multidimensional nature of land inequality highlighted in this paper underscores the challenges of data availability, as numerous dimensions and thus variables need to be covered. Data for the required indicators are not always available at the country level: for example, few countries document land ownership by gender, or take into account local traditions of land heritage. There is a need for qualitative studies at the country level to collect this type of information.

Table 1 gives examples of possible data sources for the different indicators needed for our proposed land inequality index, which are available from sources such as FAOSTAT, Prindex, and AQUASTAT. Other variables and sources could, of course, also be used and aggregated.

3 In Spanish, "extranjerizacion de la tierra".

Table 1: Dimensions, variables, and potential data sources

INDICATOR	CODING AND DESCRIPTION	POTENTIAL DATA SOURCE
PROPERTY RIGHTS DIMENSION		
u1	% of privately owned land (titled)	Prindex
u2	% of privately owned land (untitled)	Prindex
u3	% of communal land	Prindex
u4	% state-owned (and reserved land)	Prindex
LAND RENTING AND LEASING		
u5	% of rented land (formal contracts)	Prindex
u6	% of rented land (informal contracts)	Prindex
u7	% paralegal possession of land (de facto occupation)	Prindex
LAND ACCUMULATION		
u8	Acquisition of multiple adjacent parcels of land (legal or paralegal)	Prindex
SIZE OF TENURE HOLDINGS		
u9	Size rank distribution by tenant (2–5 ha; 5–10 ha, 10–15 ha, etc.)	FAOSTAT, AQUASTAT
QUALITY OF LAND		
u10	Fertility sub-index (soil quality 40%; humidity 35%; flat/steep terrain 25%)	FAOSTAT
u11	Rain-fed (mainly = over 90%)	
u12	Irrigated (mainly = over 80%)	FAOSTAT
WATER AVAILABILITY		
u13	Access available (to all >90%)	AQUASTAT, FAOSTAT
u14	Pervasive water monopolies ("water latifundia")	
ASSET ENDOWMENTS		
u15	% mechanisation and infrastructure	FAOSTAT
u16	Accessibility, roads, and location	
GENDER INEQUALITY		
u17	Degree of legislation missing for women's rights (property, legal contracts)	Prindex, FAOSTAT
u18	Intra-household workload, farming activities	Prindex

INDICATOR	CODING AND DESCRIPTION	POTENTIAL DATA SOURCE
LAND GRABBING		
U19	Land aggregation through leasing/renting by foreigners	Land Matrix
COVARIATES: GINI COEFFICIENT		
x1	Gini – size of hectares (tenure)	FAOSTAT
x2	Gini – income of farmers	Prindex
COVARIATES: RURAL/URBAN		
x3	% rural vs. urban land, population	Prindex, FAOSTAT

Prindex The Global Property Rights Index, which measures global perceptions of land and property rights. *https://www.prindex.net/*

FAOSTAT Data on food and agriculture from the Food and Agriculture Organization of the United Nations (FAO). *http://www.fao.org/faostat/en/*–home

AQUASTAT FAO's global information system on water and agriculture. *http://www.fao.org/aquastat/en/*

Note: Some of the indicators and measures used could be approximations, and we therefore face a measurement problem (i.e. variables could be measured with an ordinal scale instead of a percentage). The measurement error problem in surveys is recognised in a number of studies, and can be attributed to sampling problems or to the survey method used (Biemer et al., 2013). Such measurement errors with multiple indicators (named ui and known as manifest variables) can be assessed, which means that it is still possible to create an index of inequality for each country. This index is constructed on a continuous scale.

2

METHODOLOGICAL PROPOSAL

In this section, we first explain the use of the Gini coefficient and Theil's entropy index, and then explore the advantages and disadvantages of these measures.

Existing literature and data generally assess land inequality using the Gini or Theil coefficients. However, as already stated, recognising the complexity of land inequality implies the need for a broad range of indicators, which makes these methodologies either less relevant or not applicable at all. To overcome such limitations, confirmatory factor analysis (CFA) can be used to construct a single index from the multidimensional indicators for each country (Brown, 2015). This CFA index produces scores that enable the ranking of countries from high to low levels of inequality, based on all the indicators previously described. This is the first step of our methodology proposal. The second step involves latent class analysis (LCA) (Hagenaars and McCutcheon, 2002), which allows for the identification of country groups based on similar characteristics of inequality. Before presenting this innovative methodology, we first assess the challenges of the Gini coefficient and the Theil index.

Current measures of land inequality: Gini and Theil coefficents

The Gini coefficient

The Gini coefficient is a widely used and general measure of inequality. It is a measure of statistical dispersion intended to represent inequality within a nation or any other group of people. It was developed by the Italian statistician and sociologist Corrado Gini and first published in 1912. The Gini coefficient measures inequality between values of a frequency distribution – for example, levels of income. A Gini coefficient of zero expresses perfect equality, where all values are the same (for example, where everyone has the same income). A Gini coefficient of one (or 100%) expresses maximal inequality among values (for example, for a large number of people where a single person has all the income or consumption and all the others have none, the Gini coefficient would be nearly one). The measure is generally used for income and wealth distributions but can also be used to assess other types of inequality, such as land inequality.

Regarding land, the Gini coefficient is typically used to assess the distribution of plot size (although assets or income from agricultural sales have also been used to measure land inequality). The Gini has also been used to correlate land inequality with other issues, such as land conflicts (see *Box 2*).

Box 2: Using the Gini coefficient to assess land and other inequalities

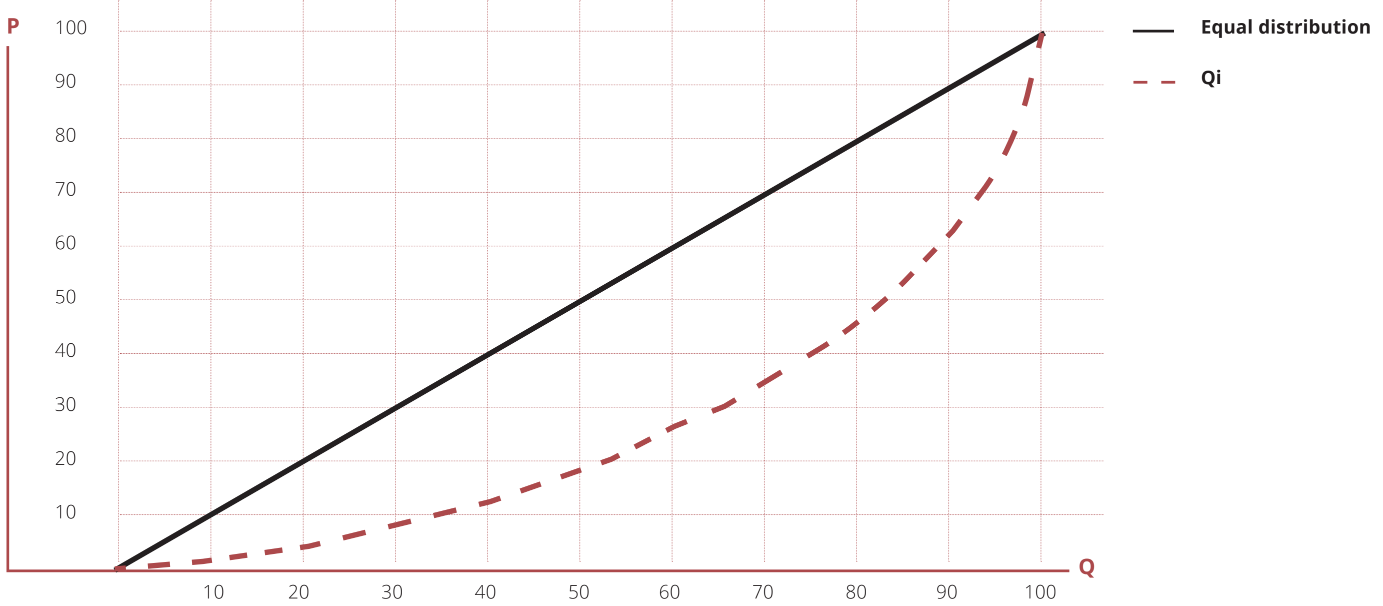
To cite some examples (but without aiming to be exhaustive), Heshmati (2004) uses the Gini coefficient to measure inequalities in income and employment but also in non-income measures such as educational opportunities, human capital, growth, health, life expectancy, welfare, and happiness.

The Gini is sometimes used to assess the correlation between land inequality and other aspects of inequality. Martinelli (2012) utilises it (along with Theil's entropy index) to link land inequality and the inefficiency of market power in Italy. Albertus, Brambor, and Ceneviva (2016) examine the Gini distribution of land holdings in 12 countries that have experienced land conflicts, highlighting the relationship between such conflicts and the Gini landholding index. Cipollina, Cuffaro, and D'Agostino (2018) use meta-analysis techniques to compare the Gini coefficient for economic growth with that for land inequality. Other studies have used it to assess water distribution, for example in South Africa (Cullis and van Koppen, 2007).

In addition to the Gini coefficient, other measures can be used to compare deciles or quartiles of land holdings in hectares. For example, Bharti (2019) uses the Palma ratio to compare the top 10%, the middle 40%, and the bottom 50% shares of assets in total wealth, as well as land inequality in deciles from 1961 to 2012 in India. Additionally, Bharti employs the Gini to calculate land distribution in several countries in order to compare land inequality around the world.

The Gini coefficient can also be represented in graphic form as the Lorenz curve, as presented in *Figure 1*.

Figure 1: The Lorenz curve



Source: Cortés (2020).

The abscissa (x) axis represents the cumulative percentage of cases (Q), after ordering them from the lowest to highest value of the variable, and the ordinate (y) axis represents the proportion of the variable (P). The diagonal line in *Figure 1* represents a situation where the accumulated percentage of cases is equal to the accumulated percentage of the variable. This diagonal is called the equidistributional line. The triangle formed by the points (0, 0), (100, 0), (100, 100), whose area is equal to ½ represents the area of maximum concentration, since it corresponds to a situation in which the last case has the entire variable (100%). The dotted line – the Lorenz curve – represents the effective distribution of the variable between the observations. The area bounded by the equidistributional line and the Lorenz curve is called the concentration area, which we will symbolize by A.

The Gini index results from dividing the concentration area by the maximum possible concentration:

$$G = \frac{A}{\frac{1}{2}} = 2A$$

From this equation, we derive that if the distribution of the variable is equitable G = 0 since A = 0, and if it is totally and absolutely concentrated A = 1/2 and therefore G = 1. Consequently, the range of possible values of G fluctuates between 0 and 1, $0 \le G \le 1$. The index assumes the value 0 if the variable is equidistributed and 1 in the case of maximum inequality.

In concrete terms, this equation is nothing more than an average of all the possible differences between pairs of magnitudes of X, scaled by the average of the variable X to prevent the measurement from being sensitive to changes in scale.

The Theil entropy index

Another popular index for measuring land inequality is the Theil entropy index, often referred to simply as the Theil index. This is a measure of land inequality based on the concept of entropy in communication theory. This index can be used to indicate the degree of inequality in land distribution, and its interpretation is similar to that of the Gini coefficient: high values of Theil entropy are associated with high levels of inequality (Cortés, Ruvalcaba, and Fernández, 2014). Additionally, Theil's entropy index compiles the decomposition of inequality in an additive way,[4] which allows decomposition in groups (Lora and Prada, 2016: 22). In this study, we use the decomposition of the Theil coefficient to break down land inequality into portions of inequality between groups of countries, obtained by means of latent class analysis. Once we have created a Theil index of inequality, it is relevant to understand the extent to which each of the groups contributes to the overall inequality.

4 The term "decomposition" refers to the separation of the variance attributed to land inequality between groups.

The Theil formula for land distribution is as follows:

$$T = \sum_{i} X_i In(X_i n)$$

where n is the number of countries $i\text{-}th$, and X_i is the concentration of the total land distribution for the i-th country. When land is distributed equally, all the X_i values take the value of $1 / n$, and $In(X_i n)$, is equal to zero (the logarithm of one is equal to zero). Therefore, when the Theil index equals zero, there is an equal land distribution among countries.

The Theil index is part of this methodological proposal because we foresee the formation of groups of countries affected by different degrees of land inequality, and we therefore propose to disaggregate the Theil index in groups. The inequality between groups tells us how heterogonous each group is compared with the others. Furthermore, the Theil decomposition can answer the question, "Which portion contributes the most to the overall inequality measure, the 'between' or the 'within' portion?". The formation of groups is discussed later in this paper.

Relevance and limitations of the Gini coefficient and the Theil index for land inequality measures

The Gini coefficient, based on the probability distribution function (PDF) of land size ownership, is particularly easy to calculate when raw data are available — indeed, that is the main reason for its popularity. In addition, such an index can be interpreted intuitively.

This being said, the Gini coefficient is a relative measure, and its proper use and interpretation can be controversial. A changing Gini coefficient with regards to land inequality can be due to structural changes in a society, such as population growth or decline (baby booms, an ageing population, increased divorce rates, extended family households splitting into nuclear families, emigration, immigration) and income mobility. Gini coefficients are simple, but this simplicity can lead to the complexities they embody being overlooked, especially in the land sector.

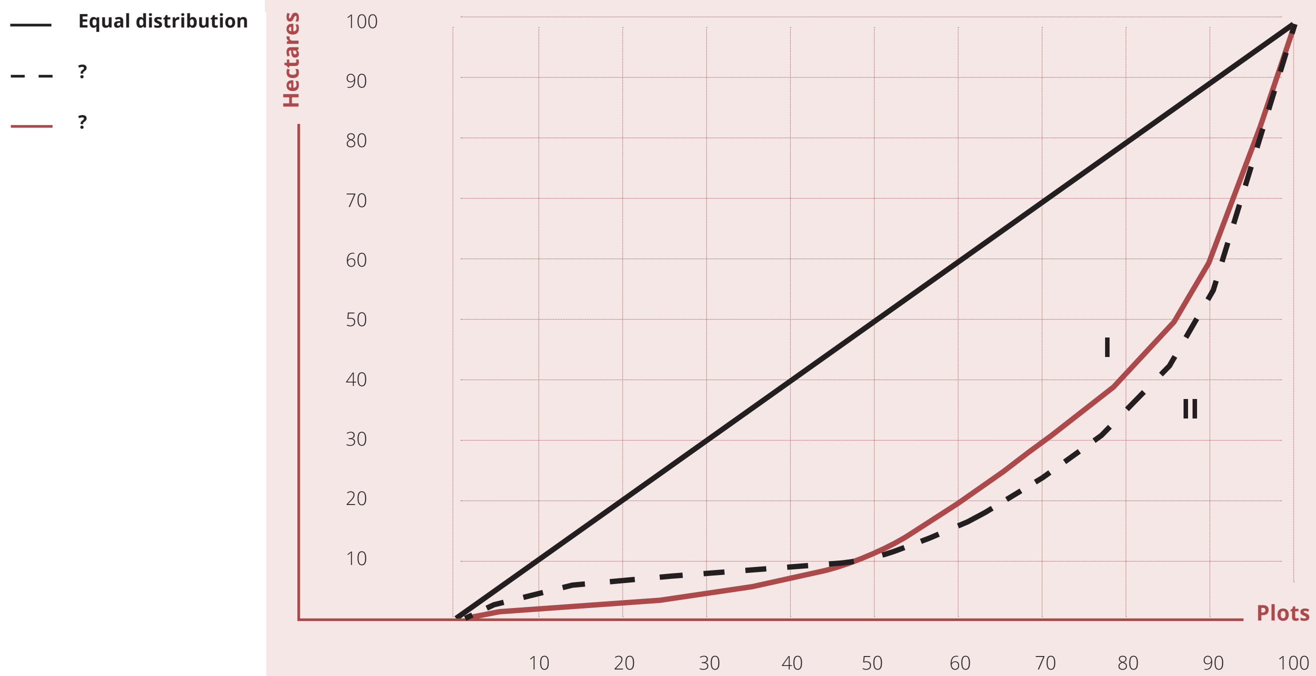
Different distributions with the same Gini coefficient

Also, even when land distribution is the same overall, in certain situations two countries with different distributions can have the same value on the Gini index. This is related to the fact that the Gini coefficient for land inequality measures the scaled average of all the possible differences between the amount of land owned or controlled by people, companies, organisations, and others. As such, it is an average and therefore the same value can be consistent with many different distributions. This means that a given Gini index could correspond to various different Lorenz curves that present a similar land concentration. Therefore, no analysis of land inequality can be exhaustive with a simple calculation of the Gini index (see *Box 3*).

Box 3: Same Gini but different (land) inequality distributions

Figure 2 represents a simulated approximation of the number of plots (x axis) and hectares (y axis). The straight line represents an equal distribution of land, and the Gini coefficient is zero. Additionally, this figure shows two hypothetical populations with the same Gini index. The area under the straight line for curve I is almost the same as for curve II. However, the distribution of the land is different and, while the degree of inequality is almost the same in both populations, the shape of land concentration is different. This shows how limited the use of the Gini coefficient can be in the measurement of land inequality.

Figure 2: Two Lorenz curves with the same coefficient but different degrees of land concentration



Source: Cortés and Rubalcava (2014: 201).

Another limitation in using the Gini coefficient – and to some degree also the Theil index – is its application to single dimensions. These methodologies do not fully comply with the additive decomposition property. To some extent the Gini index can be used, but all the sub-groups do not add up to the original disaggregated index. For that reason, we turn to the use of the Theil entropy index, which accounts for the decomposition of the inequality measure, and instead use the Gini coefficient as an explanatory variable, as explained below.

GLOBAL LAND INEQUALITY METHODOLOGICAL CONSIDERATIONS ON LAND INEQUALITY

Creating a global index of multidimensional land inequality via factor analysis (FA)

As mentioned above, the indicators required for assessing land inequality are multidimensional and take into account numerous aspects such as property rights, size of tenure holdings, quality of land, water availability, asset endowments, gender inequality, and foreign ownership.

In general, indices tend to generate an average of indicators, in an approach based on "coarse factors". The problem with this approach is that all variables are assumed to be equally important (Brown, 2015). Instead, the index we propose incorporates in its structure a correlation of inequality calculated from each variable, and generates a score as a linear combination of the variables weighted by the estimated loads for each variable. These weights are estimated using the maximum likelihood estimation (MLE) method and reflect the contribution of each variable to the global inequality index. This method is called a confirmatory factor analysis (CFA) index (Brown, 2015). Additionally, with this approach it is possible to assess the reliability and validity of the data using Ω-composite reliability. Those variables that have important weights are kept to conform the index (see the example in *Annex 1*, where this procedure is explained).

Using multidimensional data for each country, a CFA is conducted to generate an index of land distribution. This estimates the importance of each manifest variable of land inequality for each country. There might be issues with missing data in such an index, since data for all variables might not be available in all countries. However, this is not a problem for estimating the overall index, which allows all countries to be included in the analysis (see Arbuckle, 1996).

Confirmatory factor analysis

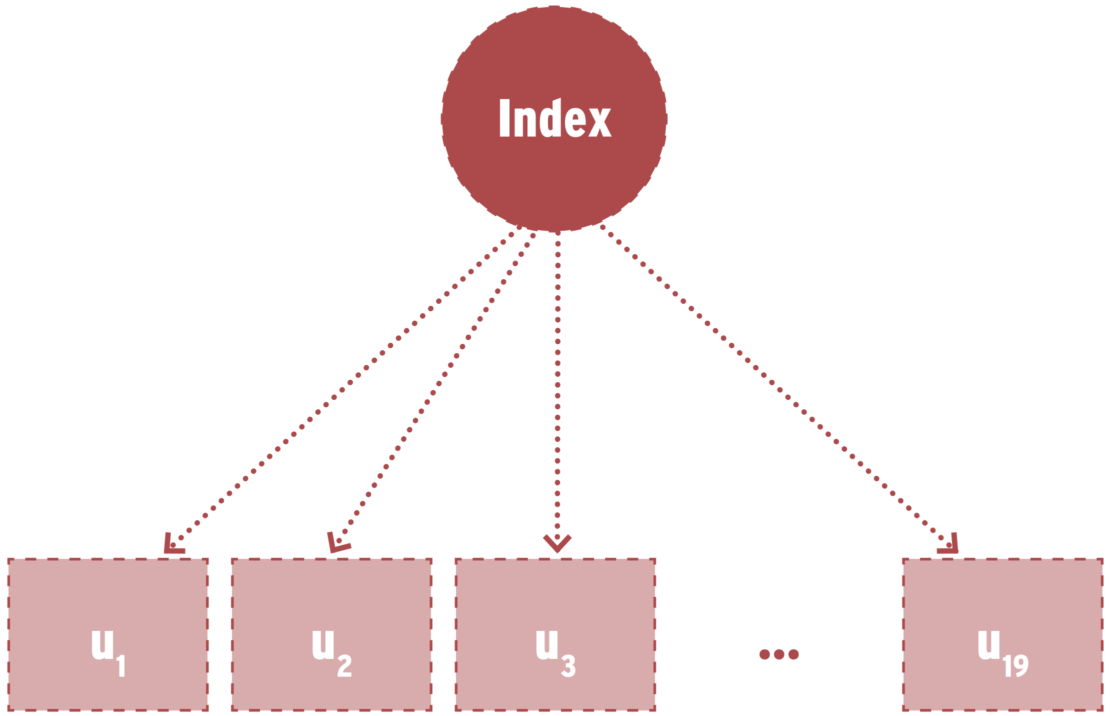
The basic idea of CFA is to explore the correlation structure of the observed variables and to generate latent variables that identify one or more different dimensions that are subjacent. Possibly, in this analysis we might identify that some variables are grouped in dimensions with a high degree of correlation. Typically, we have observed variables (or manifest variables) that we denote with the vector $\boldsymbol{u} = [u_1, u_2, ..., u_p]'$ and we assume, by construction, that the vector of factors (latent variables) exists: $\xi = [\xi_1, \xi_2, ..., \xi_n]'$ *y* $\delta = (\delta_1, \delta_2, ..., \delta_p)'$ denotes the residual variances, which are supposed to incorporate the correlation structure of the manifest variables.
The factorial model in matrix form is

$$\boldsymbol{u} = \alpha + \boldsymbol{\Lambda}\xi + \delta,$$

The matrix $\boldsymbol{\Lambda}$ has dimensions p × k and contains the estimated factor loads via maximum likelihood. CFA is a consequence of the EFA; the difference is that the first imposes restrictions on the formation of the latent variables identified in the exploratory phase and the factor loads are estimated for each indicator variable (Brown, 2015).

For the construction of the index, which we will call quality of access to land, we propose to measure this quality through the manifest variables $\boldsymbol{u}$ obtained from *Table 1* at the country level, within the obtained variables including property rights, land renting and leasing, land accumulation, size of tenure holdings, quality of land, water availability, asset endowments, and land grabbing (see the example in *Annex 1*).

Figure 3: Global index of multidimensional inequality of land distribution, via CFA



Note: Indicator variables as described in ***table 1***. Include property rights, land renting and leasing, land accumulation, size of tenure holdings, quality of land, water availability, asset endowments, gender inequality, and land grabbing.

Source: Elaborated by the authors.

Box 4: Example of an index created using CFA

The following data show a hypothetical correlation matrix between four manifest variables related to the construct of property rights measured in 80 countries. The variables are: u_1 = % of privately owned land (titled), u_2 = % of privately owned land (untitled), u_3 = % of communal land, and u_4 = % state-owned (and reserved) land.

Table 2: Simulated correlation matrix of four indicators of a property rights construct

	U_1	U_2	U_3	U_4
U_1	1.00			
U_2	0.768	1.00		
U_3	0.552	0.528	1.00	
U_4	0.582	0.615	0.707	1.00

Source: Authors' elaboration, simulated data.

Single factor loadings and residual variances are applied to this matrix, and the estimated loading weights are shown in *Table 3*.

Table 3: Factor matrix loadings and residual variance

VARIABLES	LOADINGS (Λ_i)	RESIDUAL VARIANCE ($1 - \Lambda_i^2$)
U_1	0.846	0.284
U_2	0.854	0.271
U_3	0.696	0.516
U_4	0.752	0.434
SUM	3.148	1.505

The loadings in *Table 3* are obtained by applying the MLE method to the correlation matrix shown in *Table 2*. In this hypothetical example, the variable D1 contributes the most to the property rights index (0.846) and the variable D3 contributes the least (0.696). The residual variance is obtained using the expression $1 - \lambda_i^2$, where λ_i represents the factor loadings for each variable. For example, if $\lambda_i = 0.846$, then the residual variance is $1 - \lambda_i^2 = 1 - 0.846^2 = 0.284$. The lower the residual variance, the higher the contribution to the index. This matrix shows the measurement error of the construct called property rights. Additionally, the composite reliability can be calculated using the loadings from *Table 3*; this is called omega (Ω) reliability. We use the following expression:

$$\Omega = \frac{\left(\sum_i \lambda_i\right)^2}{\left(\sum_i \lambda_i\right)^2 + \sum_i Var(e_i)}$$

where λ_i represents the factorial loads and $Var(e_i) = 1 - \lambda_i^2$ represents the residual variances. Using the results from *Table 3*, the composite reliability Ω is as follows:

$$\Omega = \frac{3.148^2}{3.148^2 + 1.505} = 0.868$$

The value of 0.868 indicates an acceptable reliability coefficient, since it is higher than 0.7 (Nunnally and Bernstein, 1994). Therefore, the factorial scores can be obtained using the linear combination from *Table 3*:

$$Score\ i = 0.846\ u_{1i} + 0.854\ u_{2i} + 0.696\ u_{3i} + 0.752\ u_{4i}, \qquad i = 1, \ldots 80$$

For the ith--country with the measures u_1 to u_4, we estimate the score so we can have a different property rights score for each of the 80 countries; in this way, we can rank the scores for all countries, from low to high. In addition, we can calculate the property rights index Gini coefficient.

We continue obtaining the factor scores for all nine dimensions listed in *Table 1* in the text using the same procedure.

Formation of groups of inequality via latent class analysis

In addition to the CFA phase, latent class analysis (LCA) is useful for conforming groups of countries with homogeneous characteristics within groups and those with heterogeneous characteristics between groups. The CFA index allows us to estimate the Gini and the entropy (Theil) indices for each group. The groups conformed will be informative of the inequalities of land distribution around the world.

Latent class analysis

This proposal includes LCA as part of the methodology for measuring multidimensional land distribution. LCA allows the formation of groups of countries based on the covariates or defined variables (as shown in *Table 1*) to explain multidimensional inequality by forming groups of countries with similar characteristics. LCA is different from multivariate clustering techniques, also called cluster analysis (Hagenaars and McCutcheon, 2002). There are many differences between the two approaches, but the main one is that LCA uses a parametric approach to the formation of groups and is based on a model, while cluster analysis uses a non-parametric approach to group observations based on an algorithm. However, in cluster analysis it is not possible to use a model for grouping the countries into typologies (see Vermunt and Magidson, 2002 for details).

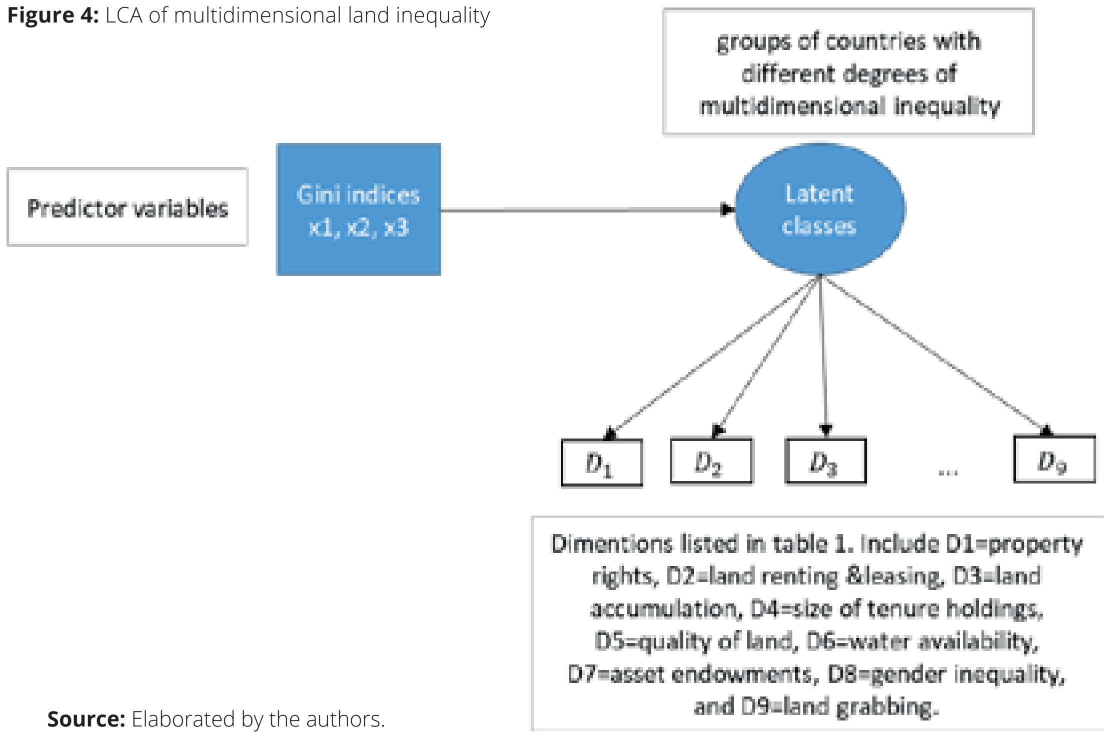
The LCA approach is used to estimate the probability of belonging of each i-th observation to each group K and to define all the characteristics that component of U that with n independent and identically distributed observations (IIDs), such that $P(u_i \mid u_j) = P(u_i)\, P(u_j)\ \forall\ i \neq j$ with a function of density (or probability) $f(u)$ (Muthén 2002; Vermunt and Magidson 2002).

The groups that we call latent classes are homogeneous within groups and heterogeneous between groups. The objective of the LCA is to show a gradient in the formation of these groups to later describe their characteristics. LCA can be studied from the perspective of models with mixtures of distributions. Actually, these are specific forms of parametric distribution functions in each of the underlying populations. It is based on the principle that there are at least two mixed populations and that more populations can be found that can be separated based on a model (Wedel and DeSabro, 2002). In our case, the essential characteristic of the latent classes consists of disaggregating the countries without identifying the region and identifying the distribution function to which they belong, thus evaluating the probability of each country belonging to each of the defined typologies, in each of the latent classes. This approach is focused on the countries, not on the variables selected to measure the quality of access to land.

LCA is very similar to factor analysis (FA). FA starts from the idea that the variables have a correlation structure between them, which can generate latent variables (called factors) that group a set of correlated variables, and that a practical sense of interpretation can be given to each identified factor; this approach is focused on variables. Meanwhile LCA focuses on the formation of groups of countries that are homogeneous within and heterogeneous between groups.

Additionally, we propose a conditional form of LCA. This model is shown in *Figure 4*, where the Gini coefficients and other covariables at the country level are relevant in improving the formation of groups. The covariates are indicated as xi in this model.

Figure 4: LCA of multidimensional land inequality



Source: Elaborated by the authors.

Sample size: It is necessary to collect information for the indicators listed in *Table 1* from at least 100 countries, as the method requires this amount of data.

Missing values: The multidimensional indicators in this table are named u_i, and the predictors are named x_i. This information is needed using a percentage scale as indicated; if the information is missing, researchers should avoid the use of zeroes but instead code this as a missing value, using an empty cell. Observations from countries with missing values need to be kept in the data set. Information is needed for each country, and the sample size should be at least 100 countries. This method can handle the missing values and no country is left out of the analysis (see Arbuckle, 1996 for further explanation).

Box 5: An example of forming groups using LCA

Once we know the scores for the nine dimensions obtained via the previous step, we can apply LCA to obtain groups of countries. Using simulated data that contain the scores for the nine dimensions of all 80 countries in the model described in *Box 4*, we first determine the number of latent classes (see *Table 4*).

Table 4 has been calculated for four class solutions. Five criteria are used to determine the number of classes (see Hagenaars and McCutcheon, 2002): 1) the lower the Bayesian Index Criterion (BIC); 2) the higher entropy index (closed to 1); 3) the minimum percentage of countries has to be higher than 5% (to avoid countries belonging to an outlier group); 4) the higher probability of correct classification into each class (closer to 1 is better); and 5) the p-value less than 0.05 for the Lo–Mendell–Rubin (LMR) Test. The solutions for three classes meet all five of these criteria.

Table 4: Determination of the number of latent classes

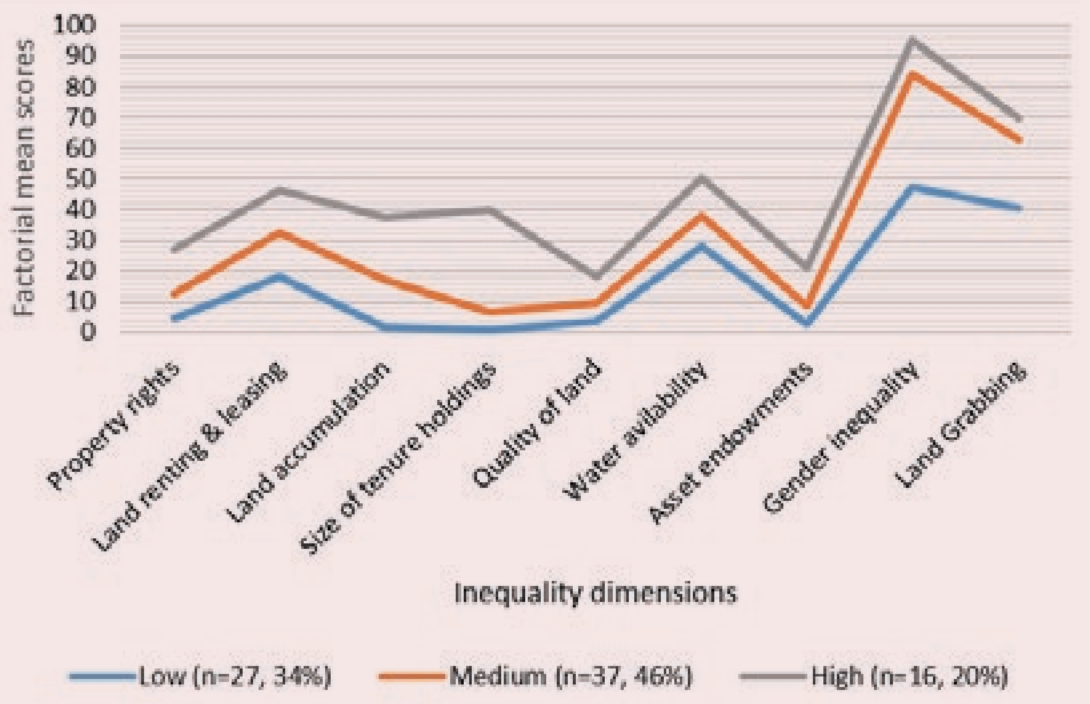
CLASSES	BIC	ENTROPY	% MIN–MAX	PROBABLE CLASSIFICATION	LMR-ADJUSTED
1	167,594.3				
2	161,591.9	0.864	41.3%-58.6%	.952–.966	0.0004
3	158,775.3	0.883	20%, 34%, 46%	.938–.952	<0.0001
4	157,051.7	0.907	1.1%, 15%, 38%, 45.9%	.937–.985	0.2141

Using the three class solutions, we obtain the mean scores of all nine dimensions. For the simulated data, we might classify the 80 countries into three categories of land inequality. As shown in *Table 5*, we have 27 countries (34%) in the low inequality group, 37 countries in the medium inequality group (46%), and 16 in the high inequality group (20%).

Table 5: Means for the three class solutions for the nine dimensions of land inequality

	MEANS		
DIMENSIONS	LOW (N=27, 34%)	MEDIUM (N=37, 46%)	HIGH (N=16, 20%)
D1 PROPERTY RIGHTS	4.7	12.4	26.8
D2 LAND RENTING AND LEASING	18.2	32.6	46.2
D3 LAND ACCUMULATION	1.7	16.9	37.6
D4 SIZE OF TENURE HOLDINGS	0.8	6.5	39.9
D5 QUALITY OF LAND	3.8	9.6	18.2
D6 WATER AVAILABILITY	28.2	38.2	50.3
D7 ASSET ENDOWMENTS	2.8	8.5	21.2
D8 GENDER INEQUALITY	47.7	84.2	95.3
D9 LAND GRABBING	40.3	63.0	70.0

Figure 5: Graphical representation of the means of land inequality, using LCA



The means data listed in *Table 5* is represented in graphical form in *Figure 5*. All 80 countries are classified in terms of their dimensions of inequality, and 16 can be identified as being highly unequal.

Decomposition of inequality

In the decomposition of the Theil index, three components are observed:

» the inequality within each group, which is entropy;

» the contribution of inequality within each group to total inequality, which results from the degree of entropy and the proportion of income received by each group (which is a result of the income level and group size); and

» the contribution of inequality between groups to total inequality.
In this regard, we are able to describe the contribution to the inequality index, comparing inequality within and between groups of countries.

3 CONCLUSION

Land inequality is multidimensional and, to measure it, it is necessary to consider a number of variables, as described in this paper. Such variables relate to diverse tenure arrangements, which entail differences in access to land as well as the quality of land, endowments, and assets. Together, these variables allow us to comprehend these various dimensions and to measure land inequality more precisely.

This proposal also includes the construction of a global land inequality index that would take into account property rights, land renting and leasing, land accumulation, size of tenure holdings, quality of land, water availability, asset endowment, gender inequality, and land grabbing. The calculation of the index would be based on confirmatory factor and latent class analyses, which allow us to rank countries from high to low inequality respectively, according to the indicators selected, and to compile groups of countries with intra- and inter-group variance to assess the degree of contribution of each of these variables to land inequality.

REFERENCES

Albertus, M., Brambor, T., and Ceneviva, R. (2016). Land inequality and rural unrest: Theory and evidence from Brazil. SAGE Journals, University of Chicago. doi: 10.1177/0022002716654970

Arbuckle, J.L. (1996). Full information estimation in the presence of incomplete data. In G.A. Marcoulides and R.E. Schumacker (eds). Advanced Structural Equation Modeling: Issues and Techniques (pp.243-277). Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.

Bharti, N.K. (2019). Local Level Land Inequality in India, pp.1-31. PSE, University of Namur.

Biemer, P.P., Groves, R.M., Lyberg, L.E., Mthiowetz, N.A., and Sudman, S. (eds) (2013). Measurement Errors in Surveys. New Jersey: John Wiley & Sons.

Brown, T.A. (2015). Confirmatory Factor Analysis for Applied Research. The Guilford Press.

Cipollina, M., Cuffaro, N., and D'Agostino, G. (2018). Land Inequality and Economic Growth: A Meta-Analysis. Sustainability MDPI, 1-20. doi: 10.3390/su10124655

Cortés, F., Rubalcava, R.M., and Fernández, T. (2014). Estadística Social Básica (second edition). Montevideo, Uruguay: Universidad de la República.

Cortés, F. (2020). A cerca del coeficiente de Gini. Personal correspondence with Delfino Vargas.

Cullis, J. and van Koppen, B. (2007). Applying the Gini Coefficient to Measure Inequality of Water Use in the Olifants River Water Management Area, South Africa. Colombo, Sri Lanka: IWMI Research Report 113.

De Janvry, A. and Sadoulet, E. (2016). Development Economics: Theory and Practice. New York: Routledge.

Deininger, K. and Squire, L. (1998). New ways of looking at old issues: inequality and growth. Journal of Development Economics, Vol. 57, pp.259-287.

FAO (2013). 2000 World Census of Agriculture: Analysis and International Comparison of the Results (1996–2005). Rome: Food and Agriculture Organization of the United Nations.

Gini, C. (1935). Curso de Estadística. Barcelona: Editorial Labor.

Hagenaars, J.A. and McCutcheon, A.L. (2002). Applied Latent Class Analysis. Cambridge, UK: Cambridge University Press.

Heshmati, A. (2004). Inequalities and their Measurement. MTT Economic Research and IZA Bonn, IZA Discussion Paper No. 1219, 1-17.

Lora, E. and Prada, S.I. (2016). Técnicas de medición económica, Metodología y Aplicaciones en Colombia (5th edition). Cali, Colombia: Universidad Icesi.

Martinelli, P. (2012). Latifundia Revisited. Market Power, Land Inequality and Efficiency in Interwar Italian Agriculture. EHES Working Papers in Economic History, No. 20, 1-60.

Marshall, A. (1920). Principles of Political Economy. London.

Merlet, M. (2020). Les régulations des marchés fonciers et de l'usage de la terre: des outils pour réduire les inégalités. Rome: ILC, Land Inequality Initiative.

Muthén, B. (2002). Beyond SEM: General latent variable modeling. Behaviometrika, 29(1), 81-117.

Nunnally, J.C. and Bernstein, I.H. (1994). Psychometric Theory (third edition). New York: McGraw Hill.

Piketty, T. (2020). Capital and Ideology (trans. A. Goldhammer). Harvard University Press.

Scalise, E. (2020). Land Inequalities: Assessing and Measuring the Gender Gap. Rome: ILC, Land Inequality Initiative.

Stiglitz, J. (1974). Incentives and Risk Sharing in Sharecropping. Review of Economic Studies, Vol. 41(2), pp.219-255.

Vellvé, R. and Rakotondrainibe, M. (2018). The Daewoo-Madagascar land grab: ten years on. GRAIN. *http://www.thisisplace.org/i/?id=ad4dd531-1500-4589-bdb7-2ba721d4*

Vermunt, J.K. and Magidson, J. (2002). Latent class cluster analysis. In J.A. Hagenaars and A.L. MacCutcheon (eds). Applied Latent Class Analysis (pp.89-106). Cambridge, UK: Cambridge University Press.

Wedel, M. and DeSabro, W.S. (2002). Mixture regression models. In J.A. Hagenaars and A.L. McCutheon (eds). Applied Latent Class Analysis. Cambridge, UK: Cambridge University Press.

Wegerif, M. and Anseeuw, W. (2020). Unearthing the less visible trends in land inequality. Rome: ILC, Land Inequality Initiative.

PAPERS IN THIS RANGE

Series jointly edited by ILC and Oxfam

SYNTHESIS REPORT

» **Uneven ground: land inequality at the heart of unequal societies**
Anseeuw, W. and Baldinelli, G.M.

CASE STUDIES

» **¿Puede la concentración de la tierra ser fuente de desarrollo? Un análisis de las condiciones y bienestar de trabajadores agroindustriales de la provincia de Virú, Perú**
Araujo Raurau, A.L.

» **Assessing and measuring the gender gap in land rights under communal land associations in Karamoja**
Lakidi Achan, P.

» **Desigualdades en el acceso a la tierra y la inserción laboral de los nicaraguenses en la agricultura de Costa Rica**
Baumeister E.

» **Global financial funds, land grabs, and the (re)production of inequalities: a contribution from Brazil**
Kato, K., Furtado, F., Junior, O.A. and Siviero, J.

» **How the Talaandigs regained their ancestral lands in the Kalatungan mountain range**
Ravanera, R., Verdijo, T.C., and Gualberto, X.M.E.

» **La agricultura bimodal en el sector soyero: desentrañando la coexistencia entre pequeños y grandes productores en el oriente de Bolivia**
Colque, G. and Mamani, M.I.

» **La tierra entre la palma y las flores en los montes de María y el Oriente Antioqueño, Colombia**
Espinosa Rincón, N. and Jaramillo Gómez, O.E.

CONCEPTUAL PAPERS

- **La gobernanza colectiva de la tierra y su contribución a la reducción de las desigualdades**
  Bautista Durán, R. and Bazoberry Chali, O.
- **Assessing and measuring the gender gap**
  Scalise, E.
- **Unearthing less visible trends in land inequality**
  Wegerif, M. and Anseeuw, W.

DATA PAPERS

- **Global land inequality**
  Bauluz, L., Govind, Y., and Novokmet, F.
- **Methodological considerations on land inequality**
  Vargas, D. and Luiselli, C.

SOLUTION PAPERS

- **Acceso a la tierra en el Ecuador a partir de sus políticas públicas más recientes**
  Montesdoeca Chulde, D. and Ramos Bayas, M.
- **Building inclusive food chains: pathways beyond land inequality through collective action**
  Oberlack, C., Zambrino, L.A., Truong, Q.C; Dang, B.T, Vu, X.V., Blare, T.
- **Stratégies pour garantir un meilleur respect de l'environnement et des droits de l'homme par les investisseurs et entreprises**
  Nguiffo, S.
- **Régulation des marchés fonciers et de l'utilisation de la terre : des outils pour réduire les inégalités**
  Merlet, M.
- **Una oportunidad real: impuestos a la tierra para reducir desigualdad**
  Itriago, D.

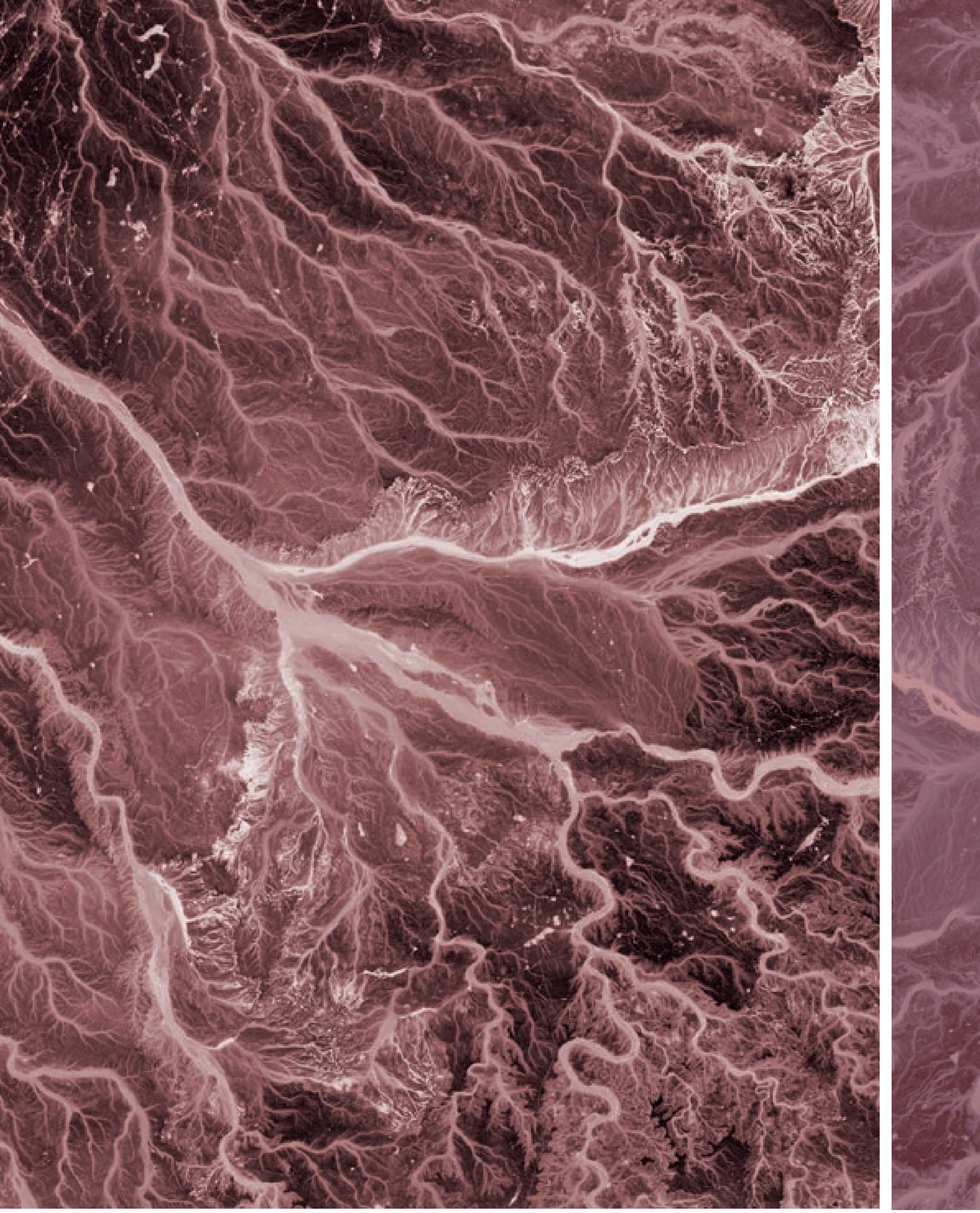
NOTES

NOTES

Land Inequality Initiative

is steered by an informal reference group, composed of experts in the field of land and wider inequalities.

Members of the reference group did provide guidance and expertise throughout the process and include the following organisations:



INTERNATIONAL LAND COALITION SECRETARIAT

c/o IFAD
Via Paolo di Dono, 44 ,
00142–Rome, Italy

tel +39 06 5459 2445
fax +39 06 5459 3445

info@landcoalition.org
www.landcoalition.org