UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE ESTUDIOS ESTADÍSTICOS

TESIS DOCTORAL

Data science applied to refining socio-economic indicators for decision-making

Data science aplicada al refinado de indicadores socio-económicos para la toma de decisiones

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

Antonio M. Salcedo Galiano

Directores

Lorenzo Escot Mangas
Gregorio Izquierdo Llanes

Madrid

FACULTAD DE ESTUDIOS ESTADÍSTICOS

UNIVERSIDAD
COMPLUTENSE
MADRID

TESIS DOCTORAL

Data science applied to refining socio-economic indicators for decision-making.

Data science aplicada al refinado de indicadores socio-económicos para la toma de decisiones.

Autor

ANTONIO M. SALCEDO GALIANO

Directores

LORENZO ESCOT MANGAS

GREGORIO IZQUIERDO LLANES

Madrid, 2019

UNIVERSIDAD COMPLUTENSE MADRID

DECLARACIÓN DE AUTORÍA Y ORIGINALIDAD DE LA TESIS PRESENTADA PARA OBTENER EL TÍTULO DE DOCTOR

D./Dña. Antonio - Maurilio Salcedo Galiano,
estudiante en el Programa de Doctorado de Análisis de Datos (Data Science),
de la Facultad de Estudios Estadísticos de la Universidad Complutense de Madrid, como autor/a de la tesis presentada para la obtención del título de Doctor y titulada:

Data science applied to refining socio-economic indicators for decision-making.

y dirigida por: Prof. Dr. Lorenzo Escot Mangas y Prof. Dr. Gregorio Izquierdo Llanes

DECLARO QUE:

La tesis es una obra original que no infringe los derechos de propiedad intelectual ni los derechos de propiedad industrial u otros, de acuerdo con el ordenamiento jurídico vigente, en particular, la Ley de Propiedad Intelectual (R.D. legislativo 1/1996, de 12 de abril, por el que se aprueba el texto refundido de la Ley de Propiedad Intelectual, modificado por la Ley 2/2019, de 1 de marzo, regularizando, aclarando y armonizando las disposiciones legales vigentes sobre la materia), en particular, las disposiciones referidas al derecho de cita.

Del mismo modo, asumo frente a la Universidad cualquier responsabilidad que pudiera derivarse de la autoría o falta de originalidad del contenido de la tesis presentada de conformidad con el ordenamiento jurídico vigente.

En Madrid, a 11 de mayo de 2019

Fdo.: SALCEDO GALIANO ANTONIO MAURILIO - 52381422A

Firmado digitalmente por SALCEDO GALIANO ANTONIO MAURILIO - 52381422A
Fecha: 2019.05.11 09:14:03 +02'00'

Esta DECLARACIÓN DE AUTORÍA Y ORIGINALIDAD debe ser insertada en la primera página de la tesis presentada para la obtención del título de Doctor.

Acknowledgements

"A prædicta civitate
castrum quoddam Montlælii
non multis millibus distat,
oppositione sua satis jocundum."
Historie de Gigny (1158)

This thesis is dedicated to my family.

First of all I would like to thank my soul mate, my dear wife, without you I would not have enough motivation and strength to keep walking, nor would I have been able to do all that I have managed to do, particularly this research work.

I would also like to thank my dear children, you have been one of the keys to the achievement of this work, accompanying me on many occasions in this long journey full of difficulties and personal efforts.

I'm thankful for my beloved parents and My sister, you are always the light that guides me at all times, my reference, thank you for being there for me every single day.

It has been a real pleasure to return to the Complutense University of Madrid, where I studied Mathematical Sciences not long since, learned how to live and made very extraordinary friends.

Thanks to the Faculty of Statistical Studies for hosting me to undertake this third cycle programme. I feel proud of being admitted in this really innovative initiative that, in my view, is extremely useful to complement other PhD programmes as well as filling an existing gap in statistical research.

I am also very grateful to my Directors, Prof. Dr. Lorenzo Escot and Prof. Dr. Gregorio Izquierdo, your technical advice, help, friendship, and encouragement in the most difficult moments have been essential to put this thesis to the end, thank you very much for trusting and giving me this opportunity.

I would also like to thank the program coordinators, Prof. Dr. Cristóbal Pareja and Prof. Dr. Daniel Gómez, and all Professors and other staff that are participating in this project, sharing their knowledge and research work with the doctoral students, particularly during seminars and training activities included in this PhD programme.

Finally, it is a real privilege for me to work at the National Statistics Institute of Spain and to be able to dedicate my efforts to the development of official statistics, which represent a factor of enormous importance in a democratic society.

It is there where I've had the opportunity to work in several different statistical topics, including the coordination of regional statistics as RESCO, then I had the chance to learn of the existence of a French town called Montluel, whose name is really similar to my dear hometown, Montiel, the place where the famous knight Don Quixote of La Mancha mounted his steed Rocinante and began to walk just only a few centuries ago.

That intriguing coincidence has sparked my curiosity so, at this stage, I hope I could apply in the future the data science techniques to the case, *i*ntroducing the issue; performing the appropriate *m*ethods and models; estimating the *r*esults to find out new empirical evidence from the data; *a*nd *d*iscussing and summing up the conclusions ...

... but that would be part of another research,
for another time ....

Contents

List of Tables

List of Figures

Acronyms

AROP	At-risk-of-poverty rate.
AROPE	At-risk-of-poverty or social exclusion rate.
COICOP	Classification of individual consumption by purpose
EUROSTAT	Statistical Office of the EU.
GDP	Gross domestic product.
ICW	Income, comsumption and wealth.
LWI	Low work intensity.
OECD	Organisation for Economic Co-operation and Development.
NUTS	Nomenclature of territorial units for statistics
PPP/PPS	Purchasing Power Parities/Purchasing Power Standard.
ROC	Receiver operating characteristic.
SMD	Severe material deprivation.
UN	United Nations.
UNECE	United Nations Economic Commission for Europe.

Chapter 1

INTRODUCTION.

1.1 Introduction to the topic.

Indicators of social and economic well-being constitute a key input to the European Pillar of Social Rights. From a general perspective, socio-economic well-being indicators cover a broad spectrum of thematic areas among which are the following ones: income; material deprivation; employment (having a job and quality of that job); education and training; health; housing; access to childcare; entitlement to leave for caring reasons; decent replacement incomes (retirement, invalidity, unemployment); a safe environment: a healthy environment (free of air pollution, noise, etc.); an absence of discrimination; access to justice. Harmonization of the production of these indicators has been a subject of significant development in the last few years, particularly after the adoption of a series of regulations, that include in detail a set of technical specifications, harmonized and mandatory for all EU countries[1]:

- Council Regulation (EC) on the organisation of a labour force sample survey in the Community.
- Regulation (EC) No 1177/2003 concerning Community statistics on income and living conditions (EU-SILC).
- Regulation (EC) No 452/2008 concerning the production and development of statistics on education and lifelong learning.
- Regulation (EC) No 1338/2008 on Community statistics on public health and health and safety at work.
- Regulation (EC) No 808/2004 concerning Community statistics on the information society.

[1]The future framework regulation that integrates social statistics (IESS) is expected to replace those regulations and to introduce other areas covered so far by means of informal agreements, such as the household budget survey or the European Union time use survey

As shown in Figure 1.1, statistical production can be represented by means of a pyramid in which the data is located at the base, the accounting systems in the middle, and the indicators at the top. Thus, in the European Statistical Program 2013-2017 the following three definitions have been adopted.[2]:

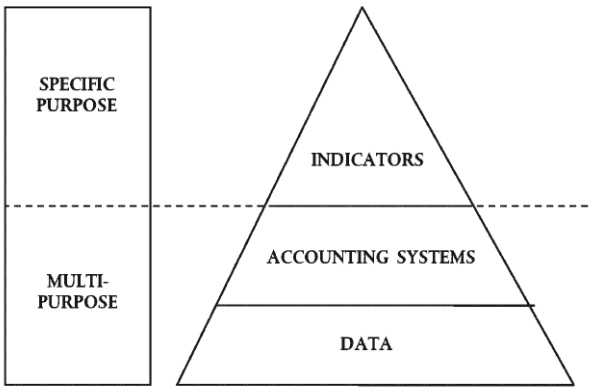


Figure 1.1: Statistical information infrastructure.

(Source: European statistical programme 2013–2017)

- "Data: information compiled by national statistical authorities, on the basis of traditional statistical activities (sample surveys, censuses, etc.) and data from other sources that are reused for statistical purposes. This information is tailored to serve needs in specific policy areas, e.g. the labour market, migration or agriculture. The term also includes data collected for administrative purposes but used by national statistical authorities for statistical purposes (usually referred to as data from administrative sources).
- Accounting systems: coherent and integrated accounts, balance sheets and tables based on a set of internationally agreed rules. An accounting framework ensures a high profile of consistency and comparability; statistical data can be compiled and presented in a format that is designed for the purposes of analysis and policy-making.
- Indicators: an indicator is a summary measure related to a key issue or phenomenon and derived from a series of observed facts. Indicators can be used to reveal relative positions or show positive or negative change. Indicators are usually a direct input into Union and global policies. In strategic policy fields they are important for setting targets and monitoring their achievement."

[2]Regulation (EU) No 99/2013 of the European Parliament and the Council on the European statistical programme 2013–2017

In official statistics we can find several types of socio-economic indicators that can be constructed from the data, accounting systems or even other indicators. When facing the construction of new socio-economic indicators, it must be borne in mind that they must offer a right empirical vision to the reality that they (themselves) represent. Thus, Eurostat has proposed a classification of indicators according to different approaches to their potential use, as presented in Table 1.1:

Why do we use indicators?	- Performance indicators. - Descriptive indicators.
What do we use indicators to measure?	- Leading, coincident and lagging indicators. - Driving force, pressure, state, impact and response indicators. - Input, output, outcome and impact indicators. - First, second, third level and contextual indicators. - Individual, composite and synthetic indicators.
How objectively and directly do indicators represent reality?	- Direct and proxy indicator. - Qualitative and quantitative indicators. - Objective and subjective indicators.

Table 1.1: Different ways of classifying indicators.

(Source: Eurostat - Towards a harmonised methodology for statistical indicators - Part I)

The measurement of income and consumption by households/individuals is a key aspect of socio-economic statistics. It has been the subject of study by numerous researchers of recognized prestige. At this stage we can refer as an example to the report made by J. Stiglitz, A. Sen and J.P. Fitoussi (" Measurement of Economic Performance and Social Progress - 2009"), that gave a significant boost to produce new social statistics and indicators. In this report it was particularly recommended to "consider income and consumption jointly with wealth" and to "give more prominence to the distribution of income, consumption and wealth". In general, international organisations and statistical offices have worked in the development and implementation of different multidimensional statistical approaches that could offer a set of indicators that are situated "beyond GDP".

In particular, poverty indicators are close related to income and consumption distribution. Historically, some authors point the origin of the poverty studies in the work undertaken by Booth and Rowntree by the end of the 19th century. However, there are earlier historical references to statistics on poverty; for example, in the Population Census of the Crown of Castile in 1752 where it was investigated the number of the so-called "solemn poor", in reference to the extremely poor who came to beg on solemn feasts and who, in general, were forced to live from begging.

When measuring poverty, the classical approach introduces a first distinction between the objective and subjective measurement of poverty. Thus, from an objective point of view, the researcher is the one who concludes whether or not a household/individual is poor, while, from a subjective point of view, the informant is who gives the answer about his or her poverty situation.

Objective measurements can, in turn, be classified into indirect, which are usually based on a low level -absolute or relative- of resources from the perspective of income or consumption and reflect a lack of resources that may cause deprivation, and direct, which are based on the living conditions of the household/individuals and can be related to a more multidimensional approach to the poverty.

In this way, there are three main approaches that dominate the scholarly literature today. The first one, which is often based on consumption or family budget surveys and is common in North America, considers the necessary income to live with an acceptable standard of living or a provision of a minimum set of necessary products (goods or services). The second one is applied in Europe and

also in OECD countries, where a relative income poverty measure is used to define the population at-risk-of poverty, that is, the people situated on the left side of the distribution of the variable "equivalent income" normally using a percentage (60% by Eurostat; 50% by OECD and UNECE) of the median as a threshold. The third one, unlike the other two, can be considered as a less indirect measure of poverty and based on non-monetary results, and it tackles the problem of the poverty measurement by studying and analysing the deprivation or lack of consumer goods and services that are necessary to enjoy an acceptable standard of living.

In 2017, in the framework of the Conference of European Statisticians, a manual was adopted with the aim to improve the international comparability of poverty statistics. Table 1.2 summarizes the different measurements currently applied. Besides, the 2030 Agenda adopted by the General Assembly of the United Nations in 2015 has provided momentum on a global scale for the compilation of socio-economic indicators for sustainable development, particularly about poverty, to monitor the degree of compliance with the 17 objectives and 169 goals that make up this global action plan.

Unidimensional	Monetary	Income based	Absolute poverty lines	National thresholds specific for individual countries, in the national currency	1. Cost of basic needs
					2. Subsistence minimum
				Internationally comparable thresholds	3. Severely poor with income below 1.9 PPP$
					4. "Just poor" with income below 3.1 PPP$
			Relative poverty lines	Share of the median (or mean) income	5. Relative low income (example: below 50% or 60% of the contemporary median equivalised income in each country)
					6. Relative low income anchored at a fixed point in time
					7. Weakly relative poverty line

(Cont.)

Unidimensional	Monetary	Expenditure based	Absolute poverty lines	National thresholds specific for individual countries, in the national currency	8. Cost of basic needs
					9. Subsistence minimum
				Internationally comparable thresholds	10. Severely poor with income below 1.9 PPP$
					11. "Just poor" with income below 3.1 PPP$
			Relative poverty lines	Share of the median (or mean) income	12. Relative low expenditure (example: below 50% or 60% of the contemporary median equivalised expenditure in each country
					13. Relative low expenditure anchored at a fixed point in time
					14. Weakly relative poverty line
	Food energy intake (FEI)				15. Nationally specific FEI - based poverty rates

(cont.)

Multidimensional	Deprivations	16. Indicator dashboards
		17. Indices of multiple deprivation, including material deprivation
	Multidimensional poverty estimates	18. Multidimensional poverty index (thresholds for the various dimensions)
	Official national multidimensional poverty indices	19. Severely poor
		20. Moderately poor

Table 1.2: Different approaches to poverty measurement.

(Source: UNECE - Guide on Poverty Measurement.)

At this point it is necessary to recall that the indicators related to well-being have been developed sometimes under a moderate availability of harmonized empirical data. However, in recent years, the implementation of different surveys in the European area, particularly in the case of EU-SILC but also the HBS, has offered researchers the possibility of performing new harmonized and comparable information treatments, which in turn it's an interesting opportunity to refine -whenever possible- the existing classical models.

1.2 Objectives and thesis structure.

The general objective of this doctoral research is to apply the data science to try to improve the construction of socio-economic indicators, putting the attention on one of the aspects with a greater degree of complexity: The current paradigm of objective measurement of poverty -direct and indirect-. Thus, it is intended to seek and to propose improvements that allow a more adequate reconciliation between the methods and results of the different indicators, especially taking into account the different opinions expressed on the current degree of linkage between the data derived from severe material deprivation and low monetary resources available to households/individuals.

In this way, through this research, the following specific issues are addressed:

- Equivalization: This technique allows to transform an economic variable from households to individuals giving a different weighting to the members of a household. It's usually applied to income but also other variables can be subject of an equivalization process, e.g., consumption. There is a wide range of equivalence scales, many of which were reviewed by Atkinson in 1995, but the most commonly used are: OECD equivalence scale or Oxford scale (1982), OECD-modified scale (1994) and Square root scale (2011).

- Thresholds: The so-called at-risk-of monetary poverty rate can be considered as an objective, indirect and unidimensional indicator based on a concept of economic distance, which is understood as a percentage p of the median equivalent income per adult. In the European Unions it's calculated by placing the percentage p at 60% in all EU countries, but the United Nations and the OECD recommend the use of a 50% of the median of equivalent income. The fact that such prestigious institutions recommend, by convention, different

values for p or thresholds in relation to the median of equivalent income to determine the risk of poverty rate is a clear indication that these percentages may vary depending on the reality of the situation under study.

- Multidimensional approach: In the European Union poverty has been measured indirectly in a one-dimensional way from a perspective based on disposable income, but this classical approach has certain limitations when representing such a complex phenomenon by means of a single indicator, producing sometimes modest results with regard to other direct poverty measurements such as severe material deprivation. It has been suggested that a multidimmensional framework based on Income-Consumption-Wealth (ICW) might provide a possible solution since it would integrate the three variables clearly linked to poverty.

The main data sources used in this doctoral research are the Living Conditions Survey (LCS) and the Household Budget Survey (HBS) anonymized microdata sets.

The Living Conditions Survey is a statistical activity equivalent to the community statistics on income and living conditions of the European Union (EU-SILC). This survey is carried out in Spain since 2004 and provides information on income and social exclusion, both transversal and longitudinal because it is a panel survey. The LCS is based on a rotating (annual) panel design consisting of four independent sub-samples. For each sub-sample a two-stage design is applied, with stratification in the first-stage units (census sections). The sample consists of about 13,000 households and 35,000 people.

The Household Budget Survey is a statistical activity of great tradition in the Spanish official statistics. Thus, the INE undertook this study for the first time six decades ago, in 1958. This survey estimates the annual consumption expenditure of private households and its distribution in accordance with the classification of goods and services ECOICOP (European Classification of Individual Consumption according to Purpose). The EPF is based on a two-stages sampling with stratification in the first stage units (census sections) while in the second stage (housing) no subsampling is done, which implies a research of some 24,000 households.

This doctoral work is structured around 8 chapters.

First, Chapter 1 offers an introduction to this work, presenting the main objectives, content and structure of this document.

Chapter 2 addresses the problem of equivalence scales when determining the variable "at risk of monetary poverty rate", which is an indirect measure of poverty, and its relation to the variable " population rate in severe material deprivation ", which is a direct measure. The existence of a very unequal panorama regarding the availability of data between the different Member States is confirmed, measured in terms of the Nomenclature of Statistical Territorial Units (NUTS). Thus, in some cases, such as Germany or France, there is not any regional data or, in other cases, they offer a degree of association measured in terms of correlation coefficients close to zero (Netherlands, 0.1) or even negative (Sweden, -0.3). Then, a model is proposed to generalize the OECD equivalence scales, which are constant by convention in all the countries in order to determine the households equivalence consumption units. As a result, it is shown that equivalence scales are not constant but vary over time, and the application of the new model substantially improves both the estimation of the expenditure of the consumption units by type of household and the link between the two previous poverty variables, what is particularly significant in the case of its application by autonomous communities. This chapter is published in *Social Indicators Research*, July 2018, Volume 138, Issue 2, pp 623–638, https://doi.org/10.1007/s11205-017-1675-4 .

In Chapter 3 we study an optimal cut-off point selection method based on the results of the variable "population rate in severe material shortage". This issue arises from the fact that different international organizations (Eurostat and OECD / UNECE) recommend, by convention, different percentages p of the median of the equivalent income (p=60% and p=50%) to determine the monetary poverty thresholds as an indirect measure of poverty. From the analysis of the results by countries, first we can see that in some cases the results of the risk of poverty are lower than those in material deprivation and, second, that the value p that best adjusts the distance to the material deprivation rate could vary from one year to the other, either increasing or decreasing. Based on the application of ROC curves, an estimation of the optimal cut-off point is proposed and it's tested internally, in terms of its sensitivity, specificity and accuracy, and externally, in terms of the different degrees of correlation. Additionally, the degree of linkage and elasticity with other exogenous variables, such as unemployment, is assessed. In conclusion, it is observed that it is feasible to apply statistical methods to microdata sets that

allow us to determine optimum cut points that can vary from year to year, in one direction or another, depending on the direct observation of the severe material deprivation. This chapter is published in *Quality & Quantity* January 2019, Volume 53, Issue 1, pp 513–527, https://doi.org/10.1007/s11135-018-0766-x .

In Chapter 4 a study is carried out on the classifier used in the determination of the indirect measurement of poverty, starting from an analysis of the issue from a multidimensional perspective. The issue arises from the UNECE recommendations to measure poverty, which considers as valid the poverty measurements both from the perspective of the variable "income" and the variable "consumption". In the analysis a solution based on the multidimensional paradigm Income-Consumption-Wealth (ICW), seen from a perspective of complementarity rather than substitutability, is proposed. Based on the low sensitivity observed in the poverty risk indicator in some European countries, first a two-dimensional extension, income-consumption, of the poverty risk indicator is carried out, what allows to reveal a new typology depending on the situation of the income of households / individuals according to the different thresholds. To conclude the study, a new model is proposed to introduce the wealth variable derived from housing tenure. The new indicator offers better results in both the temporal (years) and spatial dimensions (autonomous communities) measured in terms of its link to severe material deprivation as well as depending on its sensitivity, specificity and effectiveness. This chapter has been sent to a *Social Indicators Research* and it's under revision to be published as a scientific article.

In Chapter 5 the conclusions of the doctoral thesis are offered, including some possible lines of research that may give continuity to the work done in this thesis. Chapter 6 presents an extension to the measures of poverty and social exclusion in the form of tables and graphs. Thus, the vision of the parameters of sensitivity and specificity as well as the intersections of the components of the indicator of poverty risk and social exclusion, ordered according to the wealth of each country in terms of GDP in PPS, constitutes an additional approach that is annexed to this work. References and bibliography mainly used to compile this thesis are included in Chapter 7. Lastly, in Chapter 8 the abstract of the thesis is presented in English, which is the working language used in this research, and also in Spanish according to the general requirements of the Complutense University of Madrid.

Chapter 2

EQUIVALIZATION REVIEW.

Abstract.

This chapter investigates the issue of measuring the proportion of the population affected by the phenomenon of poverty in Spain on the basis of the relationship between the severe material deprivation rate and the at-risk-of-poverty rate by regions. The current definitions of equivalence units and equivalent income were analysed and an innovation to allow the introduction of a parametric model, based on the revealed annual consumption of the main different type of households and on the regional purchase power parities, was subsequently proposed. In the case of Spain it was discovered empirically that the equivalence units are not constant, as maintained until now, but that they vary over time. An improvement between the relations of the two rates was obtained by means of the application of these revealed scales. The regional composition was also improved. The results could serve as a base for future studies to provide more detailed information about the breakdown of the relative poverty within each country and, thereby, to meet the relevant information requests, at national and international level, to support the implementation of public policies in this area.

Acknowledgement.

Antonio M. Salcedo & Gregorio Izquierdo, 2018. "An Empirical Approach to the Poverty Indicators Based on Revealed Parameters: The Case of Spain" Social Indicators Research: An International and Interdisciplinary Journal for Quality-of-Life Measurement, Springer, vol. 138(2), pages 623-638, July.
https://doi.org/10.1007/s11205-017-1675-4 .

2.1 Introduction.

The EU statistics on income and living conditions (EU-SILC) represent a valuable source of quantitative and qualitative information on income, poverty, social exclusion and other social conditions of households and individuals in the European Union. They are based on several international standards, in particular, Regulation (EC) No 1177/2003 of the European Parliament and of the Council of 16 June 2003 concerning Community statistics on income and living conditions (and subsequent regulations).

Thus, these statistics are one of the fundamental pillars that underpin decision-making in different policy areas. As an example, in the European Union we can find the initiative Europe 2020 strategy for smart, sustainable and inclusive growth, whose objectives include "fighting poverty and social exclusion: at least 20 million fewer people in or at risk of poverty and social exclusion" (European Commission, 2010). Besides, now from a global perspective, the data obtained would serve to provide high-quality and comparable information in the framework of the implementation of the Agenda 2030 for sustainable development (United Nations, 2015).

The EU-SILC offers two important estimators which, from different points of view, provide information about the ratio of the population φ affected by the phenomenon of poverty[1]:

[1]In December 1984 the following definition was agreed by the EU Council of Ministers: 'the poor' shall be taken to mean persons, families and groups of persons whose resources (material, cultural and social) are so limited as to exclude them from the minimum acceptable way of life in the Member States in which they live.

(a) The severe material deprivation rate (designated as T1 for the purposes of this article). It is a direct estimation of poverty and covers those households/persons declaring that they cannot afford or lack at least four of the nine key items specified in the questionnaires. These items are, in general, related to the consumption possibilities of families (Guio, A. et al., 2010). Therefore this variable reflects, in some way, the capacities and the use of the available resources or income of interviewees, in relation to the disposal of a minimum set of goods and services.

(b) The at-risk-of-poverty rate (denominated as T2 in this article). It represents a conventional estimate of "relative poverty$\hat{A}\acute{}\hat{A}\acute{}$ (Ravallion, M., 2011), including total income and different sizes of households. It should be borne in mind that, since this variable is based on total income, it is not only limited to the use of resources and consumption but also to household/individual savings during the reference period.

Given the fact that these two variables provide, one way or another, information about the population and the poverty of a country, some degree of relation between them would be expected from a statistical viewpoint.

Indeed, our review of the situation allows us to conclude, in general terms, the existence of a sufficient degree of correlation between the two rates according to most of the classification variables of EU-SILC. However, this is not the case with the analysis carried out by European regions, where there is a quite different picture regarding the availability of data from the Member States. This situation remains unusual taking into account the fact that the European regions constitute a fundamental pillar to the construction and design of public policies within the EU. In this manner, some authors have chosen to focus their work towards a multidimensional vision of poverty in the European regions (Weziak-Bialowolska et al., 2014).

A first review of the datasets on severe material deprivation (T_1) and at-risk-of-poverty (T_2) rates available on the Eurostat website, shows that in some cases -for example France or Germany- there is no accessible information on these rates by regions, according to levels 1 or 2 of the Nomenclature of Units for Territorial Statistics (NUTS) (Eurostat, 2015). The databases currently offer a complete set of statistical information about the two population rates covering around half of the Member States.

Moreover, when studying the available data by European regions in more depth, it is observed that, sometimes, the correlation coefficients between both estimators would not be satisfactory enough and, therefore, they could be improved. As examples, in Spain in the year 2011, the Pearson correlation coefficient between these two poverty estimators is 0.3 at NUTS2 level; in The Netherlands, in 2014, it scores 0.1 at NUST1 level; or in Sweden, in 2015, it reaches -0.3 at NUTS2 level.

These results lead us to reflect on the current situation and to consider the existence of possible elements or factors, included in some international standard definitions which, in some way, might be influencing the results of the estimators negatively. Since material deprivation is a direct outcome of the survey, that is object of an adequate and timely editing and depuration made by the national statistical offices, at this stage it is hard to believe that this could be the origin of the problem and, consequently, we think we should turn our attention to the elements or factors which, in a conventional way, determine the at-risk-of-poverty rate.

Different approaches to the measurement of the poverty.

The measurement of poverty has been the subject of study by numerous researchers of recognized prestige, many of which point to its origin in the work undertaken by Booth and Rowntree by the end of the 19th century (Gillie, 2008).

A first distinction can be drawn between the objective and subjective measurement of poverty (Heikkilä et al., 2004). Thus, from an objective point of view, the researcher is the one who concludes whether or not a household/individual is poor, while, from a subjective point of view, the informant is who gives the answer about his or her poverty situation. Objective measurements can, in turn, be classified into indirect, which are usually based on a low level -absolute or relative- of resources from the perspective of income or consumption and reflect a lack of resources that may cause deprivation, and direct, which are based on the living conditions of the household/individuals and can be related to a more multidimensional approach to the poverty, since many authors and institutions consider that poverty is a complex phenomenon and it cannot be measured solely by taking into account a single variable of interest, for example, the income of the population (Alkire and Santos, 2013).

In this way, there are three main approaches that dominate the scholarly literature today (Mood, C. and Jonsson, J.O., 2016). The first one, which is often based on consumption or family budget surveys and is common in North America, considers the necessary income to live with an acceptable standard of living or a provision of a minimum set of necessary products (goods or services). The second one is applied in Europe and also in OECD countries, where a relative income poverty measure is used to define the population at-risk-of poverty, that is, the people situated on the left side of the distribution of the variable "equivalent income" normally using a percentage of the median as a threshold. The third one, unlike the other two, can be considered as a less indirect measure of poverty and based on non-monetary results, and it tackles the problem of the poverty measurement by studying and analysing the deprivation or lack of consumer goods and services that are necessary to enjoy an acceptable standard of living.

This latter direct approach has been used in different statistical surveys, for example in the EU-SILC and, thereby, it has been possible to offer a joint measurement of poverty from a direct and indirect point of view with the same statistical tool, which is fundamental for the purposes of this article and allows us to work with the estimates of material deprivation and relative income poverty.

Finally, in relation to the subjective measurements of the poverty, they are based on the perception that households / individuals have about their own living conditions and surrounding environment, by answering questions such as "is your household poor?". This approach allows us to gather the interpretation of those affected by the phenomenon of poverty, which in turn presents certain drawbacks derived from the own subjective perception and the different understanding of the concept of the poverty that people might have (Ravallion, M. et al., 2016).

Elements to estimate the at-risk-of-poverty rate from the perspective of income.

The current definition of the population at-risk-of-poverty includes all people whose equivalised disposable income is below the at-risk-of-poverty threshold, that

is, 60% of the median calculated at national level.

At this point, it should be noted that the equivalised disposable income is not a direct outcome of the questionnaires used to compile national surveys but is calculated in a conventional way taking into account, from one side, the household's total disposable income (numerator) and, from the other side, the so-called equivalent size of the household (denominator).

The total disposable income is a variable that is either declared by the respondents of the questionnaire or collected from administrative records, such as those available in tax agencies. Therefore, a priori, the results of this variable should be considered to be sufficiently contrasted so that they would not be expected to influence or misrepresent the final outcome in any way.

However, the calculation of the equivalent size of each household h is currently calculated by using an international convention (OECD, 2013), applied in European countries without taking into consideration their own structure or features. The modified equivalence scale of the OECD ($eqScale_{h,OECD}$) assigns to each household h a value of 1.0 to the first household member; 0.5 to the rest of members over 14 years old living in the household; and 0.3 to each child under the age of 14.

The equivalent scales constitute a good tool to compare income or expenditure between households of different sizes and compositions. These scales are based on the theories referring to the existence of scale economies and equivalent consumption (Buhmann, B. et al., 1988). The introduction of the equivalence scales has meant in practice, in the Member States of the European Union, the inclusion of a series of income/consumption parameters that are constant and unchanged over time and geographical areas or regions.

The equivalence scales and their use have been subject to detailed analysis in the related literature. The generally accepted interpretation can be found in the different European regulations on living conditions and in the methodological guidelines of the EU-SILC (Atkinson and Marlier, 2010) as well as the related analysis of material deprivation (Guio, A. et al., 2012). The robustness of the method at regional level in

terms of sampling errors has been studied by different authors (Verma et al., 2010). Some others have suggested possible further improvements showing, for example, that the poverty measures are sensitive to the choice of reference demographics, calling into question the reference to a single adult, as this is an untypical household (Ravallion, 2015). Poverty and time has also be the objective of analysis by other researchers (Bossert, W.a et al., 2012).

2.2 Data and methods.

Elements to calculate the equivalised income per household and individual

As presented in the previous section, in the European Union the at-risk-of poverty threshold ($Arpt_{p,t}$) in a concrete year t is determined, in general, by means of a percentage p (normally $p = 60\%$) of the Median[2] (Mdn) of the equivalised disposable income after social transfers in that year t based on the modified OECD equivalence scale ($eqInc_{OECD,t}$), according to the expression:

$$Arpt_{60,t} = 60\% * Mdn(eqInc_{OECD,t}) \tag{2.1}$$

In Europe the at-risk-of-poverty thresholds ($Arpt_{p,t}$) can be calculated, besides, in euros, in the national currency or in Purchasing Power Parities (Eurostat and OECD, 2012). The Purchasing Power Parities make it possible to convert the different national currencies into a "single" currency, controlling the level of market prices among Member States.

Coming back to the expression (2.1), the equivalised disposable income in a household h in the year t ($eqInc_{OECD,t}$) represents the total income of the household available for consumption or savings after tax and deductions ($Inc_{h,t}$), divided by the number de equalised adults in terms of the modified OECD equivalence scale ($eqScale_{h,OECD}$):

[2]The Mean (M) of the distribution could also be used instead of the Median (Mdn).

$$eqInc_{OECD,t} = \frac{Inc_{h,t}}{eqScale_{h,OECD}} \tag{2.2}$$

After that, and also in a conventional way, the ratio calculated in (2.2) is set not only to the household but also to each and every member of the household, regardless, for example, of their ages or relationship with economic activity.

Towards an adjustment of the equivalence scales by NUTS2 regions.

The use of equivalence scales in the framework of the measurement of poverty has been subject to several studies, as mentioned in the section 2.1 (introduction). The OECD modified equivalence scale ($eqScale_{h,OECD}$) of a household h made up of m members (aged fourteen years or older) and n members (aged thirteen years old or less) can be expressed as:

$$eqScale_{h,OECD} = (1 + 0.5 * (m - 1) + 0.3 * n) \tag{2.3}$$

This expression can be easily generalized to a parametric equivalence scale (revealed) in any year t as follows:

$$eqScale_{h,Rev,t} = (1 + a_{1,t} * (m - 1) + a_{2,t} * n)^{a_{3,t}} \tag{2.4}$$

for $0 \leq a_{i,t} \leq 1$, i=1,2,3.

This equation (2.4) could be interpreted as follows: the parameter $a_{1,t}$ would represent the number of equivalent adults of the second and successive adults with respect to the first adult in a household; the parameter $a_{2,t}$ would correspond to the number of equivalent adults of the children under 14 years old; and the parameter $a_{3,t}$ would mean, inversely, the economy scale of a household since, if $a_{3,t}$ is equal to 0, then there will be a "perfect" and constant scale economy for any household size

and composition but, if $a_{3,t}$ is equal to 1, there will not be any consumption scale in terms of the number of equivalent adults in a household. It should be noted that the revealed equivalence scale in those households that consists of a single adult (m=1 y n=0) is always equal to 1 (Betson and Muellbauer, 2004).

This parametric model generalises not only the OECD-modified scale but other scales by replacing the parameters $a_{1,t}$, $a_{2,t}$ and $a_{3,t}$ as shown in the following table 2.1.

$eqScale_{h,Rev,t}$	OECD equivalence scale 1982	OECD equivalence scale 1994	Square root scale 2008
$(1+a_{1,t}*(m-1)+a_{2,t}*n)^{a_{3,t}}$	$a_{1,t}$=0.7 $a_{2,t}$=0.5 $a_{3,t}$=1.0	$a_{1,t}$=0.5 $a_{2,t}$=0.3 $a_{3,t}$=1.0	$a_{1,t}$=1.0; $a_{2,t}$=1.0; $a_{3,t}$=0.5

Table 2.1: The OECD scales as a particular case of the parametric model.

To obtain an optimal solution of the equation (2.4) in a year t, we will make use of the anonymised microdata of the Household Budget Survey, published by the National Statistics Institute of Spain on their website, as an exogenous source of information providing annual data on the consumption expenses and patterns of Spanish households.

For any given year t, it would be possible to approximate the average consumption expenditure of the households h made up of m adults (aged 14 years old or older) and n children (aged 13 years old or under), to be denominated as $\bar{C}_{m,n,t}$ in this article, by multiplying the average consumption expenditure of the households made up of 1 adult and 0 children in the year t, abbreviated as $\bar{C}_{1,0,t}$ by the revealed equivalence scale ($eqScale_{h,Rev,t}$), in accordance with the following expression:

$$\bar{C}_{m,n,t} \cong \bar{C}_{1,0,t} * eqScale_{h,Rev,t} \tag{2.5}$$

Substituting (2.4) in (2.5) we obtain a system of non-lineal equations, one for each type of household. At this stage this system of equations can be easily solved by applying the algorithm Generalized Reduced Gradient (GRG) Nonlinear. A brief introduction to the use of this algorithm can be found in several articles (Walsh, S. and Diamond, D., 1995) as well as the application to several case studies (Roberts, G.P. et al., 2001). To obtain a better estimation and to avoid adding extra complexity to the problem, the non-lineal equation system will be restricted to the type of households that represents at least 1 per thousand of the total households in the year t (that is, $w_{m,n,t} > 1‰$). This election guarantees that the solution of the algorithm covers almost 100% of the total households. Thus, the objective function F to minimise using the previous algorithm is:

$$minF = min \sum_{m,n} w_{m,n,t}((1 + a_{1,t} * (m-1) + a_{2,t} * n)^{a_{3,t}} - \frac{\bar{C}_{m,n,t}}{\bar{C}_{1,0,t}})^2 \tag{2.6}$$

The solution of the system of non-lineal equations allows us to obtain the optimal values $\hat{a}_{1,t}$, $\hat{a}_{2,t}$ and $\hat{a}_{3,t}$ and, then, a solution to the expression (4) for a given year t to be used to calculate the number of equivalent adults of the complete set of households made up of m adults and n children.

Towards an adjustment of the equivalence scales by NUTS2 regions.

As mentioned previously in subsection 2.1, the at-risk-of-poverty threshold is calculated throughout the European Union at Member State level both in national currency and in purchasing parities. It's worth mentioning that these (national) "poverty" thresholds are not under the influence of the variability of the income in households of different countries, since the design and sampling of the surveys are independent and, in this regard, again as an example, the distribution of the equivalent income of Danish households does not influence the poverty line of Spanish households at all, and vice versa.

However, this is not the situation within each individual country at region level (NUTS). Indeed, the EU-SILC Regulation does not foresee the production of independent surveys by region within each Member State; this would require an important increase of the available resources of the National Statistics Institutes and also mean an increase in the response burden, so we cannot rule out that, within a country, the regions with higher income may influence those with lower incomes increasing the poverty rates of the latter and, vice versa, the regions with lower incomes may influence in the regions with higher income reducing the poverty rates, which would be particularly relevant in those countries with a large income gap among regions.

The situation described above could be solved by introducing regional purchasing power parities (PPP). It is feasible to estimate international poverty lines from comparable national thresholds by using purchasing power parity (Jolliffe and Prydz, 2015). The regional PPP approach has also been proposed by other authors in other economic areas, such as the proposal of a reform of a fiscal system taking into account the differences in governmental purchasing powers across the German states (Zimmer, 2015).

The equivalent income of individuals living in a household in the region r in the year t ($eqInc_{Rev,t,r}$) can be standardised by the use of PPP in a year t in the region r ()$PPP_{t,r}$) and, consequently, it would be determined as follows:

$$eqInc_{Rev,t,r} = \frac{Inc_{h,t}}{eqScale_{h,Rev,t} * PPP_{t,r}} \tag{2.7}$$

Finally, the at-risk-of-poverty threshold based on revealed parameters can be obtained by replacing in (1) the revealed equivalent income:

$$Arpt_{60,t} = 60\% * Mdn(eqInc_{Rev,t,r}) \tag{2.8}$$

2.3 Results.

Data sources and reference years.

The study presented in this paper has been conducted by using three main data sources:

The first one is the Life Conditions Survey ("Encuesta de Condiciones de Vida"[3], INE). This survey is based on the harmonised methodology and criteria established in the EU-SILC, and provides information on income distribution and social exclusion.

The second data source is the Household Budget Survey ("Encuesta de Presupuestos Familiares"[4], INE). This survey covers 24,000 households per year and provides information on the consumption expenses (including self-consumption), as well as some other features of the living conditions of Spanish households.

Finally, the third source of information is the paper on the regional purchasing parities presented in 2015 in the International Conference of Regional Science (Costa, A. et al., 2015). Here we can find an estimation of the purchasing power parity of the Autonomous Communities for the purpose of comparing economic indicators, such as regional GDP, following international recommendations.

[3] $http://www.ine.es/dyngs/INEbase/en/operacion.htm?c=Estadistica_C\&cid=1254736176807\&menu=ultiDatos\&idp=1254735976608$

[4] $http://www.ine.es/dyngs/INEbase/en/operacion.htm?c=Estadistica_C\&cid=1254736176806\&menu=ultiDatos\&idp=1254735976608$

Two reference periods are used in this research. One covers the period 2006-2015 and affects the calculation of the revealed equivalence scales. The other is the year 2012, for the variables regarding poverty. This year is considered one of the worst of the recent financial crisis in Spain so it would be expected that the 2012 data regarding material deprivation and at-risk-of-poverty should be very representative of the poverty proportion, both at national and regional levels.

The raw data used in this article can be downloaded from the INE website[5] where the Spanish statistical office offers, free of charge, to researchers and other advanced users, a wide range of microdata covering different statistical areas, in particular, about "Living conditions" as specified in that webpage.

Estimation of the revealed equivalence scales, period 2006-2015.

As a first step, due to the fact that average household consumption is not constant but varies yearly, we will begin by calculating the consumption units for the period 2006-2015, covered by the available anonymised microdata of the Household Budget Survey on the INE's website.

After solving the non-lineal equations system, the optimal solution obtained for the parameter $(\hat{a}_{1,t})$ gave the value 1.0 in all cases, which indicates that the equivalent consumption in households with more than one adult is not bound up with the consumption of the "first" adult (1.0) and subsequent adults (0.5), but that all adults contribute "equally" when calculating the equivalent consumption scale; this seems quite logical and does not presuppose any type of ' discrimination", based on the household member chosen as the "first adult" of the household.

The parameters $(\hat{a}_{2,t})$ (representing the children aged 13 and under) and $(\hat{a}_{3,t})$ (household scale economies) showed a gradually decline during the years of the financial crisis, but in the year 2015 this situation experienced a turning point

[5] $http://www.ine.es/en/prodyser/microdatos_en.htm$

and values started to increase. Even though these three parameters might be interpreted individually, we think an interpretation of the revealed equivalence scale as a whole, that is, the number of equivalent adults by type of household, is more appropriate.

Year t	$\hat{a}_{1,t}$	$\hat{a}_{2,t}$	$\hat{a}_{2,t}$	$eqScale_{h,Rev,t}$	$\bar{C}_{1,0,t}$
2006	1.00	0.66	0.69	$(m + 0.66n)^{0.69}$	16,725.66
2007	1.00	0.56	0.64	$(m + 0.56n)^{0.64}$	17,632.68
2008	1.00	0.45	0.62	$(m + 0.45n)^{0.62}$	18,502.70
2009	1.00	0.45	0.57	$(m + 0.45n)^{0.57}$	18,436.63
2010	1.00	0.48	0.59	$(m + 0.48n)^{0.59}$	17,758.53
2011	1.00	0.45	0.56	$(m + 0.45n)^{0.56}$	18,239.40
2012	1.00	0.46	0.54	$(m + 0.46n)^{0.54}$	17,959.84
2013	1.00	0.39	0.55	$(m + 0.39n)^{0.55}$	17,325.31
2014	1.00	0.36	0.54	$(m + 0.36n)^{0.54}$	17,435.98
2015	1.00	0.41	0.56	$(m + 0.41n)^{0.56}$	17,368.04

Table 2.2: Revealed equivalence scale by years (period 2006-2015).

The estimation of the revealed consumption units gives us a vision of the evolution of the consumption pattern by type of households in the period 2006-2015. In table 2.2 we can see the progression of the equivalent consumption units of three representative types of households composed of 2 adults; 2 adults and 1 child; and 2 adults and 2 children. It is worth pointing out that, if applying the modified OECD scale to the case, the number of equivalent adults would have remained unchanged in the values 1.5, 1.8 and 2.1 in all years 2006-2015, thus it is feasible to calculate the differences (overestimation or underestimation, as percentages) between both scales as shown in table 2.3.

Year	2 adults		2 adults and 1 child		2 adults and 1 children	
	Revealed Scale	Dif. OECD (%)	Revealed Scale	Dif. OECD (%)	Revealed Scale	Dif OECD (%)
2006	1.61	7.40	1.96	8.81	2.28	8.57
2007	1.56	4.09	1.83	1.58	2.08	-1.18
2008	1.53	2.27	1.74	-3.49	1.93	-8.27
2009	1.49	-0.87	1.67	-7.31	1.84	-12.57
2010	1.50	0.09	1.70	-5.44	1.89	-10.12
2011	1.47	-1.83	1.65	-8.29	1.81	-13.58
2012	1.46	-2.83	1.63	-9.40	1.79	-14.77
2013	1.46	-2.45	1.62	-10.26	1.76	-16.36
2014	1.46	-2.77	1.60	-11.36	1.72	-17.93
2015	1.48	-1.41	1.64	-8.79	1.79	-14.61

Table 2.3: Compared evolution of the consumption units of three representatives type of households (period 2006-2015).

(Note: Generated by the authors based on the outcome of the revealed equivalence scale by years. To keep a homogeneous comparison with the modified OECD scale, children are considered as those households members aged 13 years old or under.)

Finally, in the next table 2.4 we can see the impact of the equivalence scales when estimating the consumption expenditure in the twelve main types of households in the year 2012.

Type of household: Adults (A), Children (C)	Observed average consumption	Estimated consumption using the Rev. scale	Difer. (%)	Estimated consumption using the modified OECD scale	Difer. (%)
1A	17,960 €	17,960 €	0.00	17,960 €	0.00
1A1C	19,976 €	22,057 €	**10.42**	23,348 €	**16.88**
2A	26,962 €	26,186 €	-2.88	26,940 €	-0.08
2A1C	29,469 €	29,301 €	-0.57	32,328 €	**9.70**
3A	32,687 €	32,648 €	-0.12	35,920 €	**9.89**
2A2C	33,644 €	32,159 €	-4.41	37,716 €	**12.10**
3A1C	33,702 €	35,277 €	4.67	41,308 €	**22.57**
4A1C	35,976 €	40,503 €	**12.58**	50,288 €	**39.78**
2A3C	36,940 €	34,820 €	**-5.74**	43,104 €	**16.69**
3A2C	37,968 €	37,751 €	-0.57	46,696 €	**22.99**
4A	39,582 €	38,179 €	-3.54	44,900 €	**13.44**
5A	42,586 €	43,106 €	1.22	53,880 €	**26.52**

Table 2.4: Observed and estimated average consumption by type of household (year 2012).

The cases were the estimation is over 5% of the observed consumption, in absolute terms, are shown in bold. We can see that the revealed scale fits more accurately with the real results obtained using the Household Budget Survey. The OECD modified scale, which remains constant over time, overestimates the average consumption in the most cases, particularly when the size of the households is larger, reaching 39.78% above to the average of households composed by four adults

and one child.

Review of the main results by using the revealed equivalence scales, year 2012.

The following table 2.5 presents some key results from the Life Conditions Survey in Spain after applying the two scales:

Main indicators	OECD modified scale	Revealed scale
At risk of poverty threshold ($Euros*$)	8,114	9,328
At risk of poverty rate (T_2)	20.38 %	19.96 %
Severe material deprivation rate (T_1)	6.19 %	6.19 %
Intersection $T_1 \cap T_2$ as a % of the severe material deprivation (T1)	56.04 %	57.65 %
Gini coefficient	33.69	32.52
S80/S20 income quintile share ratio	6.29	5.90

Table 2.5: Comparison of some key indicators by scale (year 2012)

The first consequence of the introduction of the new scale was an increase of the poverty line from 8,114 euros to 9,328 euros, that is, 14.96 % more. However, the at risk of poverty rate decreased by 0.43 percentage points. In addition, it should be noted that the intersection between the severe material deprivation population and the population at risk of poverty is 1.61 percentage points higher, which represents an improvement in the predictive capacity of the model.

The next figure 2.1 shows the intersection between the two rates of population. The area A (3.57%) represents the Spanish population rate in severe material deprivation and, simultaneously, with a low equivalent income (that is, at-risk-of-poverty). The area B (2.62%) reveals the population under severe material deprivation but with an equivalent income above the poverty threshold. The area C (16.38%) is the estimation of the population with an equivalent income below the poverty threshold but without declaring severe material deprivation. Finally, the area D (77.43%) constitutes the percentage of the population without severe material deprivation and with an equivalent income above the poverty threshold.

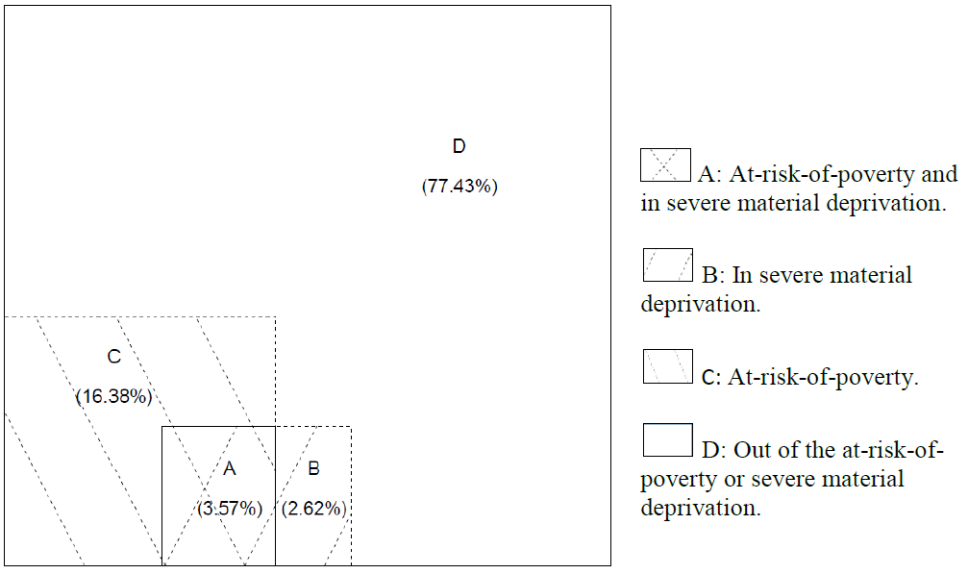


Figure 2.1: SMD and the at risk of poverty rates, year 2012

The composition of the at-risk-of-poverty rate by regions and the relation with severe material deprivation, year 2012.

To conclude this section 3 of Results, we should now examine the at-risk-of-poverty rate by regions and its relation with the population rate affected by severe material deprivation.

What stands out in Table 2.6 is a breakdown of the population rates under risk of poverty using both equivalence scales in the year 2012. In addition it also shows the population rate with severe material deprivation. The introduction

of the revealed scale, adjusted by regional purchasing power parities, changes the composition of the poverty rates by regions significantly. In this way, the at-risk-of-poverty rate of the Comunidad de Madrid, with the highest PPP (114.5), increases by +6.2 percentage points. On the contrary, the at-risk-of-poverty rate for Extremadura, which has the lowest PPP (80.3), reduces by -10.5 percentage points. The obtained outcome is more in line with the severe material deprivation observed. Moreover, we can see a lower dispersion rate of the poverty rates, which is more consistent with the assumption of mobility of persons within the same country. The use of revealed parameters, including purchasing parities, introduces changes into the composition, increasing the intersection of estimators T_1 and T_2 in the regions with higher income, particularly in Madrid and Barcelona.

Autonomous Communities (NUTS2)	Population with severe material deprivation	Population at-risk-of-poverty			PPP
		Revealed scale	Modified OECD scale	Dif. (absolute values)	
Navarra	0.61	12.9	9.9	+3.0	110.6
Castilla y León	1.81	15.5	17.5	-2.0	88.0
Aragón	2.50	16.8	16.1	+0.7	96.4
Extremadura	3.72	19.4	30.9	-10.5	80.3
Cantabria	3.88	17.2	17.8	-0.6	99.1
Asturias	4.21	13.5	14.1	-0.6	87.9
Galicia	4.61	15.4	17.2	-1.8	92.4
País Vasco	4.78	12.8	10.5	+2.3	107.7
Castilla-La Mancha	5.81	22.8	31.3	-8.5	84.8
Cataluña	6.09	18.2	13.9	+4.3	108.5
Rioja (La)	6.56	17.1	19.3	-2.2	90.4
Andalucía	6.98	25.8	29.1	-3.3	92.7
Comunidad Valenciana	7.07	21.4	23.6	-2.2	93.0
Comunidad de Madrid	7.28	19.6	13.4	+6.2	114.5
Canarias	8.50	21.9	28.4	-6.5	83.1
Murcia	8.99	22.5	26.8	-4.3	94.8
Balears, Illes	10.42	22.2	19.8	+2.4	98.9

Table 2.6: Population in severe material deprivation rate and at-risk-of-poverty by scales (%). Year 2012

The next figure 2.2 show the scatter plot of both rates after applying the two scales. Again, we can observe the effect produced after the introduction of the revealed parameters in the at-risk-of-poverty rate by regions. The new scale offers, without any doubt, more consistent results. Besides, it also provides a better explanation of the situation of severe material deprivation by regions, with a marked increase of the coefficient of determination, R^2, giving the variance of the dependent variable that can be "predicted" from the independent variable.

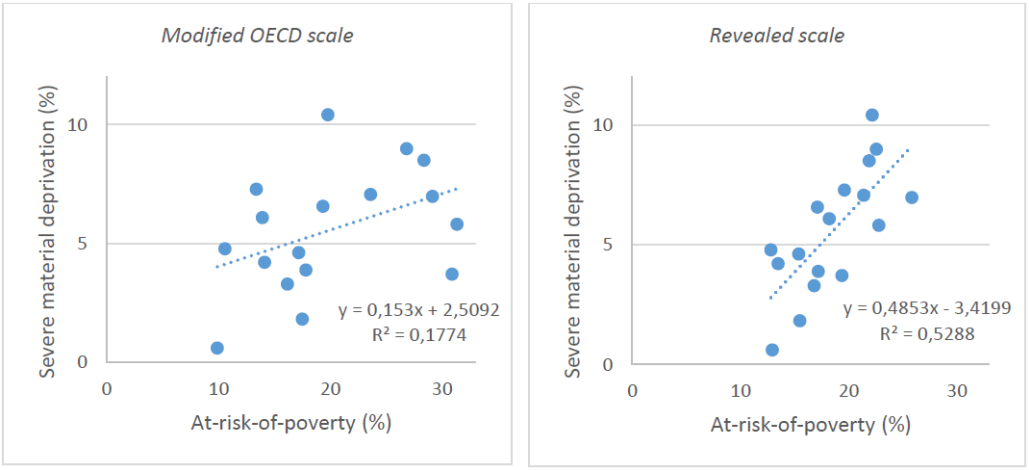


Figure 2.2: At-risk-of-poverty and severe material deprivation by regions, year 2012

This result is consistent with the positive correlation test carried out between both set of variables. Table 2.7 shows the result of the p-value of the hypothesis test. After applying the revealed scale, the hypothesis of lack of correlation between the severe material deprivation rate and the at-risk-of-poverty rate, by region of residence, is rejected.

	H_0 : ∄ correlation between T_1 and T_2 (Modif. Scale OECD)	H_0 : ∄ correlation between T_1 and T_2 (Rev. scale)
Pearson	p-value = 0.04590	p-value = 0.00047
Spearman	p-value = 0.07644	p-value = 0.00123
Kendall	p-value = 0.11810	p-value = 0.00225

Table 2.7: Results of the correlation tests.

2.4 Conclusions.

The aim of the present research was to analyse the issue of the low relation levels between the severe material deprivation rate and the at-risk-of-poverty rate using a regional approach. A review of the main features of standardised definitions was carried out to identify elements or factors that could be influencing the results at regional level.

Assuming that severe material deprivation, as an observed variable in the national surveys, is subject to an adequate and timely editing and depuration by the national statistical offices, in this article attention was focused on the variable equivalent income, which determines the population at-risk-of-poverty in a conventional way.

A parametric model was proposed in equation (2.4) that generalised not only the OECD modified scale, obtained by replacing the parameters $a_{1,t}$ by 0.5, $a_{2,t}$ by 0.3, and $a_{3,t}$ by 1 for any year t, but also other scales that have been proposed in this context.

This study has shown empirically that, in Spain, the equivalence scales aren't constant but they vary over time. Hence, the revealed equivalence scale fits the observed average consumption of the different type of households much better, which is particularly meaningful in those households with a larger size in terms of number of members (adults and children).

Regarding the evolution of the equivalent units, it was observed that the application of the OECD-modified scale has underestimated the equivalent consumption units in Spain in the years before the financial crisis, while it has overestimated the numbers of equivalent adults in the years of the financial crisis (and now).

The model proposed in this article for Spain takes into consideration not only the changes of the consumption patterns over time but also the variability of purchasing power parities in the different regions which, without any doubt, influences the results, with the effect being greater as the variability of the parities increases. In this way, the introduced model revealed a set of parameters based on the average consumption pattern of the households and on an adjustment of the equivalent income by means of the purchasing parities.

The application of the new scales led to an increase of a 15% in the at-risk-of-poverty threshold of the year 2012 (i.e., 1,214 euros more) which in turn involved a reduction of the at-risk-of-poverty rate of 0.42 percentage points (that is, 193.2 thousand of people). The form of the individual equivalent income distribution predetermines the poverty line since, by convention, this line is defined as a percentage of a percentile (the median), which is basically a measure of position of a distribution.

It is also remarkable that, by using the revealed equivalence scales, the intersection between the estimators of severe material deprivation and at-risk-of-poverty (T_1 and T_2) increased by 1.61 percentage points (that is, 45.8 thousand of people) and, consequently, the predictive capacity of the variable T_2, obtained from the income side, with respect to the severe material deprivation T_1, obtained from the consumption side, is increased.

Concerning the data by regions, the use of revealed parameters improved the outcome obtained with the OECD-modified scale substantially. The application of the parametric model partially changed the composition of the population at-risk-of-poverty, raising the rate in those regions with higher purchasing power parities and reducing the percentage in those regions with lower purchasing parities, showing more coherent results in line with the results obtained on severe material deprivation.

The proportion of the independent variable variance explained by the dependent variable (R^2) tripled from 0.1774 to 0.5288. In addition, the three independence tests (Pearson, Spearman and Kendall) gave a p-value lower than 0.05 and, therefore, the hypothesis of lack of correlation between the severe material deprivation rate and the at-risk-of-poverty rate by region of residence is rejected.

Notwithstanding the good results obtained, this paper presents several improvement opportunities and also limitations that influence the final outcome.

The election of the parametric model is itself a limitation of the study. In this sense, in this article we introduce, as a constraint, the use of a model that covers the models previously used in this framework. However, it is clear that the distinction between "children" (aged 13 or under) and "adults" (aged 14 or over) is a convention that limits the final outcome. The introduction of more parameters distinguishing between different age groups in a household (i.e, from 0 to 6 years old; 7 to 15; 16 to 24; 25 to 64; and over 64) would probably have returned different results.

Searching for a feasible solution to the parametric model of equivalence scales by means of the Generalized Reduced Gradient (GRG) Nonlinear is also another limitation of the study, since it might be possible to find a different optimal solution by applying another algorithm. Nevertheless, this was not the aim of this study; our objective was to use a reasonable method to prove that there is at least a solution of the parametric model that clearly improves the current OECD scales using an empirical method instead of a constant and invariant approach applied to European countries by convention.

Another limitation of this study involves the use of regional purchasing power parities. In the future, the performance of a harmonised method to calculate these regional PPPs would allow the production of comparable information by regions (and years) at European level. As an alternative, the definition of an agreed proxy in index form would make it possible to produce an adjustment of these rates to the regional reality.

An additional challenge to extend this model to other countries lies in the fact that the Household Budget Survey is not fully harmonised across European countries. Without any doubt it is crucial to continue working towards the harmonisation of this important survey that is not only of great value for this study but also to compile and produce other key official statistics such as the Consumer Price Index.

This research provides new insights on the estimation of poverty and could serve as a base for future studies to provide more detailed information about the incidence of this phenomenon in the population. From a statistical viewpoint, it would be of great value to manage a harmonised definition of poverty, which could be applicable across countries from the side of income and/or consumption, so that in case of continuing working with point estimations it would be possible to propose estimators based on the best statistical properties.

The methods used may be applied to other countries with annual household budget surveys and estimations of their regional PPPs. As an alternative, an increase of the surveys' sampling sizes and the introduction of a stratification at regional level could help to provide more accurate estimations. There is a clear need to compile accurate regional rates but, in any case, any set of indicators should be built on high-quality statistics and in line with the international standards, to ensure the comparability of the information produced and to meet the important requirements of information about the phenomenon of poverty.

Chapter 3

THRESHOLD ESTIMATION.

Abstract.

This study incorporates revealed severe material deprivation in the determination over time of the at-risk-of monetary poverty line from the perspective of income. Receiver operating curves taking the prevalence of severe material deprivation are used to calculate the optimal threshold. In this way, two poverty perspectives which were previously seen as independent, and to some extent mutually exclusive, are brought together. A harmonized procedure to determine the percentage p in relation to the median is applied; this percentage determines whether or not a person is at-risk-of poverty -50% in the UN and OECD but 60% in the EU. In these two cases the percentage is constant over the time by convention. Unlike the classical approach, the study identifies different thresholds for each reality (space and time), that are checked firstly using criteria for sensitivity, specificity, accuracy and the predictive values, secondly, by correlation between the new indicator and severe material deprivation, and finally, using an exogenous contrast variable.

Acknowledgement.

This is a post-peer-review, pre-copyedit version of an article published in *Quality & Quantity*. The final authenticated version is available online at: https://doi.org/10.1007/s11135-018-0766-x.

Salcedo, Antonio & Izquierdo Llanes, Gregorio. (2018). Drawing the optimal monetary poverty lines based on empirical data: an application to Spain. Quality & Quantity. 10.1007/s11135-018-0766-x.

3.1 Introduction.

Poverty is a phenomenon which has significant economic and social consequences and, therefore, it is vital to have access to statistical information of the highest quality to support the decision-making process and the evaluation of public policies. Poverty is recognized as a global phenomenon (World Bank, 2015) and, consequently, the international statistical community is dedicating an increasing amount of attention to the development of techniques and methodologies that allow us to measure poverty in a comparable and harmonized way. A specific case of this type of initiative is to be found in the framework of the 2030 Agenda for Sustainable Development of the United Nations with its seventeen goals to transform the world (United Nations, 2017), for which all countries have agreed to draw up a list of more than two hundred indicators.

In this context, the European Union's statistics office (Eurostat) and the statistics offices of the member states have worked together closely to produce harmonized statistics and indicators for poverty. Particularly noteworthy is the project aimed at designing a yearly survey to address this issue and offer a high degree of harmonization and comparability: the Statistics on income and living conditions in the European Union (EU-SILC). This survey has multiple objectives which include information on income, social exclusion, housing conditions, labour, education and health of households and individuals in the European Union (Atkinson et al., 2010). Among the results relating to poverty, we can highlight the delimitation of both the group in a situation of severe material deprivation and that at-risk-of poverty, the latter being defined by convention as those individuals whose equivalent income is below 60% of the median. It should be noted that the EU-SILC database can also be used to consider other possible poverty risks based on 70%, 50% and 40% of the median.

From a statistical point of view it is to be expected that both measures of poverty (severe material deprivation and risk of poverty) would offer a satisfactory degree of association and comparability, both in terms of their spatial dimension (countries or regions) and in terms of their temporal dimension (years). Table 3.1 shows that, in the European Union, the correlation between severe material deprivation and the risk of poverty ranges from 0.57 for p = 70% to 0.68 for p = 40%. These are certainly significant correlations but they gloss over situations which are not easily explained; for example, the fact that with p = 60% (the EU standard figure) the country with the highest material deprivation (Bulgaria) has a lower risk of poverty than other countries with a lower level of material deprivation, such as Romania, Latvia, Lithuania or Spain. On the other hand, the country with the lowest severe material deprivation (Sweden) has a greater risk of poverty than other countries with higher levels of severe material deprivation, such as Slovakia, Slovenia, the Czech Republic, France, Denmark, Austria, the Netherlands or Finland.

Having arrived at the point at which the comparability of the information is based on establishing a common value for p for all countries, it seems reasonable to ask whether it would be better to consider harmonizing the method used to determine the best p year by year in relation to the monetary income determined from the perspective of revealed material deprivation. To this end, a first approximation would involve determining a value of p for each country based on optimizing the difference between the risk of poverty obtained and the level of severe material deprivation of that country.

We can see that the value p = 70% offers a poverty risk rate closer to the severe material deprivation rate in two countries (Bulgaria and Hungary); on the other hand, the value p = 60% is closer to the rate of severe material deprivation in three countries (Romania, Greece and Cyprus); the value p = 50% is a better match in nine countries (Latvia, Lithuania, Croatia, Italy, Slovakia, Malta, Ireland, Belgium and the Czech Republic); and, finally, when p = 40% the value provides a closer estimate in fourteen countries (Portugal, Poland, Spain, the United Kingdom, Slovenia, Estonia, France, Germany, Denmark, Austria, the Netherlands, Finland, Luxembourg and Sweden).

	Smd	p=70%		p=60%		p=50%		p=40%			
		Armp	Dif.	Armp	Dif.	Armp	Dif.	Armp	Dif.	p_{min}	Armp
	(%)	(%)	(p.p.)	(%)	(p.p.)	(%)	(p.p.)	(%)	(p.p.)		(%)
BG	34.2	28.8	-5.4	22.0	-12.2	15.5	-18.7	10.0	-24.2	70%	28.8
RO	22.7	31.6	+8.9	25.4	+2.7	19.8	-2.9	14.5	-8.2	60%	25.4
EL	22.2	27.9	+5.7	21.4	-0.8	15.0	-7.2	10.1	-12.1	60%	21.4
HU	19.4	22.9	+3.5	14.9	-4.5	9.0	-10.4	4.5	-14.9	70%	22.9
LV	16.4	30.4	+14.0	22.5	+6.1	14.7	-1.7	8.9	-7.5	50%	14.7
CY	15.4	25.6	+10.2	16.2	+0.8	9.0	-6.4	3.9	-11.5	60%	16.2
LT	13.9	29.9	+16.0	22.2	+8.3	14.4	+0.5	7.8	-6.1	50%	14.4
HR	13.7	26.9	+13.2	20.0	+6.3	13.5	-0.2	7.9	-5.8	50%	13.5
IT	11.5	27.4	+15.9	19.9	+8.4	13.4	+1.9	9.0	-2.5	50%	13.4
PT	9.6	27.0	+17.4	19.5	+9.9	13.8	+4.2	8.5	-1.1	40%	8.5
SK	9.0	18.8	+9.8	12.3	+3.3	8.4	-0.6	5.7	-3.3	50%	8.4
MT	8.1	25.7	+17.6	16.3	+8.2	8.5	+0.4	3.0	-5.1	50%	8.5
PL	8.1	24.8	+16.7	17.6	+9.5	10.7	+2.6	6.3	-1.8	40%	6.3
IE	7.5	25.6	+18.1	16.3	+8.8	8.8	+1.3	3.3	-4.2	50%	8.8
ES	6.4	29.2	+22.8	22.1	+15.7	15.9	+9.5	11.2	+4.8	40%	11.2

(cont.)

	Smd	p=70%		p=60%		p=50%		p=40%			
		Armp	Dif.	Armp	Dif.	Armp	Dif.	Armp	Dif.	p_{min}	Armp
	(%)	(%)	(p.p.)	(%)	(p.p.)	(%)	(p.p.)	(%)	(p.p.)		(%)
UK	6.1	25.0	+18.9	16.7	+10.6	9.7	+3.6	4.8	-1.3	40%	4.8
BE	5.8	23.8	+18.0	14.9	+9.1	7.8	+2.0	3.4	-2.4	50%	7.8
SI	5.8	21.2	+15.4	14.3	+8.5	8.4	+2.6	3.9	-1.9	40%	3.9
CZ	5.6	17.1	+11.5	9.7	+4.1	5.3	-0.3	2.6	-3.0	50%	2.6
EE	4.5	28.9	+24.4	21.6	+17.1	12.5	+8.0	7.4	+2.9	40%	7.4
FR	4.5	21.6	+17.1	13.6	+9.1	6.5	+2.0	2.8	-1.7	40%	2.8
DE	4.4	24.5	+20.1	16.7	+12.3	10.2	+5.8	5.0	+0.6	40%	5.0
DK	3.7	20.5	+16.8	12.2	+8.5	7.1	+3.4	4.3	+0.6	40%	4.3
AT	3.6	21.8	+18.2	13.9	+10.3	8.3	+4.7	3.8	+0.2	40%	3.8
NL	2.6	20.3	+17.7	11.6	+9.0	5.8	+3.2	3.0	+0.4	40%	3.0
FI	2.2	21.3	+19.1	12.4	+10.2	5.3	+3.1	2.3	+0.1	40%	2.3
LU	2.0	23.9	+21.9	15.3	+13.3	8.2	+6.2	3.9	+1.9	40%	3.9
SE	0.7	22.7	+22.0	14.5	+13.8	8.0	+7.3	4.1	+3.4	40%	4.1

Table 3.1: Severe material deprivation (Smd) and at-risk-of monetary poverty (Armp) for different percentages p for EU countries: 2015.

In a complementary way, we can extend the previous cross-sectional comparative analysis by incorporating the evolution of the indicators over time. For the sake of simplicity we will limit the analysis to two specific countries whose size allows us to identify them as representative; at the same time they have experienced clear variations in their levels of severe material deprivation in recent years and that, in addition, these differences have been of opposite signs: in Poland it has reduced while in Italy it has increased.

Thus, in the case of Poland (table 3.2) we observe that there has been a clear fall in the level of severe material deprivation over the years, so that the percentage p which minimizes the difference between the two rates would have decreased from p = 70% in 2007 to p = 40% in 2015. In contrast, Italy (table 3.3) saw an increase in the population suffering severe material deprivation between 2007 and 2015, which would imply that the value of p should have increased from p = 40% to p = 50% for the population at risk of monetary poverty to have followed suit. Once again, the values of the correlation coefficient for pmin provide the best results in terms of the Pearson correlation.

Poland	Smd	Armp	Dif.	Armp	Dif.	Armp	Dif.	Armp	Dif.	p_{min}	Armp
	(%)	(%)	(p.p)	(%)	(p.p)	(%)	(p.p)	(%)	(p.p)		(%)
2015	8.1	24.8	+16.7	17.6	+9.5	10.7	+2.6	6.3	-1.8	40%	6.3
2014	10.4	24.8	+14.4	17.0	+6.6	10.7	+0.3	5.8	-4.6	50%	10.7
2013	11.9	24.9	+13.0	17.3	+5.4	10.8	-1.1	5.8	-6.1	50%	10.8
2012	13.5	24.9	+11.4	17.1	+3.6	10.5	-3.0	5.5	-8.0	50%	10.5
2011	13.0	25.7	+12.7	17.7	+4.7	10.5	-2.5	5.7	-7.3	50%	10.5
2010	14.2	25.5	+11.3	17.6	+3.4	10.5	-3.7	5.7	-8.5	60%	17.6
2009	15.0	24.8	+9.8	17.1	+2.1	10.4	-4.6	5.4	-9.6	60%	17.1
2008	17.7	25.0	+7.3	16.9	-0.8	10.2	-7.5	5.2	-12.5	60%	16.9
2007	22.3	25.2	+2.9	17.3	-5.0	11.1	-11.2	6.3	-16.0	70%	25.2

Table 3.2: Severe material deprivation (Smd) and risk of monetary poverty (Armp) in Poland: 2007-2015.

Italy	Smd	Armp	Dif.	Armp	Dif.	Armp	Dif.	Armp	Dif.	p_{min}	Armp
	(%)	(%)	(p.p)	(%)	(p.p)	(%)	(p.p)	(%)	(p.p)		(%)
2015	11.5	27.4	+15.9	19.9	+8.4	13.4	+1.9	9.0	-2.5	50%	13.4
2014	11.6	26.6	+15.0	19.4	+7.8	12.7	+1.1	8.7	-2.9	50%	12.7
2013	12.3	27.5	+15.2	19.3	+7.0	12.8	+0.5	8.5	-3.8	50%	12.8
2012	14.5	27.2	+12.7	19.5	+5.0	12.4	-2.1	7.9	-6.6	50%	12.4
2011	11.1	27.3	+16.2	19.8	+8.7	12.8	+1.7	8.5	-2.6	50%	12.8
2010	7.4	26.4	+19.0	18.7	+11.3	12.1	+4.7	7.3	-0.1	40%	7.3
2009	7.3	26.3	+19.0	18.4	+11.1	11.6	+4.3	6.9	-0.4	40%	6.9
2008	7.5	26.3	+18.8	18.9	+11.4	11.9	+4.4	6.8	-0.7	40%	6.8
2007	7.0	27.3	+20.3	19.5	+12.5	12.3	+5.3	6.9	-0.1	40%	6.9

Table 3.3: Severe material deprivation (Smd) and risk of monetary poverty (Armp) in Italy: 2007-2015.

The aforementioned situation is clearly seen in Figure 3.1. We can see how severe material deprivation has followed contrasting paths in Poland and Italy in recent years. As far as the at-risk-of-poverty rate is concerned, with p constant and equal to 60%, a level of around 17% is maintained continuously in the case of Poland and close to 19% in the case of Italy. The choice of a percentage p depending on the variable that minimizes the distance allows a more accurate result to be obtained.

This situation suggests that the choice of the percentage p of the median of equivalent income which determines the rate of the population at risk of poverty, and which until now has been determined in a conventional manner, always constant and invariable in space (countries) and over time (years), in practice has a clear impact on the coherence of the results obtained. Intuitively it would seem logical for there to be a connection between the two population rates (Fusco, A. et al., 2010), in our view meaning a certain amount of similarity and, simultaneously, a high degree of correlation. In fact, it would be desirable for the determination of

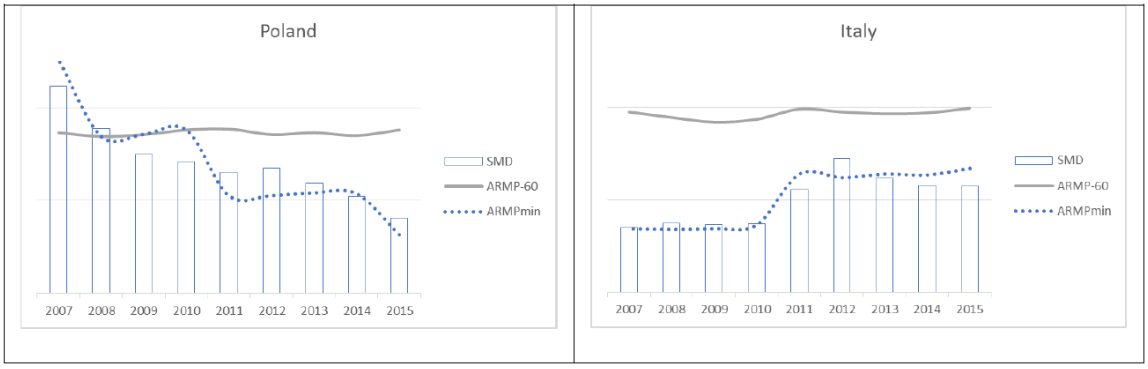


Figure 3.1: Rates of severe material deprivation (Smd) and risk of poverty (60% and below) for Poland and Italy: 2007-2015.

a rate of risk (in this case, the risk of poverty) to somehow be in tune with the prevalence of the measurement variables of direct poverty such as, in the case at hand, the rate of population in severe material deprivation in each country and each year. Building from such reflections, the objective of this article is to develop a statistical method which incorporates some innovations with respect to existing methodologies and makes it possible to calculate the optimal values of p for a given country and period of time as well as including the determination of severe material deprivation as a direct measure of poverty.

3.2 Literature review.

Poverty has been much studied throughout history. Many authors indicate the works of Booth and Rowntree in Victorian England (Gillie, 2008) as the origin of current poverty studies. There are, however, earlier historical references to statistics on poverty; for example, in the Population Census of the Crown of Castile in 1752 (INE, 1991) where, among other questions, those carrying out the study asked the mayors of the different Castilian municipalities about the number of the so-called "solemn poor", in reference to the extremely poor who came to beg on solemn feasts and who, in general, were forced to live from begging. The concept of poverty has evolved from initial notions of human subsistence, passing through a basic needs perspective as well as the use of different methods to calculate different lines of poverty, to a much broader approach, reaching a multidimensional vision of poverty (UNECE, 2017).

In summarized form and for exclusively statistical purposes, there are various different approaches to measuring poverty. An initial approach would allow us to distinguish between objective and subjective measures (Moisio, 2004). According to this approach, objective poverty indicates that it is the researcher who, after a detailed analysis of the information obtained, concludes whether or not an individual is in a situation of poverty, while subjective poverty refers to those situations in which the informant him/herself reports on his/her own situation of poverty. The subjective approach allows us to obtain a direct opinion from those who consider themselves to be in a state of poverty; although it should not be forgotten that individuals' perceptions of poverty are different and this will influence each person's comprehension of and way of responding about their true situation (Ravallion, M. et al., 2016).

A second approach would make it possible to differentiate between direct

and indirect measures of poverty (Lang and Lingnau, 2015). Using the direct measurement of poverty we can obtain information regarding whether households have succeeded in satisfying their basic needs; while with indirect measures we are concerned with evaluating the resources of households, mainly income and consumption, which are considered sufficient to attain an acceptable standard of living.

A third approach to the measurement of poverty would allow us to distinguish between unidimensional and multidimensional poverty. Measures of unidimensional poverty respond to the complex challenge of attempting to summarize the poverty situation of a specific group of people, for example, a country, in a single numerical variable. This numerical variable can then be broken down according to different classification variables, for example, sex, age, type of dwelling, etc. Unidimensional measures are usually approached from the perspective of income or consumption, often introducing absolute or relative thresholds (Moatsos, 2016) in order to classify individuals as poor or not poor. On the other hand, many authors are of the opinion that poverty is too complex to be summarized by means of a single economic variable (Battiston et al., 2013) and, consequently, it would be necessary to establish a multidimensional image providing much more comprehensive information than a simple unidimensional vision; this information could include other fundamental aspects such as health, education, work, etc. (Alkire, Sabina et al., 2015).

In the European Union, material deprivation is considered as an objective, direct and multidimensional measure of poverty by means of which the situation of people / households is studied with respect to the following nine items: not being able to afford a holiday of at least one week a year; not being able to afford to eat meat, chicken or fish at least every other day; not being able to afford to maintain an adequate temperature at home; not having the capacity to deal with unforeseen expenses (of 650 euros); having experienced delays in the payment of bills related to the main place of dwelling (mortgage or rent, gas receipts, community expenses ...) or hire purchase operations in the last 12 months; not being able to afford a car; not being able to afford a telephone; not being able to afford a television; and not being able to afford a washing machine. Severe material deprivation corresponds to a synthesis of material deprivation in which the threshold is established as those cases in which an individual reports suffering at least four of the nine previous situations.

The so-called at-risk-of monetary poverty, in turn, can be considered as an

objective, indirect and unidimensional measure of poverty, based on a concept of economic distance, and which is understood as a percentage p of the median (or average) equivalent income per adult. This at-risk-of-poverty indicator in the EU can be called a synthetic indicator (Eurostat, 2017a) and is calculated by placing the percentage p at 60% in all EU countries. However, as agreed by the United Nations and the OECD, this p should be set at 50% of the median of equivalent income. In order to transform income per household into equivalent income per adult it is necessary to weight the income of each household by the equivalent number of individuals; various scales have been proposed (OECD, 2013) and, by convention, these remain unchanged and invariable for all countries and years.

The fact that such prestigious institutions recommend different values for p or thresholds in relation to the median of equivalent income to determine the risk of poverty is a clear indication that these percentages may vary depending on the reality of the situation under study. In other words, as these values are clearly different for these three institutions, we can state that if these organizations have fixed their percentage p through some process of optimization, the optimal value for p can vary depending on the reality of the situation under study, which justifies the exploration of methodologies to determine this percentage.

3.3 Data and methods.

The problem addressed in this article can be formulated as shown in Figure 3.2. Let us consider that the distribution of the Equivalent income variable of the entire population of a country or region is given by the function $eqInc$, while the variable $eqIncSmd$ indicates the equivalent income of that part of the population in a situation of severe material deprivation. The selection of a cut-off point on the abscissa axis, $X = c$, automatically determines four clearly differentiated areas:

- A: The percentage of people at risk of monetary poverty (Armp+) but who are not in a situation of severe material deprivation (Smd-).
- B: The percentage of people who are at risk of monetary poverty (Armp+) and are also in a situation of severe material deprivation (Smd+).
- C: The percentage of people who are neither at risk of monetary poverty (Armp-) nor in a situation of severe material deprivation (Smd-).
- D: The percentage of people who are not at risk of monetary poverty (Armp-) but who are in a situation of severe material deprivation (Smd+).

-

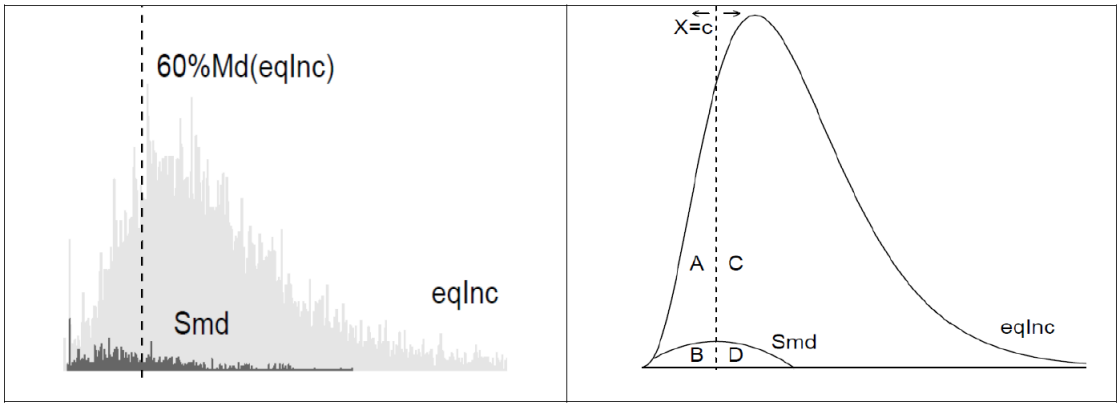


Figure 3.2: Histograms of sample frequencies obtained from the variables equivalent income and population in a situation of severe material deprivation. Year 2016.

Thus, the choice of a very high threshold c involves moving the cut-off line to the right, so that areas A and B would be larger while areas C and D would be smaller and, therefore, the risk of poverty would be higher but, on the other hand, this decision would include many people in the group at risk of poverty who are not in a situation of severe material deprivation. On the contrary, a lower c threshold would mean that areas C and D would be larger and, thus, the proportion of people at risk of poverty, whether or not they are in a situation of severe material deprivation, would be reduced.

Taking the previous approach into account, we can define the sensitivity (3.1) of the value of c as the probability that a person is at risk of monetary poverty on the condition that they are in a situation of severe material deprivation:

$$Se(c) = P(eqInc(x) \leq c \mid Smd_{+}) = P(Armp_{+} \mid Smd_{+}) = \frac{B}{B+D} \tag{3.1}$$

The specificity (3.2) of the value of c represents the probability that a person is not at risk of monetary poverty, conditioned to the fact that he / she is not in a situation of severe material deprivation, which can be expressed as follows:

$$Sp(c) = P(eqInc(x) > c \mid Smd_{-}) = P(Armp_{-} \mid Smd_{-}) = \frac{C}{A+C} \tag{3.2}$$

The accuracy (3.3) or global value of the risk of poverty is represented by the probability that a person is either in a situation of severe material deprivation and at risk of monetary poverty (simultaneously), or, in a situation of neither severe material deprivation nor at risk of poverty (simultaneously).

$$Accuracy = P(Smd_{+} \wedge Armp_{+}) + P(Smd_{-} \wedge Armp_{-}) = \frac{B+C}{A+B+C+D} \tag{3.3}$$

The positive predictive value (PPV) of the risk of monetary poverty represents the probability that a person at risk of monetary poverty will be in a situation of severe material deprivation:

$$PPV = P(Smd_{+} \mid Armp_{+}) = \frac{B}{A+B} \tag{3.4}$$

Finally, the negative predictive value (NPV) of the risk of monetary poverty represents the probability that a person who is not at risk of poverty will not be in a situation of severe material deprivation:

$$NPV = P(Smd_{-} \mid Armp_{-}) = \frac{C}{C+D} \tag{3.5}$$

The sensitivity and specificity, as defined above, are linked by the so-called ROC curve (Receiver Operating Characteristic). The ROC curve (figure 3.3) is defined by the pair (x,y) = (1-Sp(c), Se(c)) for all possible values of the threshold

c. Similarly, the PROC curve can be defined by the pair (x,y) = (1-VP-, VP+).

ROC curves have been used in many areas of science including signal detection, medicine and economics (Agarwal and Taffler, 2008). The area below the ROC curve takes values 0.5 to 1.0 and is an estimator of how good the fit is, equivalent to the Mann-Whitney test (Hand and Till, 2001); the closer it is to 1, the more accurate it is. Predictive values, meanwhile, are linked by the PROC (Predictive Receiver Operating Characteristic) curve (Shiu and Gatsonis, 2008).

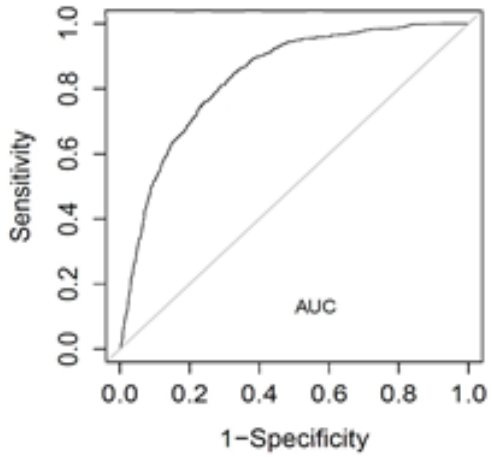


Figure 3.3: ROC curve and area under de curve (AUC) for severe material deprivation over at-risk-of monetary poverty in Spain. Year 2016.

At this point several different criteria can be proposed to determine the optimal value based on the proportion ϕ of people in a situation of severe material deprivation and the importance given to each of the four areas (A, B, C and D). For the purposes of this article, we have analyzed the results of the optimization method known as prevalence matching (López-Ratón et al., 2014) since this method makes it possible to incorporate severe material deprivation directly and use it to set a threshold C for which the predicted prevalence is practically equal to the observed prevalence, so that the optimal cut-off point is provided by the value of c which minimizes the following equation 3.6:

$$min|\ \varphi * (1 - Se(c)) - (1 - \varphi) * (1 - Sp(c))\ | \tag{3.6}$$

Once the optimal value of c has been obtained, the next step is to calculate the optimal percentage value for p of the median (Md) of equivalent income which, in turn, determines the poverty line, obtained by solving the following equation 3.7:

$$c = p * Md(eqInc(x)) \leftrightarrow p = \frac{c}{Md(eqInc(x))} \tag{3.7}$$

It should be noted that for the purposes of this article we have limited ourselves exclusively to the use of the median of equivalent income as this is the most commonly used value. However, it should be pointed out that it would be possible to replace the median (Md) with the mean (Me) in equation (7). In order to analyze the degree of association of the results obtained, the Pearson, Spearman and Kendall correlation coefficients will be used as complementary indicators as well as the classic indicators of point elasticity and arc elasticity, which provide us with a measure of the variability of one variable depending on another (Dür et al., 2014). Likewise, the prevalence matching (PM) method has been chosen due to the fact that adjusting the observed prevalence of risk of poverty by country against the severe material deprivation rate offered results which are more coherent than the current ones. It should be noted, however, that other methods could have been selected; for example, the minimax method which maximizes the areas with exactly the same estimate of poverty for the material deprivation variable and equivalent income, or another of the methods based on sensitivity, specificity or predictive values considered in the literature (Fawcett, 2006).

The validation of the results of the method applied will allow us to confirm the improvement with respect to the method currently followed. The existence of an alternative method to that selected which better optimizes the results cannot be ruled out. It should be pointed out, however, that our initial thesis was precisely to prove that a common and harmonized methodology to determine the percentage p, whether using the method proposed here or another, insofar as it includes severe material deprivation, will predetermine different values of p for each spatial and / or temporal reality. It is understood that the results are better if they support external contrasts with variables both endogenous and exogenous to the model, as discussed below for a specific case.

3.4 An application to Spain.

Main results.

Let us now apply the methodology introduced in the previous section to the specific case of Spain, using the Life Conditions Survey (LCS) prepared by the National Institute of Statistics as our data source. The LCS is completely harmonized and compatible with the EU-SILC; it also meets all the design and accuracy requirements of the European Union (Osier et al., 2013). It contains a complete set of anonymized, longitudinal and transversal microdata which can be freely consulted by researchers or other advanced users. Microdata can be downloaded directly from the INE website. This article is based specifically on transversal microdata for the period 2008-2016.

Table 3.4 presents the series of optimal cut-off points (thresholds) obtained using the prevalence matching (PM) method for the set of years 2008 to 2016. From these thresholds and considering the estimated population median, the optimal value of p can be estimated using equation (7), which, as previously noted, is calculated using material deprivation. It can be seen that the optimal value of the percentage p using PM is not constant but varies slightly from year to year. The minimum value of p was 0.25745, in 2011, while the maximum was 0.30466 in 2016.

Year	Threshold(*)	p=PM
2016	4.168 €	0,30466
2015	3.912 €	0,29299
2014	4.025 €	0,30334
2013	3.992 €	0,29518
2012	3.945 €	0,28447
2011	3.586 €	0,25745
2010	4.020 €	0,27525
2009	3.906 €	0,26401
2008	3.775 €	0,27030

Table 3.4: Threshold X=c and percentage p obtained using the Prevalence Matching method.

Table 3.5 shows the rate of population in severe material deprivation together with the main indicators obtained using the thresholds of the Eurostat database (p = 70%, p = 60%, p = 50% and p = 40%) and the optimal cut-off value obtained using PM.

The value p = 70% offers the highest threshold of all (9,577 euros per equivalent adult) and sets the rate of population at risk of monetary poverty at 29.9%, which is +24.1 percentage points higher than the Smd rate. In this case 4.6% of the population is both in a situation of severe material deprivation and at-risk-of poverty (area B), while 25.3% is at risk of poverty but is not in a situation of severe material deprivation (area A). 68.9% are neither at risk of poverty nor in a situation of severe material deprivation (area C), which is in fact the lowest figure of all the cases considered, although 1.2% claim that they are in a situation of severe material deprivation with income levels above those of the previous threshold (area D).

The value p = 60% lowers the threshold to €8,209 per equivalent adult and situates the risk of poverty rate at 22.3%, that is, +16.5 percentage points above the rate of population in a situation of severe material deprivation. 4.0% of the population (area B) is simultaneously at risk of poverty and in a situation of severe material deprivation, while 18.3% is at risk of poverty but is not affected by severe material deprivation (area A). 77.7% of the population is not at risk of poverty (areas C + D), although 1.8% of people above this figure state that they are in a situation of severe material deprivation.

The value p = 50%, meanwhile, establishes the risk rate of monetary poverty for 2016 at 15.5%, and is closer to the proportion of people in a situation of severe material deprivation, being +9.7 percentage points higher. The percentage of individuals who are both in a situation of severe material deprivation and at risk of poverty (3.5%) is reduced, as is the percentage of the population at risk of poverty but not affected by severe material deprivation (12%). The population located above the poverty risk threshold increases to 84.5%, although 2.3% are in a situation of severe material deprivation.

If p = 40% the population at risk of poverty rate is 10.7% which is +4.9 percentage points higher. Of these, 2.7% are simultaneously in a situation of severe material deprivation and at risk of poverty, while 8.0% are at risk of poverty but do not suffer severe material deprivation. The percentage of the population in a situation of severe material deprivation and without risk of poverty increases to 3.1%, while those who are neither in a situation of severe material deprivation nor at risk of poverty also rises (86.2%).

Finally, the optimal cut-off point for p obtained by PM sets the risk of monetary poverty rate for 2016 at 6.6%, which is very close to the proportion of the population in a situation of severe material deprivation (barely 0.8% higher). In this case only 1.7% of people are at risk of poverty and in a situation of severe material deprivation, but the proportion of the population without risk of either poverty or severe material deprivation reaches a high (89.3%).

	p=70%	p=60%	p=50%	p=40%	p=PM
Smd+	5,8%	5,8%	5,8%	5,8%	5,8%
Threshold(*)	9.577€	8.209€	6.840€	5.472€	4.168€
Armp+	29,9%	22,3%	15,5%	10,7%	6,6%
Dif. (p.p.)	+24.1	+16.5	+9.7	+4.9	+0.8
A (Armp+,Smd-)	25,3%	18,3%	12,0%	8,0%	4,9%
B (Armp+,Smd+)	4,6%	4,0%	3,5%	2,7%	1,7%
C (Armp-,Smd-)	68,9%	75,9%	82,2%	86,2%	89,3%
D (Armp-,Smd+)	1,2%	1,8%	2,3%	3,1%	4,1%
Total	100%	100%	100%	100%	100%
Se: B/(B+D)	0,79	0,70	0,60	0,47	0,29
Sp: C/(C+A)	0,73	0,81	0,87	0,92	0,95
Accuracy: (B+C)/100	0,73	0,80	0,86	0,89	0,91
PPV: (B/A+B)	0,15	0,18	0,22	0,26	0,27
NPV: (C/C+D)	0,98	0,98	0,97	0,97	0,96

Table 3.5: Main indicators for different levels of the parameter p: 2016.

(Note: (*) Euros per equivalent adult based on the OECD modified scale. Source: Prepared by the authors based on the LCS microdata.)

In Table 3.5 we can also observe the main indicators according to the different thresholds or cut-off points considered. The highest sensitivity is obtained when p = 70% although it includes the high figure of 25.3% of the population not in a situation of severe material deprivation. The highest specificity is obtained when p = PM, so that the model is able to classify correctly 95% of the individuals not affected by a situation of severe material deprivation. The highest accuracy, once

again, is obtained if p = PM, which means that in this case the model correctly classifies 91% of the total population. When p = PM, a higher positive predictive value is obtained as 27% (almost 1 in 3) of those at risk of poverty are in a situation of severe material deprivation. The negative predictive value is very high in all cases and always above 95%.

We can, therefore, confirm that the PM method, which establishes different values for p or thresholds in relation to the median, improves the indicator from the perspectives of specificity and effectiveness, maintaining a very high negative predictive value and is, therefore, satisfactory. We notice that the sensitivity of the new indicator is lower than the previous one, but this is inevitable since, as previously noted, in this particular case a high sensitivity can only be achieved by including as at risk of poverty many individuals who are not in a situation of severe material deprivation and, therefore, reducing specificity and accuracy. This specific situation may be indicative of the existence of other alternative variables that could be used to calculate the at-risk-of poverty.

Measures of association.

At this point we can also complement the validation of the results obtained by studying their consistency by assessing the different degrees of association between the severe material deprivation rate and the at-risk-of monetary poverty rate in Spain, calculated for all previous values of p, according to its evolution over the years.

The poverty risk rates calculated for higher percentage p returned greater levels of dissimilarity than the rate calculated using p = PM; the differences reached + 24.1 percentage points in the case of p = 70%, as seen in Table 3.6. The correlation coefficients of Pearson, Spearman and Kendall were, in all cases, higher if p = PM. This analysis suggests that the model introduced in this paper offers results which are more coherent over time than any of the other cases considered.

	Smd (%)	Armp (%)	Dif. (p.p)	Armp (%)	Dif. (p.p)	Armp (%)	Dif. (p.p)	Armp (%)	Dif. (p.p)	p_{min}	Armp (%)
2016	5,8	29,9	+24,1	22,3	+16,5	15,5	+9,7	10,7	+4,9	6,6	+0,8
2015	6,4	29,2	+22,8	22,1	+15,7	15,9	+9,5	11,2	+4,8	7,2	+0,8
2014	7,1	29,7	+22,6	22,2	+15,1	15,9	+8,8	10,6	+3,5	6,9	-0,2
2013	6,2	28,6	+22,4	20,4	+14,2	13,9	+7,7	9,3	+3,1	5,8	-0,4
2012	5,8	28,9	+23,1	20,8	+15,0	14,4	+8,6	9,6	+3,8	5,7	-0,1
2011	4,5	28,1	+23,6	20,6	+16,1	13,8	+9,3	8,7	+4,2	4,4	-0,1
2010	4,9	27,2	+22,3	20,7	+15,8	13,8	+8,9	8,8	+3,9	4,8	-0,1
2009	4,5	27,7	+23,2	20,4	+15,9	13,2	+8,7	8,1	+3,6	3,6	-0,9
2008	3,6	27,3	+23,7	19,8	+16,2	13,1	+9,5	7,4	+3,8	3,5	-0,1
Pearson		0.83		0.77		0.85		0.89		0.93	
Spearman		0.74		0.68		0.93		0.87		0.95	
Kendall		0.57		0.61		0.82		0.74		0.86	

Table 3.6: Rates of severe material deprivation and risk of monetary poverty by year.

Measures of poverty and the economic cycle.

Finally, to conclude this section, it would seem appropriate to test the consistency of the results obtained with an external and exogenous contrast variable such as, for example, unemployment, calculated using the Labour Force Survey (LFS) as a variable which is exogenous to the EU-SILC. Several authors have analyzed the relationship between poverty and unemployment, concluding that there is a positive correlation between unemployment, income inequity and poverty (Ukpere, Wilfred I. and Slabbert, Andre D., 2009). The risk of poverty has also

been directly related to unemployment through the at-risk-of poverty or social exclusion (AROPE) indicator (Eurostat, 2017b) by means of the households with low employment intensity.

The following figure shows the dispersion diagram of the at-risk-of-poverty variables calculated using p = 60% and p = PM (abscissas axis) against the unemployment rate (ordinal axis) by year. Although the time series does not contain many observations, Figure 3.4 shows that the use of Prevalence Matching returns more consistent results, with a clear increase in the R2 value from 0.17 to 0.44; this is in line with the high correlation coefficient between the rate of severe material deprivation and the unemployment rate.

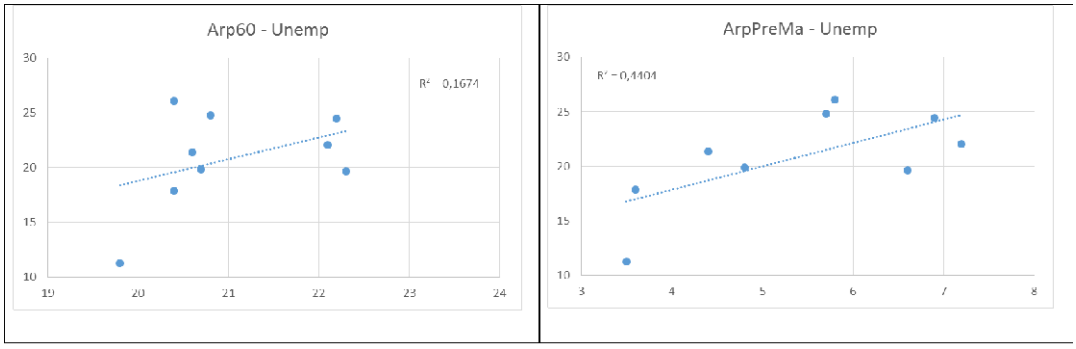


Figure 3.4: Dispersion diagrams for the at-risk-of-poverty rate with p=60% and p=PM against the unemployment rate in Spain: 2008-2016 series.

This result is also consistent with the positive correlation test carried out between the different at-risk-of-poverty rates and the severe material deprivation rate with the unemployment variable for the period 2008-2016. The Pearson correlation coefficients and the p-values obtained for the different poverty risk rates and for the Smd rate can be seen in Table 6. In the cases of Armp70, Armp60 and Armp50, the p-value obtained is greater than $\alpha = 0.05$, which leads us to reject the hypothesis that there is no correlation between the at-risk-of-poverty rates and unemployment, whereas for Armp40, ArmpPM and, especially, Smd the p-value obtained is below the significance level of 0.05 and would not allow us to reject the hypothesis of lack of correlation with the unemployment rate.

	p=70%	p=60%	p=50%	p=40%	p=PM	Smd
Pearson correlation	0.57	0.41	0.49	0.62	0.66	0.81
p-value (*)	0.0537	0.1371	0.0891	0.0366	0.0257	0.0038
Point elasticity (2008-2014)	0.07	0.10	0.18	0.37	0.83	0.83
Arc elasticity (2008-2014)	0.11	0.15	0.26	0.48	0.88	0.89

Table 3.7: Positive correlation contrast of the at-risk-of-poverty and severe material deprivation variables with the unemployment rate (*) and elasticities.

(Note: (*)$H_o : \nexists$ positive correlation with unemployment rate.)

In Table 3.7 above we can also observe the different degrees of elasticity obtained for the period 2008-2014. This period is considered given that comparable microdata for years prior to 2008 are not available and the recessive phase of the economic cycle in Spain ended in 2014, the year in which unemployment peaked at 24.4% compared to 11.25% in 2008. Thus, the rates calculated with p = 70%, p = 60% and p = 50% give an inelastic relationship with results which are very low and close to zero for the at-risk-of- monetary-poverty and unemployment rates. For p = 40% the monetary poverty rate remains inelastic but the elasticity indicator increases in value. In the case of p = PM, the elasticity value increases and is much closer to one, with values practically identical to those corresponding to unemployment and the rate of the population in a situation of severe material deprivation.

3.5 Conclusions.

The United Nations, the OECD and the European Union establish indicators for those at risk of monetary poverty which are independent from severe material deprivation and use fixed and constant percentages p. This paper aims to assess the implications of incorporating a direct poverty measurement, severe material deprivation, in the determination of an indirect poverty measurement, at-risk-of monetary poverty, from an empirical perspective and in relation to the model currently used.

Our starting point is a comparative analysis at EU level of the sensitivity of the determination of the risk of poverty to variations in the percentage value of the median equivalent income. It is observed that, when a common value for p between 40% and 70% is set, the correlation between severe material deprivation and the risk of poverty lies between 0.57 and 0.68. These results indicate that the current model for determining the risk of poverty is reasonable; our intention, however, is to improve on this model.

To this end, instead of maintaining a constant value of p to determine the risk of poverty as a mechanism to offer comparable data, we propose harmonizing the methodology used to set the value of p even if this leads to the establishment of different values of p in different countries and years. With this objective in mind we propose an initial function which attempts to minimize the dissimilarity of the risk of poverty in relation to the material deprivation function whose results provide a correlation between these two factors of 0.96; it also leads to improvements in the ordering of the various countries. The ordering problems for some countries, however, remain unresolved, which suggests that the optimal threshold is in fact not within the analyzed range of 40% to 70%. An analysis of the new indicator used to determine the median threshold over time is also presented for situations of

both increase and decrease in material deprivation, also confirming its longitudinal validity.

Once it has been shown that the establishment of different cut-off points or thresholds in relation to the median equivalent income per adult can reduce the dissimilarity and increase the correlation between the risk of poverty and severe material deprivation, the next step is to find a method which goes beyond the analysis of the sensitivity of the percentages for p previously mentioned. The proposed methodology is based on the use of receiver operating characteristic (ROC) curves to determine the optimal cut-off points based on the prevalence of material deprivation using the prevalence matching method.

The results of applying this methodology to the case of Spain show that it is feasible to obtain an optimal cut-off point $X = c$ (threshold) which is variable and a function of the annual estimate of severe material deprivation; this in turn makes it possible to infer an optimum p percentage for the median equivalent income per adult, which determines the rate of the population at risk of poverty.

Based on the previous result, the consistency of $p = PM$ and the different values of p offered by Eurostat in its databases are analyzed (p = 70%, p = 60%, p = 50% and p = 40%). The sizes of the four areas considered in the statement of the problem (A, B, C and D) are calculated in line with Figure 1 and according to the different values of p; then the indicators of sensitivity, specificity, effectiveness and the positive and negative predictive values are also calculated. An estimation of the degree of association is also made using the correlation coefficients between the different risk rates of monetary poverty and the rate of the population in a situation of severe material deprivation and their respective elasticities.

The results show that the value p = 70% yields a higher equivalent at-risk-of-poverty threshold, which means that a high number of people who are not in severe material deprivation are considered to be at risk of poverty; this means greater sensitivity but lower values for specificity, accuracy and the positive predictive value. Reducing the value of p involves lowering the threshold and, simultaneously, the size of areas A and B, reducing sensitivity while increasing specificity and effectiveness significantly since the model correctly classifies a greater number of people who are neither at risk of poverty nor in a situation of severe material deprivation (area C). Although the optimal value obtained using

prevalence matching gave the lowest sensitivity, the highest values of specificity, accuracy and positive predictive value were obtained.

Regarding the temporal dimension of the degrees of association, the rates obtained using $p = PM$ yielded levels of correlation significantly higher than those of any other case considered in this article. Of particular significance was the fact that the risk of poverty rate obtained by $p = PM$ is more consistent/coherent over time, not only with severe material deprivation, an endogenous variable in the EU-SILC, but also with unemployment, an exogenous variable obtained from the Labour Force Survey. The elasticities - understood as a measure of the variability of a variable in relation to others - obtained for the period 2008-2014 indicate that the degree of variability of severe material deprivation and the risk of poverty using $p = PM$ against the rate of unemployment is clearly higher than for constant p values, which offer a higher degree of inelasticity.

We would not like to conclude without pointing out that the proposed approach should be considered as a complement to the current harmonized standards used by international organizations for the comparative measurement of monetary poverty rather than as a substitute. In general, it would be desirable to obtain a satisfactory estimation of poverty by using a direct and multidimensional approach which would then be used to set an optimal cut-off point for the equivalent income per adult variable. Regarding this last variable, its definition provides a further limitation since it distributes the income of a household conventionally, constantly and invariably in time and space, among all household members according to the modified scale of the OECD; it is, however, possible to calculate empirically an equivalent scale based on households' annual expenditure which is not constant but varies annually and generalizes the scales proposed by the OECD, thus allowing an improved fit by type of household (Salcedo and Izquierdo Llanes, 2017).

As a final conclusion, we understand that the proposed approach will lead to new lines of additional research which, by incorporating severe material deprivation in the calculation of the rate of the population at risk of monetary poverty, make it possible to improve the desirable consistency between both and, in the case of the latter, against exogenous contrast variables, which would be useful to get a better understanding not only of cross-cutting situations but also to broaden the research on the dynamics of entries and exits into poverty. (Thorat et al., 2017).

Chapter 4

MULTIDIMENSIONAL APPROACH.

Abstract.

In the European Union poverty has been measured indirectly in a one-dimensional way from a perspective based on disposable income. This classical approach has certain limitations when representing such a complex phenomenon by means of a single variable, reaching sometimes a modest association with regard to other direct poverty measurements such as severe material deprivation rate. In this article we study the measurement of monetary poverty from a multidimensional point of view favouring a perspective of complementarity rather than one of substitutability. The joint analysis of the monetary income and consumption distribution makes it possible to identify different association patterns between these two variables for individuals located on one side or the other of the respective poverty thresholds. Expenditure on housing that is a determining factor in lower-income households and imputed rents that would be paid by the owner household of a dwelling, allow us to calculate an at-risk-of poverty rate which refines the link with material poverty in both temporal and spatial dimensions.

4.1 Introduction.

In recent decades monetary poverty has been measured, specifically, by means of the poverty risk rate based on disposable income (Atkinson, A.B. et al., 2017). This paradigm, generally accepted in the European Union (EU), has been reconsidered since the recent economic crisis, given that the indicators of severe material deprivation have shown more variation than the classical indicator of at-risk-of poverty, which in turn has led to a lower degree of association between them. One way to solve this possible dysfunction is to understand that the relationship between income and consumption has been modified by the existence of savings and/or by variations in debt service. This would lead to the need to measure the risk of poverty not only from the perspective of monetary income, but also from that of monetary consumption (Meyer, Bruce D. and Sullivan, James X., 2017). Both visions of poverty have been accepted as valid by the UNECE in its recent Manual for the harmonized measurement of poverty (UNECE, 2017).

In this sense, when applying the classical one-dimensional poverty measurement model based on income, some researchers have noted the existence of a relative modest association between the risk of poverty and material poverty (Notten and Guio, 2018), when the latter is measured in terms of the proportion of individuals in a situation of severe material deprivation, taking into account both their degree of correlation (Notten, G., 2016) and the intersection between the two subpopulations (Fusco, A. et al., 2010).

Thus, if we focus specifically on the data for a selection of EU countries in 2016 included in Table 1, we see that the sensitivity of the risk of poverty with respect to severe material deprivation stands at only 36.4 % in the case of Finland; in other words, approximately one in three of those in a situation of material poverty is at risk of monetary poverty but the other two material poor are out of risk of monetary poverty. The corresponding figure is similar in the case of Hungary (38.9%), while

Country	At risk of poverty rate (a)	Severe material deprivation rate (b)	Intersection of (a) and (b) (c)	Sensitivity (c/b)
Finland	11.7%	2.2%	0.8%	36.4%
Hungary	14.4%	16.2%	6.3%	38.9%
UK	15.9%	5.2%	2.3%	44.2%
Italy	20.6%	12.0%	5.6%	46.7%
France	13.7%	4.5%	2.3%	51.1%
Spain	22.3%	5.8%	4.0%	69.0%
Germany	16.5%	3.7%	2.6%	70.3%

Table 4.1: Intersection and sensitivity of the poverty risk and SMD rates, year 2016.

it increases for Italy (44.2%) and the United Kingdom (46.7%). In France around one out of two of those in a situation of material poverty is at risk of monetary poverty (a sensitivity of 51.1%), while the results indicate higher values in the cases of Spain (69.0 %) and Germany (70.3%), the latter being the highest value of all the EU countries.

An additional debate exists regarding whether the monetary poverty paradigm, given that it is a one-dimensional measurement system, could be improved by incorporating other dimensions (Alkire, S. et al., 2015) in order to better represent such a complex phenomenon (Serafino, P. & Tonkin, R., 2017). An Income-Consumption-Wealth (ICW) join statistical approach could provide a possible solution since it would integrate these three variables clearly linked to poverty. This join statistical approach has been reinforced at European level through the so–called Vienna memorandum on Income, Consumption and Wealth statistics, adopted in 2016, which is consistent with the ICW framework advocated by the Organisation for Economic Co-operation and Development (OECD, 2013).

At micro level, the memorandum promotes additional development of the main statistical data sources, especially EU-SILC (EU statistics on income and living conditions), HBS (Household Budget Survey) and HFCS (Household Finance and Consumption Survey).

Concerning the integration of those three variables, it is also worth noting that net wealth conditions the need for savings or the direct and indirect financing of consumption; this could explain the discrepancies between the income and consumption of some individuals. In any case, wealth, insofar as it is positive or negative, involves returns or debt service which affect income and/or consumption. In particular some authors have considered housing expenditure as an explanatory factor of some situation of poverty risk (Yang, L., 2018). The so-called income-ratio is a mainstream in the financial economy to measure accessibility, based on linking the information of defaults to indicators constructed from the relative ratio between housing expenditure and household income (Bramley, 2012), whose main weakness is that non-housing expenditures must represent a minimum proportion, which is not very applicable to households with incomes far from the average (Haffner and Heylen, 2011). But because of its potential applicability to the measurement of poverty, the alternative accessibility paradigm called residual income is particularly interesting (Stone, 2006), which is based on quantifying the absolute level of the difference between income and housing expenses, relating this difference with what is estimated as a fair standard of living. Like the economy of poverty, the residual income approach has the main difficulty of quantifying this fair standard of living since it is different for each temporal and spatial reality (Li, 2015)

Based on all previous introductory considerations, in the absence of ICW integrated empirical data we attempt an initial approach to a multidimensional model using the joint distribution of monetary income and consumption which, applied to the case of Spain, will provide the basis for the construction of a new indirect estimator of monetary poverty which represents a refinement of the classical poverty rate.

4.2 Methodology and data.

The classical approach to the measurement of monetary poverty has considered as at-risk-of-poverty those individuals whose disposable income in a year t is to be found on the left of what is known as the poverty line (Ravallion, M. and Lokshin, M., 2006). Thus, the monetary poverty risk rate is given by the proportion of individuals whose equivalent disposable income is below the poverty threshold (Lelkes, O. and Gasior, K., 2018). A percentage (p) of the median (Mdn) of the equivalent disposable income is normally used to define this poverty threshold. This percentage is conventionally set at *p=60%* in the case of the EU (Atkinson et al., 2010) as the UNECE or the OECD, recommend using values of *p=50%* for international comparisons (OECD, 2016). Methods of selection of p depending on their sensitivity, specificity and accuracy with respect to material poverty have been analysed by some authors in order to draw optimal poverty lines (Salcedo, A.M. and Izquierdo Llanes, G., 2018).

Thus, if we denote the equivalent disposable income of the individuals of a country as Y_d, the poverty line or threshold based on a percentage p of its median will be given by $y_{line,p}$, calculated as follows:

$$y_{line,p} = p\% * Mdn(Y_d) \tag{4.1}$$

The above calculation can be used for any other monetary variable, either income (Y) or consumption (C), by simply replacing the new income or consumption variable in equation 4.1. Thus, for the purpose of this article, we will denote the poverty threshold of equivalent monetary consumption for *p=60%* as $c_{line,60}$.

At this point, and before extending a one-dimensional model to a

two-dimensional model, let us consider the following proposition: "Let N be the total number of individuals in a country or region under study. Then the at-risk-of-poverty rate with p=60%, which we denote in this article as $Arop.R_{Yd,60}$, is the value of the distribution function of the equivalent disposable income (F_{Yd}) evaluated on the poverty threshold ($y_{line,60}$)". Given that $Arop.R_{Yd,60}$ represents the proportion of individuals with an equivalent income below the poverty line with p=60%, then:

$$\begin{aligned} Arop.R_{Yd,60} &= \frac{number\ of\ individuals\ with\ Y_d \leq y_{line,60}}{N} \\ &= P(Y_d \leq y_{line,60}) \\ &= F_{Yd}(y_{line,60}) \end{aligned} \tag{4.2}$$

Figure 4.1 shows the cumulative distribution function and the poverty risk rate of Spain calculated for the year 2017. This rate was 21.6% or, in other words, the risk of poverty rate $Arop.R_{Yd,60}$ is located in percentile 21.6 of the distribution function of Y_d. In case of using *p=50%* the monetary poverty rate is 15.7%.

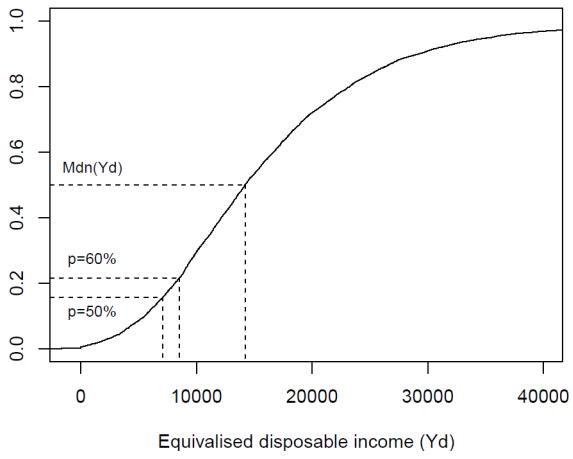


Figure 4.1: Cumulative distribution function, poverty line and at-risk-of poverty rates, year 2017.

Based on the aforementioned proposition, when considering a two-dimensional income-consumption variable we can immediately define a two-dimensional poverty risk rate as the value of the two-dimensional distribution function $F_{Y,C}$ evaluated

at the centroid determined by the respective one-dimensional poverty thresholds, as follows:

$$Arop.R_{Yd,60} = F_{Y,C}(y_{line,60}, c_{line,60}) \tag{4.3}$$

From a methodological point of view, in order to validate the results of this model externally, we will use the degree of association obtained by means of the different correlation coefficients with the rate of population suffering severe material deprivation, a direct measure of poverty (Chzhen et al., 2016). In the EU, a person who cannot afford at least four of the following nine items (Rajmil et al., 2015) is considered to be in a situation of severe material deprivation (Ayllón and Gábos, 2017): to pay rent or utility bills; to keep home adequately warm; to face unexpected expenses; to eat meat, fish or a protein equivalent every second day; a week holiday away from home; a car; a washing machine; a colour TV; a telephone. The results of this indicator have been analysed by various authors with a view to suggesting possible improvements (Guio et al., 2016). We will also apply the method of principal components to the study of the joint distribution of income and consumption variables. The parameters of sensitivity, specificity and accuracy, often used in the estimation of results using ROC curves (Fawcett, 2006) are also applied to perform an internal analysis of poverty at the microdata level. Sensitivity represents the probability that a person is at risk of monetary poverty while also being materially poor. Specificity, meanwhile, constitutes the probability that a person who is not materially poor is not at risk of poverty. Accuracy refers to the probability that a person is both monetary and materially poor (simultaneously) or neither monetary nor materially poor (simultaneously).

Regarding data, this study uses empirical information from two main sources: the Household Budget Survey (HBS) and the Living Conditions Survey (LCS); all the anonymized microdata files can be downloaded free on the website *http* : *//www.ine.es/en/prodyser/microdatos_en.htm*. The HBS is a sample survey carried out by Spain's National Institute of Statistics; its objective is to provide information about the nature and destination of consumer spending, as well as on various aspects related to household living conditions. This survey provides annualized data for net income and monetary expenditure, it is updated annually with a sample size of approximately 24,000 households, which makes its results highly representative. Eurostat, meanwhile, collects HBS data from all European countries every five years, the most recent being 2015. The LCS is Eurostat's equivalent of the European Union Statistics on Income and Living Conditions (EU-SILC). Its main objective is to provide information about income distribution

and social exclusion at national and European level, including the material deprivation of the population. In Spain, the LCS investigates around 13,000 households each year.

4.3 Results.

The joint distribution of monetary income and consumption.

We begin with the study of the joint distribution of the monetary income and consumption of Spanish households. Figure 4.2 shows the two-dimensional scatter diagram of these two variables at microdata level obtained from the HBS with 2017 as year of reference. This figure shows the representation of the 2D density lines. In the upper part and on the right, the marginal density functions of income and consumption are also shown. The two poverty lines of net income ($y_{line,60}$) and monetary consumption ($c_{line,60}$) calculated with p=60% have also been added. The inclusion of the two poverty lines makes it possible to visualize the two-dimensional centroid ($y_{line,60}, c_{line,60}$) as the intersection of the one-dimensional income and consumption poverty thresholds, respectively. This, in turn, means the quadrant can be divided into four clearly differentiated areas.

In area I, all individuals are below the two poverty thresholds. Given that everyone in this zone experiences low levels of both income and consumption, a high degree of correlation between the risk of poverty and the rate of the population in severe material deprivation would be expected. In area II, individuals have a low level of income but their levels of monetary spending are medium-high, since they are located above the poverty line for consumption. This situation could be related to the sale of household goods, the reduction of previously accumulated savings, indebtedness, family assistance or might even suggest the existence of illegal shadow economy activities (Eurostat, 2018). The individuals in area III have a low level of monetary expenditure but their income levels are medium-high since they are located above the income poverty line. They could be saving and/or facing debt service. It should be noted that low levels of monetary consumption could be significantly affected by the different price levels (PPP) to be found in

Spain's autonomous communities (Salcedo and Izquierdo Llanes, 2017), which could condition the measurement of the risk of poverty. Finally, the individuals in area IV have medium-high levels of income and monetary spending; they are all located above the two poverty lines. This situation indicates that these individuals are not at risk of poverty.

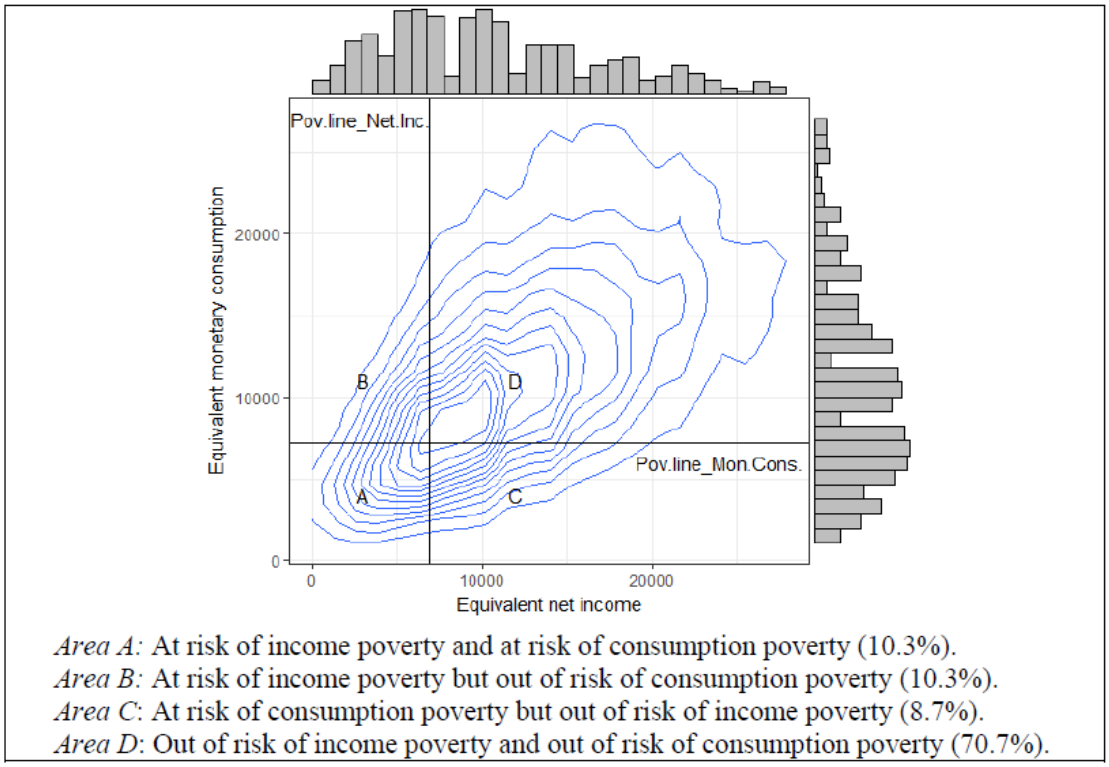


Figure 4.2: 2D-density scatter plot of equivalent net income and monetary expenditure.

Given the existence of a high degree of association between household income and expenditure, it would be expected, a priori, that the percentage of people at risk of income and consumption poverty would be very high in relation to the total population in one or another risk. However, we observe that only 1 in 3 of those at risk of income or consumption poverty (the total of areas I + II + III is 30.1%) is simultaneously at risk of income and consumption poverty (area I, 10.5%); this seems to suggest an anomalous situation in the one-dimensional models of income or consumption poverty when these are considered separately.

Table 4.2 shows the correlation coefficients obtained between the proportion of

people in a situation of severe material deprivation and the monetary poverty risk rates in Spain based on EU-SILC and HBS data. The period analysed spans the years 2008 to 2017, which is especially significant since it covers the whole period affected by the recent financial crisis. It can be observed that the two-dimensional model offers a very high degree of association, surpassing even the good results obtained from the one-dimensional models, in particular the standard used in the EU-SILC.

	Direct poverty measurement	Indirect poverty measurement		
		One dimension		Two dimensions
Year	SMD rate (*)	$Arop.R_{Yd,60}$ (*)	$Arop.R_{Yc,60}$ (**)	$Arop.R_{YC,60}$ (**)
2017	5.1%	21.6%	19.3%	10.5%
2016	5.8%	22.3%	19.0%	10.3%
2015	6.4%	22.1%	19.6%	10.6%
2014	7.1%	22.2%	19.1%	10.7%
2013	6.2%	20.4%	18.1%	9.8%
2012	5.8%	20.8%	18.2%	9.1%
2011	4.5%	20.6%	18.2%	9.1%
2010	4.9%	20.7%	18.6%	8.8%
2009	4.5%	20.4%	17.9%	8.7%
2008	3.6%	19.8%	17.8%	8.3%

(cont.)

	Direct poverty measurement	Indirect poverty measurement		
		One dimension		Two dimensions
Year	SMD rate (*)	$Arop.R_{Yd,60}$ (*)	$Arop.R_{Yc,60}$ (**)	$Arop.R_{YC,60}$ (**)
- Pearson corr. coef.		0.73	0.59	0.83
- Spearman corr. coef.		0.68	0.61	0.88
- Kendall corr. coef.		0.60	0.48	0.75

Table 4.2: Direct and indirect poverty rates (period 2008-2017).

(Note: Based on the EU-SILC database (*) and HBS microdata (**).)

Unfortunately, the indicator based on a two-dimensional model in *Table 4.2* cannot be completely validated using this approach since the HBS does not offer information at microdata level of any direct measure of poverty, in particular, that of severe material deprivation. But the results of this table and the joint distribution income-consumption suggest the possible existence of an indicator, based on a linear combination of the variables of income and consumption, which could offer a better approximation to the measurement of poverty than the one-dimensional classical indicator based exclusively on disposable income, as it is applied in the European Union among others. At this point it will be necessary to return to the EU-SILC using multivariate analysis techniques and, thus, be in a position to apply external and internal tests in both temporal and spatial dimensions.

Income characterized by expenditure on housing.

Following an analysis of the joint distribution of the equivalent income and monetary consumption in *Figure 4.2* and the poverty measurement results presented in *Table 4.2*, we proceed to a principal components analysis of the income and consumption data for 2017, which provides us with the following standardized linear equations:

$$\begin{cases} \text{PC1: } 0.707 * Y + 0.707 * C \\ \text{PC2: } 0.707 * Y - 0.707 * C \end{cases} \tag{4.4}$$

It can be seen that the first principal component (PC1) provides an eigenvector on the diagonal of the first quadrant. In *Table 4.3* we show the cumulative proportion of total variability explained by this component is 76.93%, which can be considered as significant and indicates that most of the two-dimensional variability is concentrated in this first component, that is, along the straight line on which standardized income and consumption are equal.

	PC1	PC2
Standard deviation	1.2404	0.6792
Proportion of Variance	0.7693	0.2306
Cumulative Proportion	0.7693	1.0000

Table 4.3: Summary of principal components analysis, year 2017.

The second principal component (PC2), meanwhile, explains 23.06% of the remaining variability with a subtraction, indicating a contrast between net income and monetary expenditure; this could be interpreted as the different levels of monetary savings of households. According to this second principal component, in the case of simultaneously low values of Y and C, the range of variation of savings (positive or negative) is also low; this in turn implies the existence of a low capacity of indebtedness of households and could result in situations of poverty and/or financial exclusion (Krumer-Nevo et al., 2017) affecting the financial well-being of

households (Lee and Sabri, 2017).

It should be pointed out, following on from the previous reflection, that there is a wide range of financial ratios for households calculated for different purposes (Harness, Nathaniel J. et al., 2008). The European Central Bank, for example, has considered various consumption-to-income ratios in the scope of the Household Finance and Consumption Survey (ECB, 2016). Besides, in the framework of the EU-SILC, a transformation of disposable income is also frequently used by adding the imputed income from the dwelling to the equivalent income (Törmälehto and Sauli, 2013), in order to offer a complementary measure of monetary poverty; although imputed rents are not, by definition, part of equivalent income, it can be considered as aggregate income in national accounting terms (Eurostat, 2013). In this context and for the purpose of this article we denote as Y_{id} the variable disposable income adding imputed rents and equivalised following the usual procedures.

Given that housing is usually purchased using a loan and, if income is not adjusted with financial expenses this could have the perverse effect that someone who bought a home with a loan of 100%, and whose imputed income was dedicated to servicing the loan, would be considered to have a greater income than just before buying the home, that coincides with the temporary moment when that person did not pay any mortgage although he or she could be facing the payment of a rent (Attanasio et al., 2012). This possible dysfunction leads to the incorporation of the expenses related with housing, mainly debt service and rent, into the indicators used to calculate the at risk of poverty. In addition, in the case at hand, the expression of the first principal component of the joint distribution of income and monetary consumption leads us to search for linear combinations, in the form of differences between income and consumption, in order to obtain the greatest variability possible.

Taking into account all of the above, we analyse the HBS to identify the item of highest monetary expenditure in the lowest income households, based on the international classification COICOP (Berardi, N. et al., 2017) which breaks down household expenditure into the following twelve groups: 1. Food and non-alcoholic beverages; 2. Alcoholic beverages, tobacco and narcotics; 3. Clothing and footwear; 4. Housing, water, electricity, gas and other fuels; 5. Furniture, household equipment and ordinary expenses for the maintenance of the dwelling; 6. Health; 7. Transport; 8. Communication; 9. Leisure, performances and culture; 10. Education; 11. Restaurants, cafes and hotels; 12. Miscellaneous goods and services.

Of these twelve groups, spending on group 4 (housing) is clearly the largest of all expenditure items in households with the lowest income. *Table 4.4* shows the proportion of expenditure on housing (including rent, interest payments on mortgages, water, electricity, gas and other fuels) by income quintile in five European countries in 2015. We can see that the percentage of monetary expenditure associated with this group is around 40% of total expenditure for households in the first income quintile, while in the case of households in the top quintile this percentage decreases by between -9.5 and -15.3 percentage points.

	Income quintile		
Country	Q1	Q5	Diff. (p.p.)
Bulgaria	39.7%	28.6%	-11.1
Finland	39.1%	27.0%	-12.1
Germany	43.3%	28.0%	-15.3
Hungary	46.1%	31.0%	-15.1
Spain	38.6%	29.1%	-9.5

Table 4.4: Percentage of monetary expenditure in housing, water, electricity, gas by income quintile (year 2015)

It is clear that, unlike other COICOP items such as alcoholic beverages and tobacco, leisure and culture or eating out, this item of expenditure is obligatory for households and its high proportion in the lower income quintile clearly conditions the capacity to pay for other fundamental goods or services; this could be related to situations of severe material deprivation in low-income households.

For all these reasons and based on the above results, we define the income characterized by expenditure on housing, which henceforth we will call Y_{dc}, as the disposable income of the household once the total expenditure on housing facing the household has been deducted; this latter figure is reflected in the EU-SILC at the microdata level and it has been recently used by Eurostat to calculate other poverty rates that differ from the standard use of equation (1). This variable has also commonalities with the concept of residual income (Stone, 2006). Finally, the equivalised imputed income characterized by expenditure on housing (Y_{idc})

is also defined analogously to Y_{dc} but adding imputed rents to the characterized income. In the next section we investigate whether the characterized income offers an improvement over the classical poverty risk estimator based solely on disposable income.

Refining the classical measurement of the monetary poverty.

To check the quality of the estimation of the monetary poverty risk based on characterized income, we will take the last available year (2017) as our reference year and, using the classical estimator based on Y_d and with *p=60%* of the median, we will carry a comparative study of the poverty rates based on the three income variables previously presented in this paper, that is, Y_{id}, Y_{dc} and Y_{idc} and also with *p=60%* of their respective medians according to equation (1).

Firstly, we verify that the areas below the ROC curve (López-Ratón et al., 2014) obtained with the variables Y_d, Y_{id}, Y_{dc} and Y_{idc} in 2017 are 0.82, 0.84, 0.83 and 0.84, respectively. It can be shown that the area under the ROC curve (AUC), which takes values between 0.5 and 1.0, is equivalent to that of the Mann-Whitney test (Hand and Till, 2001). We can see that in this case Y_{id}, and Y_{idc} offer the highest values of the AUC.

Our second test consists of an analysis -external and internal- of the temporal dimension. Table 4.5 shows the proportion of individuals in a situation of severe material deprivation as well as the poverty risk rates obtained using the variables Y_d, Y_{id}, Y_{dc} and Y_{idc} for the decade 2008-2017. It is observed that the monetary poverty rate derived from disposable income by adding imputed rents Y_{id} is lower than the classical one, between -1.7 and -3.1 percentage points. On the contrary, the disposable income characterized by expenditure in housing (Y_{dc}) increased the rates from +3.3 to +4.1 percentage points. The inclusion of imputed rents in the characterized income (Y_{idc}) offers more similar rates than the classical indicator, with differences ranging from +0.3 to +2.1 percentage points. The correlation coefficients obtained are very high in all cases, although the variable Yidc offered very high values (0.94, 0.91 and 0.81).

Year	SMD rate	$Arop.R_{Yd,60}$	$Arop.R_{Yid,60}$	$Arop.R_{Ydc,60}$	$Arop.R_{Yidc,60}$
2017	5.1	21.6	19.7	25.6	22.6
2016	5.8	22.3	19.8	25.8	22.6
2015	6.4	22.1	19.5	25.4	22.8
2014	7.1	22.2	19.9	26.1	23.5
2013	6.2	20.4	18.7	24.5	22.5
2012	5.8	20.8	19.0	24.9	22.3
2011	4.5	20.6	17.8	24.6	21.6
2010	4.9	20.7	17.6	24.3	21.6
2009	4.5	20.4	17.3	24.0	21.2
2008	3.6	19.8	17.1	23.6	20.7
- Pearson coef.		0.73	0.82	0.78	0.94
- Spearman coef.		0.68	0.80	0.74	0.91
- Kendall coef.		0.60	0.66	0.61	0.81

Table 4.5: Severe material deprivation and at-risk-of poverty rates (%).

The situation is similar when an internal study -at micro level- of the sensitivity, specificity and accuracy of the four variables is carried out. Table 4.6 shows that, in all cases, the characterized income is more sensitive than that of Y_d, reaching a maximum of 76.2% in 2016; this is an indication that the intersection between the risk of poverty rate and severe material deprivation is greater with this variable. As far as specificity is concerned, the highest values are obtained when considering imputed rents only (84.9% in 2008 and 2009), that is, this variable offers the largest intersection between individuals that are not materially poor and out of risk of poverty, simultaneously. Finally, the accuracy of the variable Y_{id} is again the highest of the four cases considered, with a maximum of 83.4% in 2009. This table also shows that all sensitivity results for Y_{idc} are greater than those for the classical Y_d with p = 60%, reaching +10.8 percentage points in 2011, while the specificity and accuracy are rather similar, around 80% every year.

		2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
Se.	$Y_{d,60}$	57.8	58.8	58.9	53.9	57.9	56.0	63.1	62.0	69.6	63.8
	$Y_{id,60}$	62.3	58.6	59.0	54.2	59.7	61.1	64.1	64.7	69.9	62.6
	$Y_{dc,60}$	70.1	67.1	66.3	63.6	67.7	65.5	72.0	72.3	76.2	70.4
	$Y_{idc,60}$	68.3	66.5	64.2	64.7	65.3	65.9	70.4	70.6	74.1	68.7
Sp.	$Y_{d,60}$	81.6	81.4	81.3	80.9	81.5	82.0	80.9	80.6	80.6	80.7
	$Y_{id,60}$	84.6	84.6	84.5	83.9	83.5	84.1	83.5	83.6	83.3	82.6
	$Y_{dc,60}$	78.1	78.0	77.8	77.3	77.8	78.2	77.4	77.8	77.3	76.8
	$Y_{idc,60}$	81.1	80.9	80.6	80.5	80.4	80.4	80.0	80.4	80.5	79.9
Ac.	$Y_{d,60}$	80.7	80.4	80.2	79.7	80.1	80.4	79.6	79.4	79.9	79.8
	$Y_{id,60}$	83.8	83.4	83.3	82.6	82.2	82.7	82.1	82.4	82.5	81.6
	$Y_{dc,60}$	77.8	77.5	77.3	76.7	77.2	77.4	77.0	77.5	77.2	76.5
	$Y_{idc,60}$	80.7	80.2	79.8	79.8	79.5	79.5	79.4	79.8	80.2	79.3

Table 4.6: Sensitivity (Se.), specificity (Sp.) and accuracy (Ac.)with regard to the severe material deprivation (%).

Finally, as a third test, we studied the spatial dimension, focusing on the results calculated for the seventeen Spanish autonomous communities at the *NUTS2* level with reference year 2017. In the internal analysis, *Table 4.7* shows the severe material deprivation and poverty risk rates obtained from the equivalent disposable income and the equivalent characterized income for all regions. To simplify this analysis, only the sensitivity (*Se.*) of the poverty risk rate with respect to severe material deprivation is used.

NUTS2	SMD	$Y_{d,60}$		$Y_{id,60}$		$Y_{dc,60}$		$Y_{idc,60}$	
		Arop.R	Se.	Arop.R	Se.	Arop.R	Se.	Arop.R	Se.
Total	5.1	21.6	63.8	19.7	62.6	25.6	70.4	22.6	68.7
ES11 Galicia	2.4	18.7	80.9	16.6	77.5	20.3	80.9	17.8	78.7
ES12 Asturias	3.5	12.6	78.4	12.3	78.0	15.7	88.4	14.9	83.3
ES13 Cantabria	2.2	17.6	84.5	13.0	84.5	21.9	*100.0*	18.8	*100.0*
ES21 P. Vasco	3.7	9.7	35.7	8.6	40.7	14.0	50.7	11.4	58.6
ES22 Navarra	0.3	8.3	68.3	8.4	68.3	11.6	88.3	11.4	88.3
ES23 Rioja	2.9	9.7	23.2	11.2	30.4	16.2	64.0	14.2	64.0
ES24 Aragón	0.5	13.3	*88.6*	10.2	*88.6*	16.0	*100.0*	12.4	*88.6*
ES30 Madrid	5.4	16.9	67.8	16.6	61.4	22.7	74.9	20.8	71.4
ES41 C.León	1.0	15.4	52.4	14.1	57.4	20.2	76.3	16.6	76.3
ES42 C.Mancha	4.4	28.1	50.6	26.9	38.8	31.6	57.5	31.1	48.8
ES43 Extremad.	5.6	38.8	64.5	33.5	70.8	41.0	75.0	33.7	79.9
ES51 Cataluña	5.0	15.0	60.0	13.3	54.3	20.2	66.9	18.5	66.0
ES52 C. Valenc.	7.4	25.6	64.3	24.2	65.0	29.1	67.2	25.4	66.9
ES53 I. Balears	6.9	21.3	61.2	23.8	64.0	28.6	64.6	26.8	64.0
ES61 Andalucía	5.2	31.0	71.1	27.5	77.1	33.8	80.1	28.4	74.5

(cont.)

NUTS2	SMD	$Y_{d,60}$		$Y_{id,60}$		$Y_{dc,60}$		$Y_{idc,60}$	
		Arop.R	Se.	Arop.R	Se.	Arop.R	Se.	Arop.R	Se.
ES62 Murcia	6.2	30.1	67.9	25.9	69.8	35.3	77.3	27.9	73.8
ES70 Canarias	13.6	30.5	58.0	25.9	52.7	32.2	59.0	32.1	62.0

Table 4.7: At risk of poverty rates and sensitivity (%) with regard to the population on severe material deprivation, year 2017.

(Note: (highest sensitivity values in italics and underlined))

The classical variable Y_d offers low sensitivity in regions ES23 (Rioja) and ES21 (Basque Country) of only 23.2% and 35.7%. The inclusion of imputed rents Y_{id} increases the sensitivity in eight of the seventeen autonomous communities, is unchanged in three and reduces in six. All sensitivity values improve significantly when the characterized equivalent income is considered. It is also noteworthy that in the autonomous communities ES13 (Cantabria) and ES24 (Aragón) the new variable reaches a sensitivity of 100%; that is, in these cases the maximum possible intersection is achieved. Besides, ES30 (Madrid) and ES43 (Extremadura) are the regions with highest and lowest GDP per capita in Spain respectively; they have a severe material deprivation rate rather similar (5.4% and 5.6% respectively, +0.2 percentage points only) but the situation is quite different when checking the classical risk of monetary poverty (16.9% and 38.8%, that is, +21.9 percentage points). After adding imputed rents the poverty rate doesn't change too much in Madrid but in Extremadura the risk of poverty is reduced to 33.5%. If deducing housing costs, Madrid increases the monetary poverty to 22.7% and Extremadura to 41.0%. The combined effect of imputed rents and housing costs (Y_{idc}) set the risk of poverty in 20.8% in Madrid and 33.7% in Extremadura, reducing the difference to +12.9 percentage points. In this last case it is remarkable that the sensitivity is also increased to 71.4% in Madrid and 79.9% in Extremadura.

Regarding the analysis evaluated via different degrees of association, in Figure 4.3 it can be observed that the correlation between poverty risk rates and severe material deprivation by regions is increased by using Y_{idc}, with the coefficient of determination rising from 0.39 to 0.53, which means a greater proportion of variability which can be explained using the new variable. The Spearman and Kendall correlation coefficients, meanwhile, also improve from 0.69 and 0.51 with the classical poverty rate to 0.76 and 0.54 respectively with the estimator based on Y_{idc}.

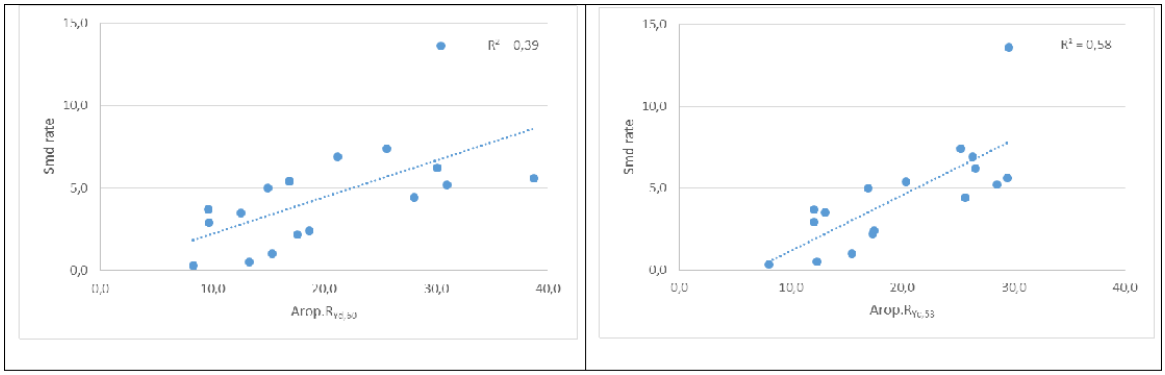


Figure 4.3: At-risk of poverty rates and severe material deprivation by region, year 2017.

To conclude the analysis of the spatial dimension, Table 4.8 shows the poverty rates by degree of urbanisation. Severe material deprivation rate is higher in very populated areas (cities, 6.0%) than in medium populated or rural areas (4.9% and 3.7%, respectively). On the contrary, the classical at-risk-of poverty rate is lower in cities (19.2%) than in towns (22.1%) and rural areas (25.9%). The risk of poverty based on Yidc increases the poverty rate in cities (+1.9) and towns and suburbs (+1.3) but decreases the poverty rate in rural areas (-1.0). The sensitivity is increased in all cases and, in this regard, it is worth noting that in rural areas the monetary poverty rate based on Y_{idc} (24.9%) is lower than the classical one (25.9%) but the sensitivity is increased +7.0 percentage points (75.4%).

		$Y_{d,60}$		$Y_{idc,60}$	
Degree of urbanisation	SMD Rate	Arop.R	Se.	Arop.R	Se.
1. Cities	6.0	19.2	63.3	21.1	68.0
2. Towns and suburbs	4.9	22.1	61.4	23.4	65.0
3. Rural areas	3.7	25.9	68.4	24.9	75.4
Total	5.1	21.6	63.8	22.6	68.7

Table 4.8: At risk of poverty rates and sensitivity (%) by degree of urbanisation, year 2017.

4.4 Conclusions.

This chapter investigates the extension of the classical monetary poverty measurement to a multidimensional approach trying to refine the current link with material deprivation, which is a direct poverty measurement. The study broadens the classical one-dimensional disposable income model and makes it applicable to other monetary variables, for example monetary consumption, via the distribution function due to the fact that the poverty risk rate coincides with the value of this distribution function evaluated on the poverty threshold. Building from here, a two-dimensional poverty risk rate (income-consumption) based on the centroid determined by the respective one-dimensional thresholds is defined. This rate is seen to show a stronger association in terms of correlations with material poverty than the two one-dimensional variables it is based on.

In this context the join distribution of monetary income and consumption at micro data level is explored, paying special attention to the left side of the distribution based on the two poverty thresholds that determine the centroid $(y_{line,60}, c_{line,60})$. The analysis of the two-dimensional poverty risk rate (income-consumption) makes it possible to determine two typologies. On the one hand, of those individuals whose consumption is more clearly linked to their income, both those who are located below both poverty thresholds (*area I*), and those whose levels of income and expenditure are above the two poverty lines should be considered (*area IV*). On the other hand, of those individuals with a less clear association between income and consumption that, additionally, allow us to consider another two different situations: the first consists of individuals who have a low level of equivalent income but whose levels of monetary expenditure are above the consumption poverty line (that is, *area II*), which could conceal situations of consumption financed by means of previously accumulated wealth, debts, family assistance or even informal economy activities, which would mean an infra declaration of income and that such individuals could not be really in a situation of material deprivation; the second would consist of individuals with a low level of monetary expenditure but whose income levels are medium-high since, in this case, they are to be found above the income poverty line (that is, *area III*), who are normally individuals facing debt service, usually a mortgage linked to home

purchase. The interpretation of the latter situation provides an additional reason for the incorporation, with monetary variables, of the expenses and/or income related with net wealth as carried out in this study.

After conducting an ACP analysis to check the component that accumulate a greater variability, we analyse whether a linear combination of monetary income and consumption may offer a refinement of the classical approach to the monetary poverty. Since expenditure on housing is determinant in households in the first income quintile, and with the restriction of using empirical information based on official sources of statistics, the solution applied is to consider in the EU-SILC area the equivalised income characterized by expenditure on housing, with and without imputed rents.

The area under the curve obtained for Y_d, Y_{id}, Y_{dc} and Y_{idc} in 2017 are 0.82, 0.84, 0.83 and 0.84, respectively. These results are between 0.8 and 0.9 and can be considered as excellent (Mandrekar, 2010) particularly in the cases of Yid and Yidc since they offer a slight improvement of the AUC compared to the classical Y_d. Concerning the temporal dimension, which covers the decade 2008-2017, the associations measured via the correlation coefficients between severe material deprivation and the risk of poverty rates for this period offer better coefficients being obtained with the characterized income adding imputed rents, Y_{idc}. The internal test at micro level, meanwhile, also throws up the result that, once again, Y_{idc} has a greater sensitivity than the classical Y_d (+10.8 percentage points in 2011) while the specificity and accuracy are always rather similar (around 80%). As far as the spatial dimension is concerned, the internal test is carried out via the analysis of the sensitivity of the indicator to severe material deprivation; when using the characterized income adding imputed rents Y_{idc} this value increases in most of the autonomous communities reaching the maximum intersection of 100% in some regions. On the other hand, the external test is carried out with the results obtained in the seventeen Spanish autonomous communities at the NUTS level and leads to the conclusion that the characterized income Y_{idc} also increases the coefficient of determination and correlation with material deprivation. The results achieved are also more consistent when an analysis by degree of urbanisation is carried out, particularly in the cases of rural areas and cities. We can therefore conclude in this case that, from an empirical point of view, the poverty risk rate obtained using the equivalent characterized income adding imputed rents Y_{idc} is an indicator that succeeds in refining the good results of the classical poverty risk indicator, in both its temporal and spatial dimensions.

Notwithstanding the good results achieved there are also some opportunities and limitations to be considered. The case presented in this study focuses the analysis in a country of the European Union and, at this stage, the conclusions should be limited to a context of complementary rather than substitutability of the classical Yd, which is an international standard. In addition, the applied approach is exclusively focused on monetary poverty variables in order to better measure the effect of the refinement, but it could also be extended by adapting the percentage p introduced in equation (1) instead of considered it as a constant parameter defined by convention, 60% in the European Union, or by introducing other multidimensional indicators to measure the poor, not only in developed countries (García-Pérez, C. et al., 2016) but also by the different regions (Jurado, A. and Pérez-Mayo, J., 2012) and, especially, if regional purchase parities were applied to the equivalence scales. Influence of risk factors of income poverty and severe material deprivation (Verbunt and Guio, 2019) is another element that could be taken into consideration for widening the analysis. Finally, to be able to conclude a joint monetary income and consumption analysis it would be very interesting to have empirical data containing the two-dimensional patterns of households/individuals together with a direct measure of poverty, particularly severe material deprivation.

This study allows us to continue a line of research that seeks to improve the measurement of monetary poverty from a multidimensional perspective (Santos and Villatoro, 2018), on this occasion by integrating the two visions of monetary poverty based on income and/or consumption according to UNECE, favouring the development of an integrated Income-Consumption-Wealth statistical framework, and also laying the foundations of a potential conceptual convergence between residual income and characterized income indicators, with the consequent improvement of them.

Chapter 5

CONCLUSIONS.

The main conclusions of this doctoral thesis are highlighted below:

- The issue of the low relation levels between the severe material deprivation rate and the at-risk-of-poverty rate using a regional approach is analyse in Chapter 2. A review of the main features of standardised definitions was carried out to identify elements or factors that could be influencing the results at regional level. Assuming that severe material deprivation, as an observed variable in the national surveys, is subject to an adequate and timely editing and depuration by the national statistical offices, the attention was focused on the variable equivalent income, which determines the population at-risk-of-poverty in a conventional way.

- A parametric model was proposed that generalised not only the OECD modified scale, obtained by replacing the parameters $a_{1,t}$ by 0.5, $a_{2,t}$ by 0.3, and $a_{3,t}$ by 1 for any year t, but also all other scales that have been proposed in this context.

- It was shown empirically that the equivalence scales aren't constant but they vary over time. Hence, the revealed equivalence scale fits the observed average consumption of the different type of households much better, which is particularly meaningful in those households with a larger size in terms of number of members (adults and children). Regarding the evolution of the equivalent units, it was observed that the application of the OECD-modified scale has underestimated the equivalent consumption units in Spain in the years before the financial crisis, while it has overestimated the numbers of equivalent adults in the years of the financial crisis (and now). The model proposed for Spain takes into consideration not only the changes of the consumption patterns over time but also the variability of purchasing power parities in the different regions which, without any doubt, influences the results, with the effect being greater as the variability of the parities increases. In this way, the introduced model revealed a set of parameters based on the average consumption pattern of the households and on an adjustment of the equivalent income by means of the purchasing parities.

- The application of the new scales led to an increase of a 15% in the at-risk-of-poverty threshold of the year 2012 (i.e., 1,214 euros more) which in turn involved a reduction of the at-risk-of-poverty rate of 0.42 percentage points. It is also remarkable that, by using the revealed equivalence scales,

the intersection between the estimators of severe material deprivation and at-risk-of-poverty (T_1 and T_2) increased by 1.61 percentage points (that is, 45.8 thousand of people) and, consequently, the predictive capacity of the variable T_2, obtained from the income side, with respect to the severe material deprivation T_1, obtained from the consumption side, is increased.

- Concerning the data by regions, the use of revealed parameters improved the outcome obtained with the OECD-modified scale substantially. The application of the parametric model partially changed the composition of the population at-risk-of-poverty, raising the rate in those regions with higher purchasing power parities and reducing the percentage in those regions with lower purchasing parities, showing more coherent results in line with the results obtained on severe material deprivation. The proportion of the independent variable variance explained by the dependent variable (R^2) tripled from 0.1774 to 0.5288. In addition, the three independence tests (Pearson, Spearman and Kendall) gave a p-value lower than 0.05 and, therefore, the hypothesis of lack of correlation between the severe material deprivation rate and the at-risk-of-poverty rate by region of residence is rejected.

- There are several improvement opportunities and also limitations that influence the final outcome. The election of the parametric model is itself a limitation of the study. Searching for a feasible solution to the parametric model of equivalence scales by means of the Generalized Reduced Gradient (GRG) Nonlinear is also another limitation of the study, since it might be possible to find a different optimal solution by applying another algorithm. Nevertheless, this was not the aim of this study; our objective was to use a reasonable method to prove that there is at least a solution of the parametric model that clearly improves the current OECD scales using an empirical method instead of a constant and invariant approach applied to European countries by convention. Another limitation of this study involves the use of regional purchasing power parities. In the future, the performance of a harmonised method to calculate these regional PPPs would allow the production of comparable information by regions (and years) at European level. As an alternative, the definition of an agreed proxy in index form would make it possible to produce an adjustment of these rates to the regional reality.

- The optimal cut-off point is studied in Chapter 3. The aim was to assess the implications of incorporating a direct poverty measurement, severe material deprivation, in the determination of an indirect poverty measurement, at-risk-of monetary poverty, from an empirical perspective and in relation to the model currently used in the United Nations, the OECD and the European

Union, that establish indicators for those at risk of monetary poverty which are independent from severe material deprivation and use fixed and constant percentages p. Even though the results indicate that the current model for determining the risk of poverty is reasonable, the intention, however, is to improve on the model. To this end, instead of maintaining a constant value of p to determine the risk of poverty as a mechanism to offer comparable data, we propose harmonizing the methodology used to set the value of p even if this leads to the establishment of different values of p in different countries and years.

- The receiver operating characteristic (ROC) curves are applied to determine the optimal cut-off points based on the prevalence of material deprivation using the prevalence matching method. The results of applying this methodology to the case of Spain show that it is feasible to obtain an optimal cut-off point $X = c$ (threshold) which is variable and a function of the annual estimate of severe material deprivation; this in turn makes it possible to infer an optimum p percentage for the median equivalent income per adult, which determines the rate of the population at risk of poverty. Based on this apprach, the consistency of $p = PM$ and the different values of p offered by Eurostat in its databases are analyzed (p = 70%, p = 60%, p = 50% and p = 40%). The sizes of the four areas considered in the statement of the problem (A, B, C and D) are calculated in line with Figure 1 and according to the different values of p; then the indicators of sensitivity, specificity, effectiveness and the positive and negative predictive values are also calculated. An estimation of the degree of association is also made using the correlation coefficients between the different risk rates of monetary poverty and the rate of the population in a situation of severe material deprivation and their respective elasticities.

- The results show that the value p = 70% yields a higher equivalent at-risk-of-poverty threshold, which means that a high number of people who are not in severe material deprivation are considered to be at risk of poverty; this means greater sensitivity but lower values for specificity, accuracy and the positive predictive value. Reducing the value of p involves lowering the threshold and, simultaneously, the size of areas A and B, reducing sensitivity while increasing specificity and effectiveness significantly since the model correctly classifies a greater number of people who are neither at risk of poverty nor in a situation of severe material deprivation (area C). Although the optimal value obtained using prevalence matching gave the lowest sensitivity, the highest values of specificity, accuracy and positive predictive value were obtained. Regarding the temporal dimension of the degrees of association, the rates obtained using $p = PM$ yielded levels of correlation significantly higher than those of any other case considered in this

article. Of particular significance was the fact that the risk of poverty rate obtained by $p = PM$ is more consistent/coherent over time, not only with severe material deprivation, an endogenous variable in the EU-SILC, but also with unemployment, an exogenous variable obtained from the Labour Force Survey. The elasticities - understood as a measure of the variability of a variable in relation to others - obtained for the period 2008-2014 indicate that the degree of variability of severe material deprivation and the risk of poverty using $p = PM$ against the rate of unemployment is clearly higher than for constant p values, which offer a higher degree of inelasticity.

- This approach should be considered as a complement to the current harmonized standards used by international organizations for the comparative measurement of monetary poverty rather than as a substitute. In general, it would be desirable to obtain a satisfactory estimation of poverty by using a direct and multidimensional approach which would then be used to set an optimal cut-off point for the equivalent income per adult variable. Regarding this last variable, its definition provides a further limitation since it distributes the income of a household conventionally, constantly and invariably in time and space, among all household members according to the modified scale of the OECD; it is, however, possible to calculate empirically an equivalent scale based on households' annual expenditure which is not constant but varies annually and generalizes the scales proposed by the OECD, thus allowing an improved fit by type of household. By incorporating severe material deprivation in the calculation of the rate of the population at risk of monetary poverty, it makes it possible to improve the desirable consistency between both and, in the case of the latter, against exogenous contrast variables, which would be useful to get a better understanding not only of cross-cutting situations but also to broaden the research on the dynamics of entries and exits into poverty.

- In Chapter 4 the limitatios of the poverty risk rate based on disposible income was studied. It has some limitations; for example, in terms of its degree of association or its sensitivity with other direct measures of poverty which have led to new research seeking to improve the results currently available by extending and/or modifying the classical model. For example, one of the adjustments employed involves a revision of the value of the percentage of the median used to determine the poverty risk threshold, conventionally 60% in the case of the EU but reduced to 50% by the OECD and the UNECE. To extend the classical model it was considered a joint approach based on monetary income and consumption, which was called characterized income. The approach applied was one of complementarity rather than substitutability

- This new model makes it possible to determine a new typology. On the

one hand, of those individuals whose consumption is linked to income, both those who are located below both thresholds (*area I*), and those individuals whose income and levels of expenditure are above the two lines should be considered (*area IV*). On the other hand, there are two types of situations for those individuals whose consumption is less linked to income: the first consists of those who have a low level of equivalent income but whose levels of monetary expenditure are above the consumption poverty line (*area II*), which could conceal situations of consumption financed by means of previously accumulated savings, debts, family assistance or shadow economy activities, which would mean an infra declaration of income and that such individuals were not really in a situation of material deprivation; the second would consist of individuals with a low level of monetary expenditure but whose income levels are medium-high since, in this case, they are to be found above the income poverty line (*area III*); these are normally individuals facing debt service, usually a mortgage linked to home purchase. The interpretation of the latter situation provides an additional reason for the incorporation, with monetary variables, of the expenses and/or income related with net wealth as carried out in this study. With the restriction of using empirical information based on official sources of statistics, the solution is to apply multivariate analysis techniques to reduce the problem to a single dimension in the EU-SILC area, in which a dual approach is used to validate the results: on the one hand, an external test based on whether the new monetary indicator, when consumption is incorporated, improves or worsens its degree of association with severe material deprivation compared with the classical model of monetary income is used; on the other hand, by means of an internal test at the micro level to check whether the results have a greater sensitivity, specificity and accuracy according to the statistical procedures based on the estimation of ROC curves.

- Principal components methodology, applied to the joint distribution of the equivalent monetary income and consumption, obtained from the HBS of 2017 in Spain, makes it possible to obtain formulae for the two standardized components. In the EU-SILC context income transformations are already applied within the classical paradigm, for example, by adding the imputed income associated with home ownership to equivalent income in an attempt to measure monetary poverty more accurately. It was identified the highest and/or most significant monetary expenditure item in the lowest income households. Information from HBS analyses in several European countries show that this item is spending on housing, at around 40%, which clearly conditions the ability of these households to afford other essential goods or services. This could be related to situations of severe material deprivation in low-income households, which justifies the consideration of difficulties to face the expenses derived from housing (whether rent or mortgages) as material deprivation variables.

- Characterized income was defined as disposable income minus the total expenditure on housing faced by the household. The new poverty risk estimator derived from this characterized income is contrasted with the classical monetary income estimator in the temporal and spatial dimensions. The external test of the temporal dimension of both variables covers the decade 2008-2017 and is resolved by means of the association measured via the correlation coefficients between severe material deprivation and the risk of poverty for this period, with better coefficients being obtained with the characterized income, Y_c. The internal test of the two variables considered over time, meanwhile, also throws up the result that, once again, the characterized income Y_c has a greater sensitivity, specificity and accuracy than the variable Y_d. As far as the spatial dimension is concerned, the external test is carried out with the results obtained in the seventeen Spanish autonomous communities at the NUTS level and leads to the conclusion that the characterized income Y_c increases the coefficient of determination and correlation with material deprivation. On the other hand, the internal test is carried out via the analysis of the sensitivity of the indicator to severe material deprivation; when using the characterized income Y_c, this value increases in most of the autonomous communities reaching the maximum intersection of 100% in some regions.

- Chapter 6 introduces an annex with a numeric and graphic review of the estimation of AROP sensitivity and specificity with regard to the SMD, that is useful to better understand the situation in countries ordered by GDP per capita in PPS and suggests the need to continuing the current research line in the direction of improving the specificity without compromising the sensitivity.

This work offers several additional options for future research. The calculation of yearly PPPs may extend the equivalence scales proposed in this doctoral thesis. Besides, incorporating severe material deprivation in the calculation of the rate of the population at risk of monetary poverty make it possible to improve the desirable consistency between both indicators and, in the case of the latter, by exogenous contrast variables, which would be useful to get a better understanding not only of cross-cutting situations but also to broaden the research on the dynamics of entries and exits into poverty. Finally, the application of the characterized income also allows to continue a line of research that seeks to improve the measurement of monetary poverty from a multidimensional perspective, by integrating the two visions of monetary income and consumption from a perspective of complementarity rather than substitutability. This in turn allows us to take an additional step forward towards an improved study and representation of the joint distribution of household income, consumption and wealth.

Chapter 6

ANNEX: SENSITIVITY AND SPECIFICITY.

6.1 Introduction.

In wealthier countries, and more precisely in the European Union, the relative measurement of the poverty is extended to an additional measurement rate by adding the people at-risk-of poverty to the people in social exclusion, represented by those severely materially deprived or living in a household with low work intensity. These are the three components of the so-called AROPE rate, an indicator used under the Europe 2020 strategy aiming to lift European people out of the risk of poverty and social exclusion. The AROPE represents the union of following three indicators:

- People with a equivalized income under the 60% of the median, after social transfers.
- People in severe material deprivation (lack of 4 out of 9 deprivations items).
- People aged 0-59 living in households where the adults (aged 18-59) work 20% or less of their total potential work.

At this stage, for the purpose of this doctoral research, using as starting point the intersection of the previous sub-populations, it's worth to calculate the sensitivity and sensibility of AROP with regard to the SMD in different European countries - EU, EFTA and Candidate Countries - since it provides a more clear picture of the coherent of the SMD and AROP results, not only in countries with very low SMD rate - usually countries with higher GDP per capita in PPS - but also in countries with higher SMD rate - usually countries with lower GDP per capita in PPS. The classical model, with constant p in the spatial and temporal dimensions, generates an important diversity of results by countries, offering generally modest results in terms of sensitivity and specificity.

6.2 Tables.

This section includes sensitivity, specificity and AROPE intersections of the following regions and countries:

- European Union: EU; Luxembourg; Ireland; The Netherlands; Austria; Denmark; Germany; Sweden; Belgium; Finland; United Kingdom; France; Italy; Malta; Spain; Czech Republic; Slovenia; Cyprus; Lithuania; Estonia; Portugal; Slovakia; Poland; Hungary; Greece; Latvia; Romania; Croatia; Bulgaria.
- EFTA countries: Switzerland; Norway.
- Candidate countries: Turkey; Former Yugoslav Republic of Macedonia; Serbia;

EUROPEAN UNION — GDP per capita in PPS = 100

Year	Se	Sp	$AROP \wedge \neg SMD \wedge \neg LWI$	$AROP \wedge \neg SMD \wedge LWI$	$AROP \wedge SMD \wedge \neg LWI$	$AROP \wedge SMD \wedge LWI$	$\neg AROP \wedge \neg SMD \wedge LWI$	$\neg AROP \wedge SMD \wedge \neg LWI$	$\neg AROP \wedge SMD \wedge LWI$	$\neg AROP \wedge \neg SMD \wedge \neg LWI$
2008	46.4	86.2	10.1	2.5	2.5	1.4	2.6	4.0	0.5	76.3
2009	47.6	86.4	9.9	2.6	2.5	1.4	2.6	3.8	0.5	76.7
2010	47.6	86.3	9.6	2.9	2.4	1.6	2.8	3.9	0.5	76.2
2011	48.3	86.2	9.7	2.9	2.6	1.7	2.9	4.0	0.6	75.7
2012	47.5	86.6	9.4	2.7	2.8	1.9	2.7	4.5	0.7	75.2
2013	47.4	86.5	9.4	2.8	2.7	1.9	2.8	4.3	0.8	75.4
2014	51.1	86.1	9.8	2.9	2.6	2.0	2.8	3.7	0.7	75.6
2015	53.1	85.8	10.2	2.9	2.5	1.8	2.7	3.2	0.6	76.2
2016	52.6	85.6	10.3	3.0	2.3	1.7	2.6	3.1	0.5	76.5
2017										

Source: Based on Eurostat database extracted on 25.09.18

Table 6.1: Sensitivity, specificity and intersections by year.

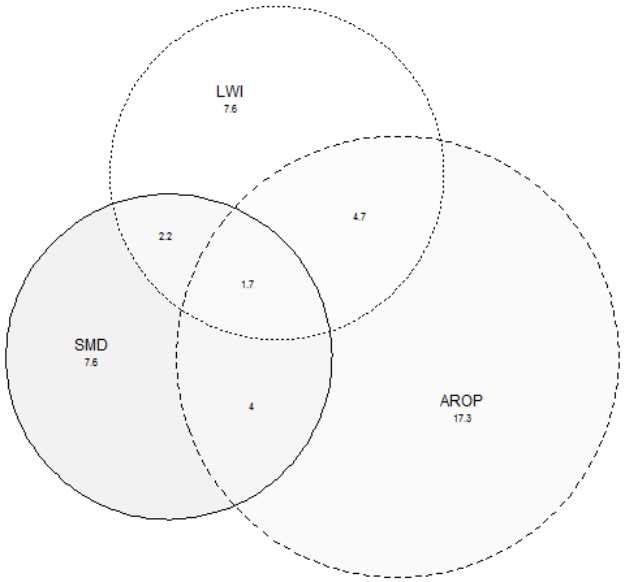
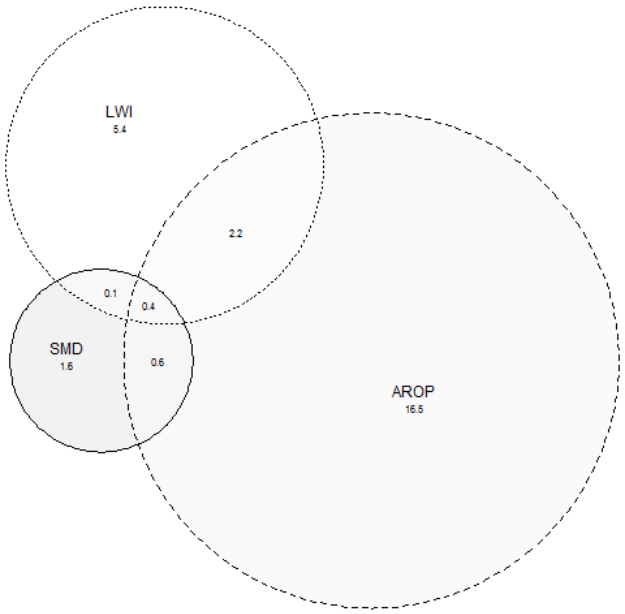


Figure 6.1: Instersection of subpopulations (AROP, SMD and LWI).

LUXEMBOURG							GDP per capita in PPS = 253			
Year	Se	Sp	$AROP \wedge \neg SMD \wedge \neg LWI$	$AROP \wedge \neg SMD \wedge LWI$	$AROP \wedge SMD \wedge \neg LWI$	$AROP \wedge SMD \wedge LWI$	$\neg AROP \wedge \neg SMD \wedge LWI$	$\neg AROP \wedge SMD \wedge \neg LWI$	$\neg AROP \wedge SMD \wedge LWI$	$\neg AROP \wedge \neg SMD \wedge \neg LWI$
2008	83.3	87.1	11.2	1.6	0.3	0.2	1.9	0.1	0.0	84.5
2009	80.0	85.8	11.9	2.1	0.5	0.3	2.7	0.2	0.0	82.2
2010	80.0	85.7	12.2	2.0	0.3	0.1	2.4	0.1	0.0	82.9
2011	58.3	87.0	11.0	1.8	0.6	0.1	2.7	0.4	0.1	83.2
2012	58.3	85.4	12.4	2.0	0.5	0.2	2.7	0.5	0.0	81.6
2013	55.6	84.8	12.4	2.5	0.7	0.3	2.4	0.6	0.2	81.0
2014	61.5	84.2	13.2	2.4	0.4	0.4	2.0	0.5	0.0	81.0
2015	68.4	85.7	12.3	1.7	1.1	0.2	2.5	0.5	0.1	81.5
2016	62.5	84.2	13.3	2.2	0.6	0.4	2.7	0.5	0.1	80.2
2017	:	:	:	:	:	:	:	:	:	:

Source: Based on Eurostat database extracted on 25.09.18

Sensitivity, specificity and intersections.

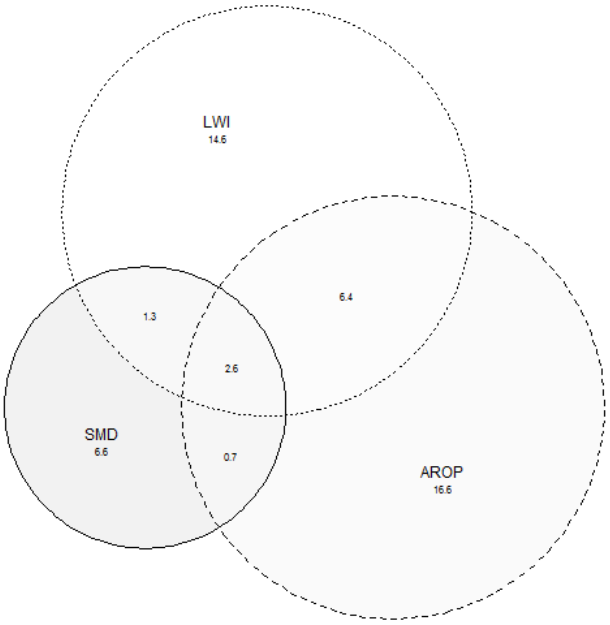


IRELAND — GDP per capita in PPS = 184

Year	Se	Sp	$AROP \wedge \neg SMD \wedge \neg LWI$	$AROP \wedge \neg SMD \wedge LWI$	$AROP \wedge SMD \wedge \neg LWI$	$AROP \wedge SMD \wedge LWI$	$\neg AROP \wedge \neg SMD \wedge LWI$	$\neg AROP \wedge SMD \wedge \neg LWI$	$\neg AROP \wedge SMD \wedge LWI$	$\neg AROP \wedge \neg SMD \wedge \neg LWI$
2008	39.3	85.9	9.5	3.8	0.6	1.6	4.8	2.1	1.3	76.3
2009	41.0	86.7	6.7	5.8	0.6	1.9	7.1	1.9	1.7	74.3
2010	43.1	86.4	6.7	6.1	0.4	2.1	8.8	1.3	2.0	72.7
2011	31.6	86.2	6.2	6.5	0.4	2.1	8.9	3.1	2.3	70.6
2012	34.3	85.3	6.1	7.2	1.0	2.4	7.2	4.2	2.3	69.7
2013	31.3	86.0	5.8	6.8	0.7	2.4	7.4	4.0	2.8	70.1
2014	37.3	85.7	6.7	6.4	0.6	2.5	6.0	3.2	2.0	72.5
2015	40.0	85.7	7.0	6.2	0.9	2.1	5.3	2.7	1.8	74.0
2016	50.0	85.8	6.9	6.4	0.7	2.6	4.3	2.0	1.3	75.8
2017	:	:	:	:	:	:	:	:	:	:

Source: Prepared by the author based on Eurostat database, extracted on 25.09.18

Sensitivity, specificity and intersections.

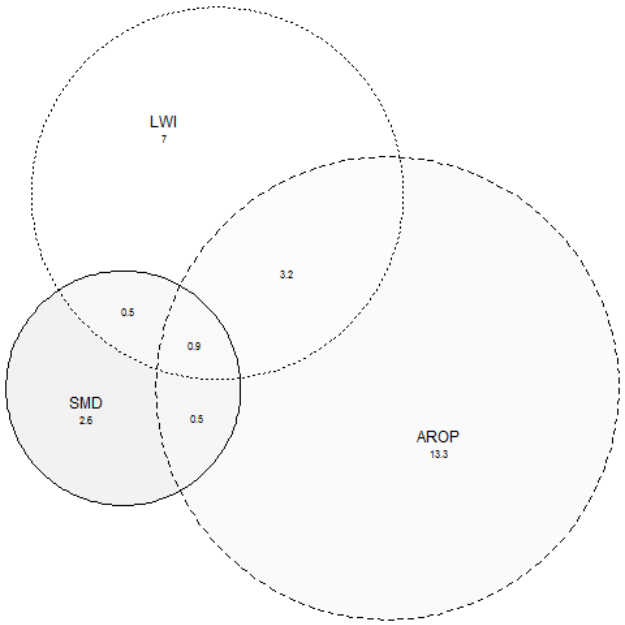


THE NETHERLANDS — GDP per capita in PPS = 128

Year	Se	Sp	$AROP \wedge \neg SMD \wedge \neg LWI$	$AROP \wedge \neg SMD \wedge LWI$	$AROP \wedge SMD \wedge \neg LWI$	$AROP \wedge SMD \wedge LWI$	$\neg AROP \wedge \neg SMD \wedge LWI$	$\neg AROP \wedge SMD \wedge \neg LWI$	$\neg AROP \wedge SMD \wedge LWI$	$\neg AROP \wedge \neg SMD \wedge \neg LWI$
2008	46.7	90.1	7.7	2.0	0.2	0.5	3.6	0.5	0.3	85.1
2009	40.0	89.4	7.7	2.8	0.3	0.3	3.3	0.6	0.3	84.9
2010	36.4	90.3	7.7	1.8	0.2	0.6	3.4	0.7	0.7	84.9
2011	54.2	90.2	7.6	2.0	0.4	0.9	3.6	0.8	0.3	84.3
2012	41.7	90.8	6.7	2.3	0.6	0.4	3.6	0.9	0.5	85.0
2013	26.9	89.9	7.3	2.5	0.3	0.4	3.6	1.2	0.7	84.1
2014	45.2	89.5	7.1	3.1	0.7	0.7	3.1	0.9	0.8	83.5
2015	44.0	89.2	7.5	3.0	0.4	0.7	3.3	0.7	0.7	83.6
2016	46.2	88.1	8.2	3.4	0.5	0.7	2.5	0.8	0.6	83.3
2017	53.8	87.8	8.7	3.2	0.5	0.9	2.4	0.7	0.5	83.0

Source: Based on Eurostat database extracted on 25.09.18

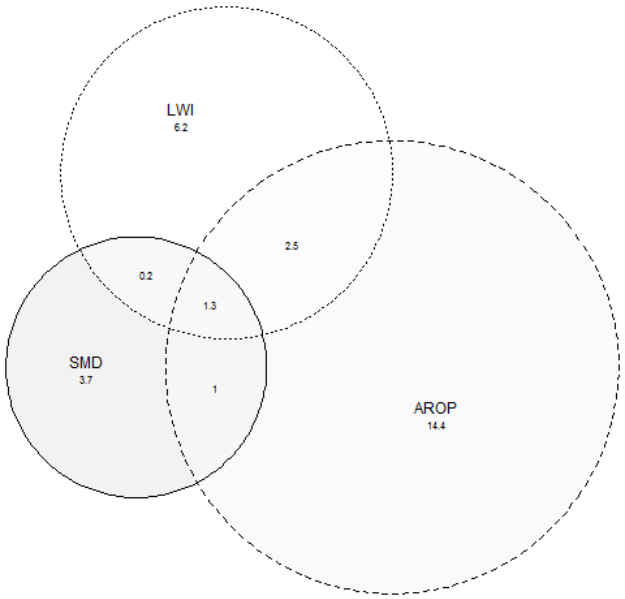
Sensitivity, specificity and intersections.



AUSTRIA							GDP per capita in PPS = 128			
Year	Se	Sp	$AROP \wedge \neg SMD \wedge \neg LWI$	$AROP \wedge \neg SMD \wedge LWI$	$AROP \wedge SMD \wedge \neg LWI$	$AROP \wedge SMD \wedge LWI$	$\neg AROP \wedge \neg SMD \wedge LWI$	$\neg AROP \wedge SMD \wedge \neg LWI$	$\neg AROP \wedge SMD \wedge LWI$	$\neg AROP \wedge \neg SMD \wedge \neg LWI$
2008	48.3	86.9	10.7	1.6	1.6	1.2	2.4	2.5	0.5	79.4
2009	52.2	87.3	10.4	1.7	1.3	1.1	2.4	1.9	0.3	80.9
2010	54.8	87.2	10.1	2.2	1.2	1.1	2.4	1.6	0.3	81.1
2011	55.0	87.2	10.1	2.2	1.1	1.1	2.8	1.4	0.4	80.8
2012	55.0	87.4	9.9	2.2	1.3	0.9	2.3	1.3	0.5	81.5
2013	53.5	87.5	10.2	1.8	1.1	1.2	2.5	1.5	0.5	81.2
2014	45.0	87.2	9.5	2.8	0.8	1.0	2.9	1.9	0.3	80.8
2015	54.3	87.7	9.7	2.2	1.1	0.8	2.8	1.2	0.4	81.7
2016	60.0	87.3	10.1	2.2	0.8	1.0	2.7	0.9	0.3	82.0
2017	62.2	87.4	9.6	2.5	1.0	1.3	2.2	1.2	0.2	81.9

Source: Based on Eurostat database extracted on 25.09.18

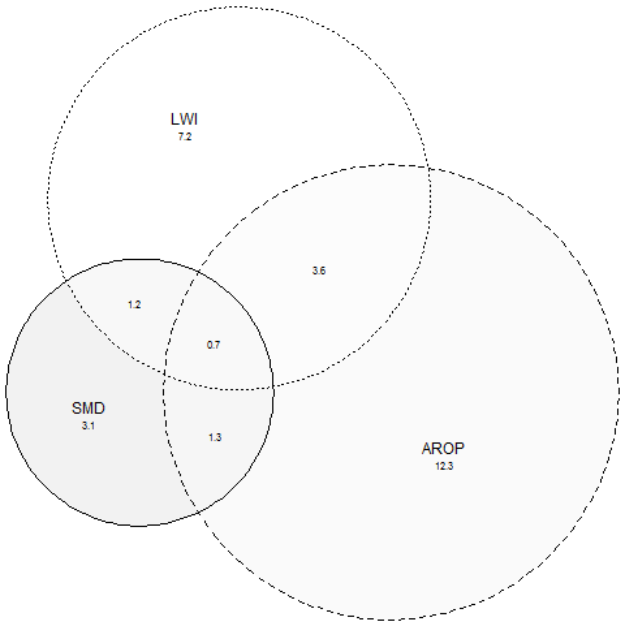
Sensitivity, specificity and intersections.



DENMARK							GDP per capita in PPS = 125			
Year	Se	Sp	$AROP \wedge \neg SMD \wedge \neg LWI$	$AROP \wedge \neg SMD \wedge LWI$	$AROP \wedge SMD \wedge \neg LWI$	$AROP \wedge SMD \wedge LWI$	$\neg AROP \wedge \neg SMD \wedge LWI$	$\neg AROP \wedge SMD \wedge \neg LWI$	$\neg AROP \wedge SMD \wedge LWI$	$\neg AROP \wedge \neg SMD \wedge \neg LWI$
2008	35.0	88.7	8.8	2.3	0.4	0.3	3.3	0.8	0.5	83.7
2009	45.5	87.6	9.7	2.4	0.5	0.5	3.3	0.8	0.4	82.4
2010	44.4	87.6	9.4	2.7	0.4	0.8	3.6	0.7	0.8	81.7
2011	39.1	88.5	8.4	2.8	0.6	0.3	4.1	0.8	0.6	82.4
2012	53.6	89.2	8.2	2.3	0.9	0.6	4.2	0.9	0.4	82.5
2013	40.5	89.2	7.8	2.6	0.6	0.9	4.3	1.3	0.9	81.7
2014	51.5	89.2	7.5	3.0	0.8	0.9	4.3	0.9	0.7	82.1
2015	52.8	89.3	7.4	2.9	1.1	0.8	3.8	0.9	0.8	82.3
2016	48.1	89.1	7.5	3.1	0.7	0.6	3.5	0.9	0.5	83.2
2017	41.9	88.6	8.1	2.9	0.6	0.7	3.1	1.3	0.5	82.8

Source: Based on Eurostat database extracted on 25.09.18

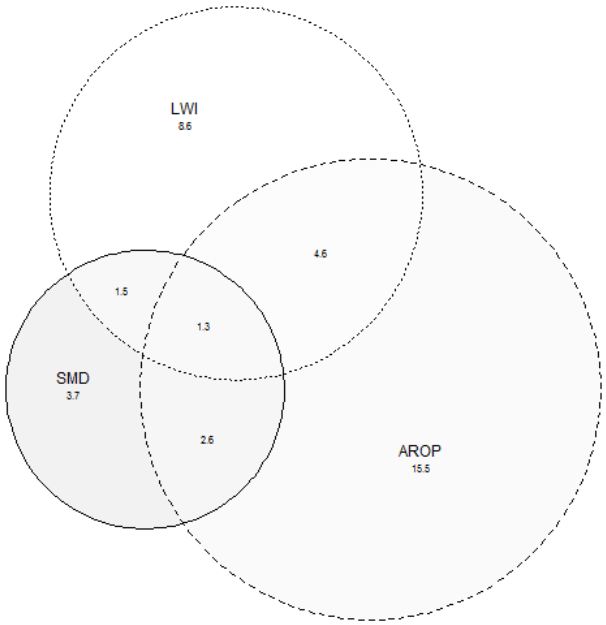
Sensitivity, specificity and intersections.



GERMANY							GDP per capita in PPS = 123			
Year	Se	Sp	$AROP \wedge \neg SMD \wedge \neg LWI$	$AROP \wedge \neg SMD \wedge LWI$	$AROP \wedge SMD \wedge \neg LWI$	$AROP \wedge SMD \wedge LWI$	$\neg AROP \wedge \neg SMD \wedge LWI$	$\neg AROP \wedge SMD \wedge \neg LWI$	$\neg AROP \wedge SMD \wedge LWI$	$\neg AROP \wedge \neg SMD \wedge \neg LWI$
2008	56.4	87.2	8.3	3.8	1.4	1.7	2.5	1.8	0.6	79.9
2009	57.4	86.9	8.8	3.6	1.3	1.8	2.1	1.8	0.5	80.0
2010	60.0	86.6	8.9	3.9	1.1	1.6	2.3	1.4	0.4	80.3
2011	66.7	87.1	8.5	3.7	1.7	1.9	2.3	1.5	0.3	80.1
2012	65.3	86.4	9.5	3.4	1.6	1.6	1.9	1.3	0.4	80.4
2013	61.1	86.5	9.7	3.1	1.8	1.5	2.2	1.6	0.5	79.7
2014	63.3	85.8	10.4	3.1	1.5	1.6	2.1	1.4	0.4	79.4
2015	68.2	85.8	10.4	3.2	1.4	1.6	1.9	1.1	0.3	80.0
2016	70.3	85.6	10.6	3.3	1.3	1.3	2.1	0.9	0.2	80.3
2017	:	:	:	:	:	:	:	:	:	:

Source: Based on Eurostat database extracted on 25.09.18

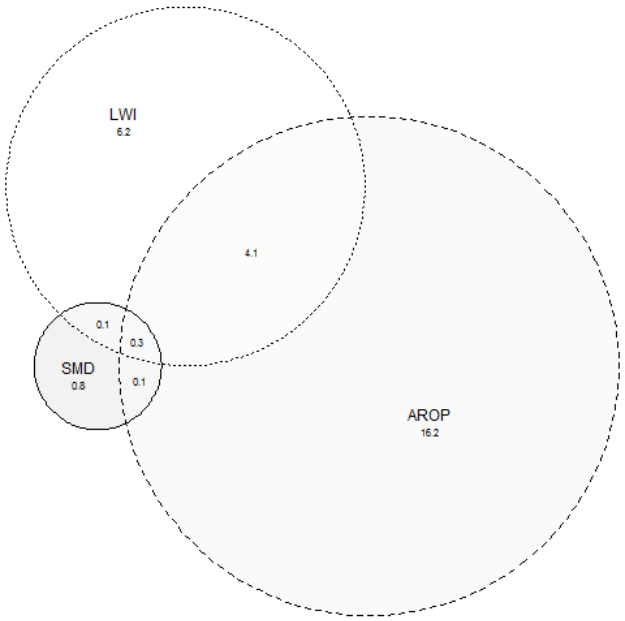
Sensitivity, specificity and intersections.



SWEDEN							GDP per capita in PPS = 122			
Year	Se	Sp	$AROP \wedge \neg SMD \wedge \neg LWI$	$AROP \wedge \neg SMD \wedge LWI$	$AROP \wedge SMD \wedge \neg LWI$	$AROP \wedge SMD \wedge LWI$	$\neg AROP \wedge \neg SMD \wedge LWI$	$\neg AROP \wedge SMD \wedge \neg LWI$	$\neg AROP \wedge SMD \wedge LWI$	$\neg AROP \wedge \neg SMD \wedge \neg LWI$
2008	44.4	87.1	10.4	2.3	0.4	0.4	2.3	0.8	0.2	83.3
2009	50.0	86.3	10.4	3.0	0.4	0.6	2.4	0.8	0.2	82.2
2010	52.6	85.9	10.5	3.3	0.3	0.7	2.0	0.7	0.2	82.3
2011	58.8	85.4	10.7	3.7	0.4	0.6	2.4	0.5	0.2	81.5
2012	61.1	85.6	10.8	3.3	0.4	0.7	1.8	0.6	0.1	82.3
2013	72.2	85.0	10.6	4.1	0.5	0.8	1.8	0.4	0.1	81.7
2014	50.0	84.8	11.0	4.0	0.2	0.3	2.1	0.3	0.2	81.8
2015	54.5	84.1	11.7	4.0	0.3	0.3	1.8	0.3	0.2	81.4
2016	50.0	84.1	11.7	4.1	0.1	0.3	1.7	0.3	0.1	81.7
2017	:	:	:	:	:	:	:	:	:	:

Source: Based on Eurostat database extracted on 25.09.18

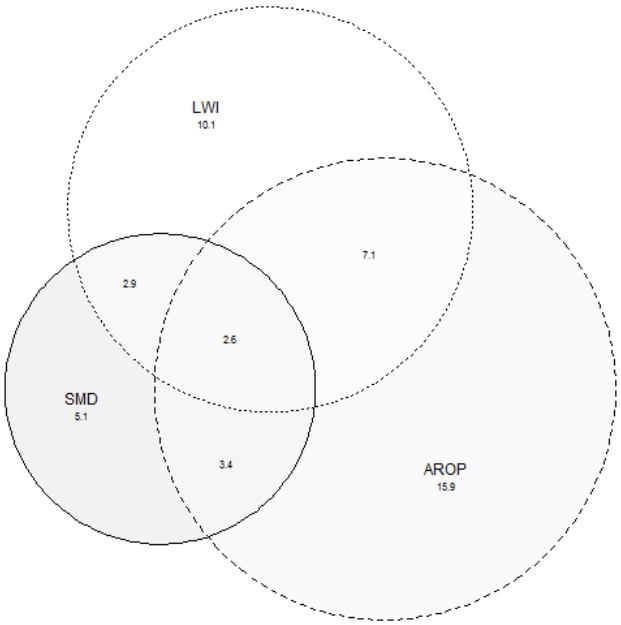
Sensitivity, specificity and intersections.



BELGIUM							GDP per capita in PPS = 117			
Year	Se	Sp	$AROP \wedge \neg SMD \wedge \neg LWI$	$AROP \wedge \neg SMD \wedge LWI$	$AROP \wedge SMD \wedge \neg LWI$	$AROP \wedge SMD \wedge LWI$	$\neg AROP \wedge \neg SMD \wedge LWI$	$\neg AROP \wedge SMD \wedge \neg LWI$	$\neg AROP \wedge SMD \wedge LWI$	$\neg AROP \wedge \neg SMD \wedge \neg LWI$
2008	58.9	87.9	8.4	3.0	1.3	2.0	3.7	1.9	0.4	79.2
2009	60.4	88.0	7.9	3.5	1.1	2.1	3.5	1.6	0.5	79.8
2010	57.6	88.0	8.0	3.3	1.2	2.2	3.7	1.9	0.6	79.2
2011	62.5	87.5	7.3	4.5	1.3	2.2	3.5	1.7	0.4	79.0
2012	56.3	87.4	7.9	3.9	1.0	2.6	3.6	2.1	0.7	78.4
2013	62.7	87.6	7.4	4.4	1.1	2.1	3.9	1.5	0.4	79.2
2014	64.4	87.6	7.3	4.4	1.1	2.7	3.7	1.5	0.6	78.8
2015	65.5	88.2	7.0	4.1	1.2	2.6	4.2	1.5	0.5	78.9
2016	68.5	87.6	7.1	4.6	1.1	2.6	3.5	1.4	0.3	79.3
2017	66.7	86.8	8.0	4.5	0.8	2.6	2.7	1.4	0.3	79.7

Source: Based on Eurostat database extracted on 25.09.18

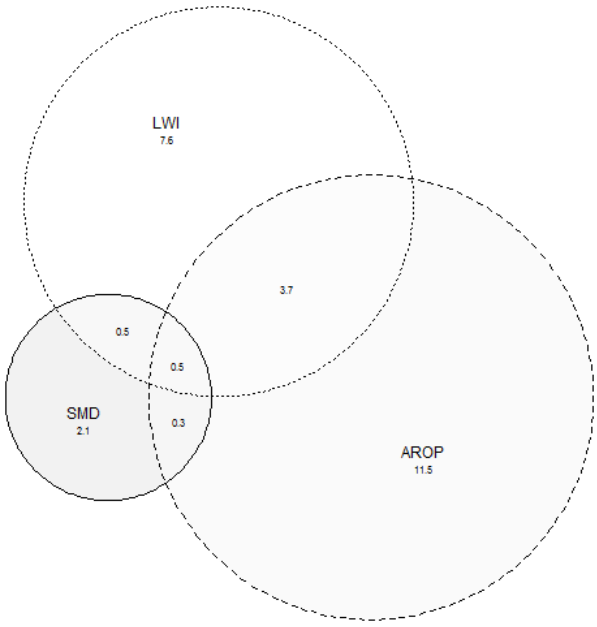
Sensitivity, specificity and intersections.



FINLAND							GDP per capita in PPS = 109			
Year	Se	Sp	$AROP \wedge \neg SMD \wedge \neg LWI$	$AROP \wedge \neg SMD \wedge LWI$	$AROP \wedge SMD \wedge \neg LWI$	$AROP \wedge SMD \wedge LWI$	$\neg AROP \wedge \neg SMD \wedge LWI$	$\neg AROP \wedge SMD \wedge \neg LWI$	$\neg AROP \wedge SMD \wedge LWI$	$\neg AROP \wedge \neg SMD \wedge \neg LWI$
2008	50.0	87.8	9.4	2.4	1.0	0.8	2.1	1.4	0.4	82.6
2009	60.7	87.4	9.3	2.9	0.6	1.1	1.9	0.8	0.3	83.1
2010	51.7	88.1	8.4	3.2	0.6	0.9	2.4	1.0	0.4	83.1
2011	46.9	87.4	8.8	3.4	0.6	0.9	2.5	1.1	0.6	82.1
2012	44.8	87.6	8.7	3.3	0.5	0.8	2.3	1.2	0.4	82.8
2013	41.7	89.0	8.0	2.7	0.4	0.6	2.7	0.9	0.5	84.0
2014	46.4	88.1	8.5	3.1	0.5	0.8	2.9	1.0	0.5	82.7
2015	50.0	88.3	7.9	3.5	0.5	0.6	3.2	0.7	0.4	83.2
2016	36.4	88.9	7.2	3.7	0.3	0.5	3.5	1.0	0.4	83.4
2017	38.1	89.1	7.0	3.7	0.3	0.5	2.9	0.8	0.5	84.3

Source: Based on Eurostat database extracted on 25.09.18

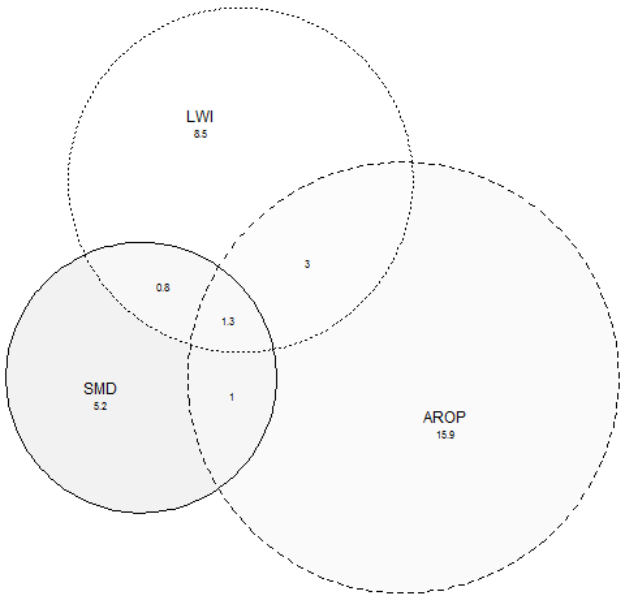
Sensitivity, specificity and intersections.



UNITED KINGDOM							GDP per capita in PPS = 105			
Year	Se	Sp	$AROP \wedge \neg SMD \wedge \neg LWI$	$AROP \wedge \neg SMD \wedge LWI$	$AROP \wedge SMD \wedge \neg LWI$	$AROP \wedge SMD \wedge LWI$	$\neg AROP \wedge \neg SMD \wedge LWI$	$\neg AROP \wedge SMD \wedge \neg LWI$	$\neg AROP \wedge SMD \wedge LWI$	$\neg AROP \wedge \neg SMD \wedge \neg LWI$
2008	48.9	82.7	12.6	3.9	1.0	1.2	2.2	1.5	0.8	76.8
2009	54.5	84.0	10.6	4.9	0.7	1.1	3.2	1.0	0.5	78.0
2010	41.7	84.1	10.7	4.4	0.6	1.4	3.2	1.7	1.1	76.8
2011	44.0	85.3	11.0	3.0	1.0	1.2	3.7	1.9	0.9	77.3
2012	44.9	86.4	9.8	2.7	1.6	1.9	3.8	2.7	1.6	75.9
2013	40.5	86.4	10.1	2.4	1.6	1.8	4.0	3.0	2.0	75.2
2014	46.6	85.6	10.8	2.5	1.3	2.1	3.4	2.5	1.4	75.9
2015	45.0	85.2	11.1	2.8	1.2	1.5	3.4	2.1	1.2	76.5
2016	44.2	85.7	10.6	3.0	1.0	1.3	3.4	2.1	0.8	77.8
2017	:	:	:	:	:	:	:	:	:	:

Source: Based on Eurostat database extracted on 25.09.18

Sensitivity, specificity and intersections.

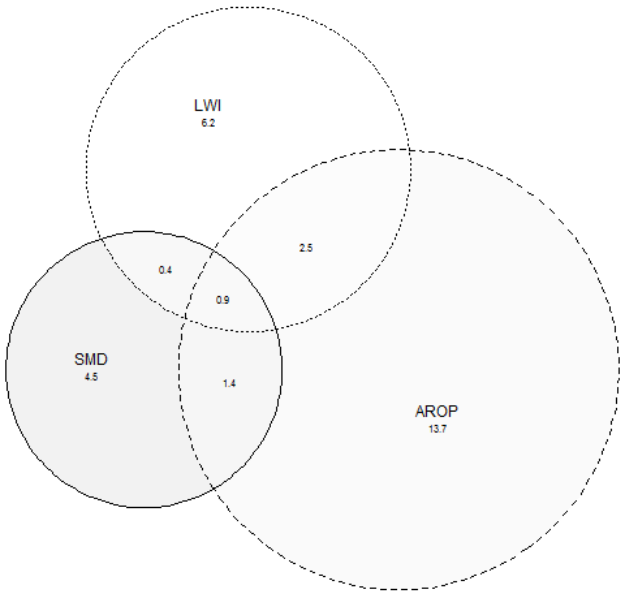


FRANCE — GDP per capita in PPS = 104

Year	Se	Sp	$AROP \wedge \neg SMD \wedge \neg LWI$	$AROP \wedge \neg SMD \wedge LWI$	$AROP \wedge SMD \wedge \neg LWI$	$AROP \wedge SMD \wedge LWI$	$\neg AROP \wedge \neg SMD \wedge LWI$	$\neg AROP \wedge SMD \wedge \neg LWI$	$\neg AROP \wedge SMD \wedge LWI$	$\neg AROP \wedge \neg SMD \wedge \neg LWI$
2008	42.6	89.2	8.1	2.1	1.1	1.2	2.9	2.6	0.5	81.5
2009	50.0	89.3	8.1	2.0	1.6	1.2	2.8	2.4	0.4	81.5
2010	51.7	89.0	7.8	2.6	1.5	1.5	3.1	2.5	0.3	80.8
2011	55.8	88.2	8.6	2.6	1.6	1.3	2.9	2.0	0.3	80.7
2012	49.1	87.9	9.2	2.3	1.3	1.3	2.3	2.3	0.4	80.9
2013	52.0	88.2	8.7	2.5	1.4	1.2	2.0	2.0	0.4	81.9
2014	52.1	88.7	8.3	2.5	1.3	1.2	2.9	1.8	0.5	81.5
2015	51.1	88.3	8.5	2.7	1.1	1.2	1.9	1.7	0.5	82.3
2016	51.1	88.1	8.9	2.5	1.4	0.9	2.4	1.8	0.4	81.8
2017	:	:	:	:	:	:	:	:	:	:

Source: Based on Eurostat database extracted on 25.09.18

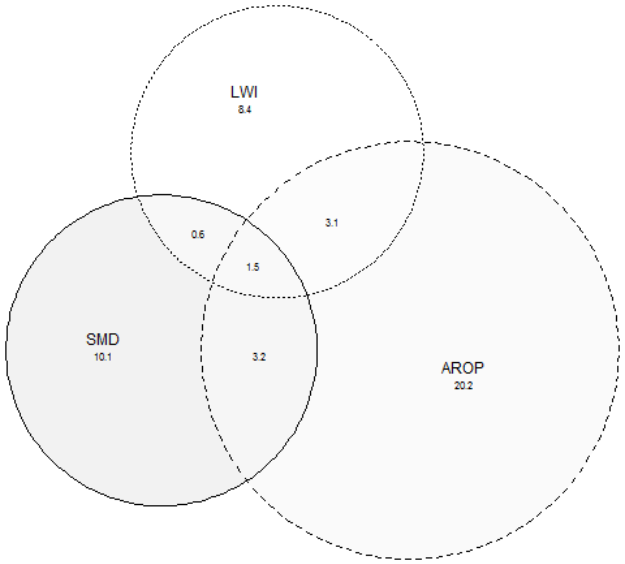
Sensitivity, specificity and intersections.



ITALY							GDP per capita in PPS = 96			
Year	Se	Sp	$AROP \wedge \neg SMD \wedge \neg LWI$	$AROP \wedge \neg SMD \wedge LWI$	$AROP \wedge SMD \wedge \neg LWI$	$AROP \wedge SMD \wedge LWI$	$\neg AROP \wedge \neg SMD \wedge LWI$	$\neg AROP \wedge SMD \wedge \neg LWI$	$\neg AROP \wedge SMD \wedge LWI$	$\neg AROP \wedge \neg SMD \wedge \neg LWI$
2008	53.3	83.9	12.1	2.8	2.6	1.4	3.1	3.1	0.4	74.5
2009	52.1	84.3	12.4	2.2	2.6	1.2	3.0	3.0	0.5	75.1
2010	55.4	84.1	11.8	2.9	2.5	1.6	3.0	3.0	0.3	75.0
2011	49.1	83.8	11.6	2.8	3.8	1.7	2.7	5.1	0.6	71.9
2012	46.9	85.1	10.5	2.2	4.7	2.1	2.7	7.0	0.7	70.1
2013	48.4	84.8	10.7	2.6	3.9	2.1	2.9	5.7	0.7	71.5
2014	48.7	84.4	10.8	3.0	3.4	2.3	2.9	5.3	0.7	71.7
2015	49.1	83.8	11.7	2.6	3.4	2.2	2.9	5.1	0.7	71.3
2016	46.7	82.9	11.8	3.2	3.3	2.3	2.9	5.6	0.8	70.0
2017	46.5	82.7	12.4	3.1	3.2	1.5	3.2	4.8	0.6	71.1

Source: Based on Eurostat database extracted on 25.09.18

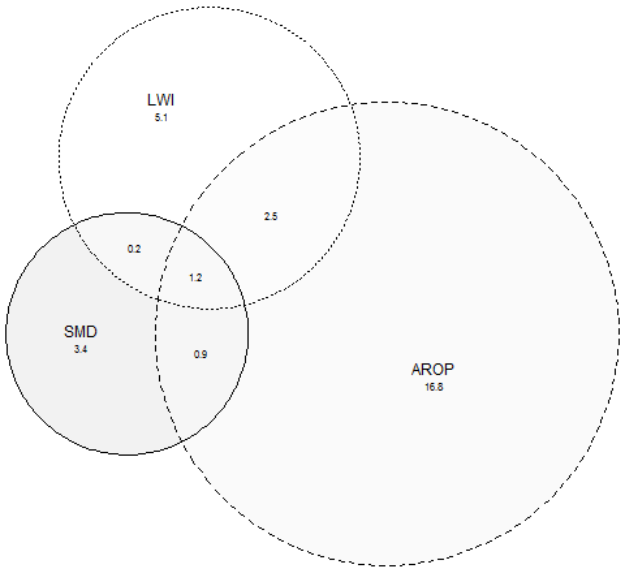
Sensitivity, specificity and intersections.



MALTA										GDP per capita in PPS = 96
Year	Se	Sp	$AROP \wedge \neg SMD \wedge \neg LWI$	$AROP \wedge \neg SMD \wedge LWI$	$AROP \wedge SMD \wedge \neg LWI$	$AROP \wedge SMD \wedge LWI$	$\neg AROP \wedge \neg SMD \wedge LWI$	$\neg AROP \wedge SMD \wedge \neg LWI$	$\neg AROP \wedge SMD \wedge LWI$	$\neg AROP \wedge \neg SMD \wedge \neg LWI$
2008	39.5	85.9	10.6	2.9	0.4	1.3	2.2	2.2	0.4	79.9
2009	46.0	86.6	9.9	2.8	1.0	1.3	2.7	2.1	0.6	79.7
2010	46.2	86.6	9.6	2.9	1.4	1.6	2.3	3.1	0.4	78.8
2011	39.4	86.1	10.1	2.9	1.3	1.3	2.4	3.7	0.3	77.9
2012	34.8	86.8	9.5	2.5	1.3	1.9	2.0	5.4	0.6	76.9
2013	34.7	86.4	9.9	2.4	1.3	2.0	2.1	5.9	0.3	76.0
2014	41.2	87.1	9.3	2.3	1.8	2.4	1.9	5.3	0.7	76.2
2015	46.9	86.3	10.0	2.6	1.6	2.2	1.7	3.9	0.4	77.6
2016	52.3	85.1	11.8	2.4	0.9	1.4	1.5	1.9	0.2	79.9
2017	61.8	84.8	12.2	2.5	0.9	1.2	1.2	1.1	0.2	80.8

Source: Based on Eurostat database extracted on 25.09.18

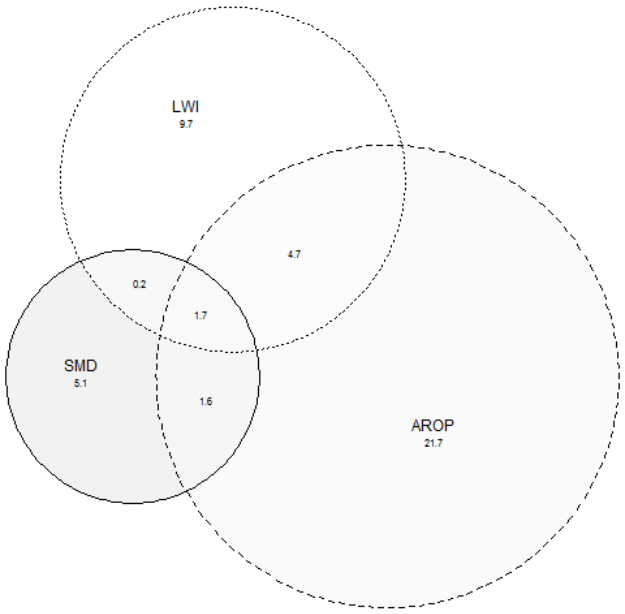
Sensitivity, specificity and intersections.



SPAIN							GDP per capita in PPS = 92			
Year	Se	Sp	$AROP \wedge \neg SMD \wedge \neg LWI$	$AROP \wedge \neg SMD \wedge LWI$	$AROP \wedge SMD \wedge \neg LWI$	$AROP \wedge SMD \wedge LWI$	$\neg AROP \wedge \neg SMD \wedge LWI$	$\neg AROP \wedge SMD \wedge \neg LWI$	$\neg AROP \wedge SMD \wedge LWI$	$\neg AROP \wedge \neg SMD \wedge \neg LWI$
2008	58.3	81.6	15.6	2.1	1.6	0.5	2.4	1.4	0.1	76.2
2009	58.7	81.5	15.3	2.4	1.8	0.9	2.5	1.7	0.2	75.3
2010	59.2	81.2	14.0	3.9	1.9	1.0	3.4	1.8	0.2	73.9
2011	54.3	80.9	13.6	4.6	1.1	1.4	4.0	1.7	0.4	73.3
2012	57.6	81.4	12.6	4.9	1.5	1.9	4.0	2.1	0.4	72.8
2013	56.5	82.0	11.3	5.6	1.6	1.9	4.2	2.2	0.5	72.7
2014	62.9	80.8	11.8	6.0	2.1	2.3	4.3	2.1	0.5	70.8
2015	61.5	80.7	13.1	5.0	1.8	2.2	4.1	2.0	0.5	71.4
2016	69.0	80.6	13.2	5.1	1.9	2.1	3.8	1.5	0.3	72.1
2017	64.7	80.6	13.7	4.7	1.6	1.7	3.1	1.6	0.2	73.4

Source: Based on Eurostat database extracted on 25.09.18

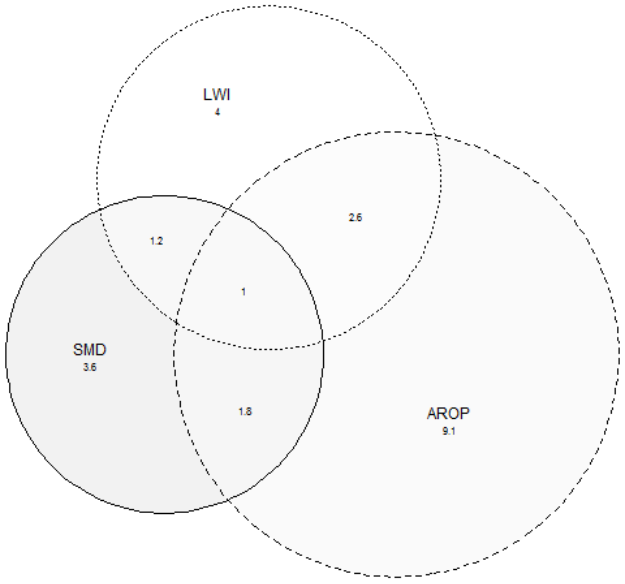
Sensitivity, specificity and intersections.



CZECH REPUBLIC							GDP per capita in PPS = 92			
Year	Se	Sp	$AROP \wedge \neg SMD \wedge \neg LWI$	$AROP \wedge \neg SMD \wedge LWI$	$AROP \wedge SMD \wedge \neg LWI$	$AROP \wedge SMD \wedge LWI$	$\neg AROP \wedge \neg SMD \wedge LWI$	$\neg AROP \wedge SMD \wedge \neg LWI$	$\neg AROP \wedge SMD \wedge LWI$	$\neg AROP \wedge \neg SMD \wedge \neg LWI$
2008	41.2	93.2	4.5	1.8	1.4	1.4	2.2	3.7	0.3	84.7
2009	41.0	93.5	4.5	1.6	1.4	1.1	1.8	3.4	0.2	86.0
2010	43.5	93.3	4.8	1.5	1.3	1.4	1.9	3.3	0.2	85.6
2011	38.7	92.1	5.7	1.7	1.1	1.3	1.7	3.4	0.4	84.7
2012	42.4	92.7	5.2	1.6	1.6	1.2	1.9	3.4	0.4	84.6
2013	40.9	93.6	4.5	1.5	1.3	1.4	2.1	3.6	0.3	85.4
2014	46.3	92.9	4.7	1.9	1.1	2.0	1.5	3.2	0.4	85.2
2015	50.9	92.8	5.2	1.6	1.1	1.7	1.5	2.5	0.2	86.0
2016	52.1	92.4	5.2	2.0	1.0	1.5	1.3	2.2	0.1	86.7
2017	50.0	92.4	5.7	1.6	0.8	1.0	1.2	1.6	0.2	87.8

Source: Based on Eurostat database extracted on 25.09.18

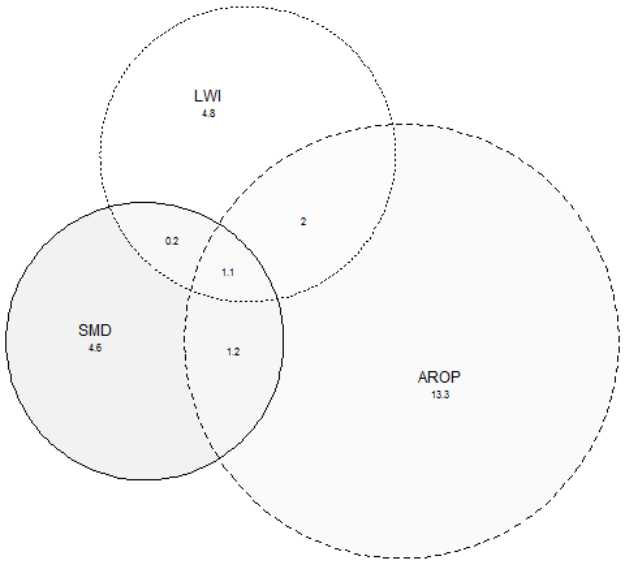
Sensitivity, specificity and intersections.



SLOVENIA							GDP per capita in PPS = 85			
Year	Se	Sp	$AROP \wedge \neg SMD \wedge \neg LWI$	$AROP \wedge \neg SMD \wedge LWI$	$AROP \wedge SMD \wedge \neg LWI$	$AROP \wedge SMD \wedge LWI$	$\neg AROP \wedge \neg SMD \wedge LWI$	$\neg AROP \wedge SMD \wedge \neg LWI$	$\neg AROP \wedge SMD \wedge LWI$	$\neg AROP \wedge \neg SMD \wedge \neg LWI$
2008	40.9	89.7	7.7	1.9	1.6	1.1	2.2	3.7	0.2	81.5
2009	37.7	90.4	7.7	1.3	1.4	0.9	2.1	3.6	0.2	82.9
2010	44.1	89.3	8.0	2.1	1.5	1.1	2.3	3.2	0.1	81.7
2011	43.3	88.3	8.7	2.3	1.4	1.2	2.2	3.1	0.3	80.7
2012	47.0	88.9	8.5	1.9	1.8	1.3	2.5	3.4	0.1	80.4
2013	49.3	88.0	9.2	2.0	1.8	1.5	2.5	3.2	0.2	79.6
2014	47.0	87.8	8.6	2.8	1.7	1.4	2.4	3.3	0.2	79.6
2015	49.2	87.9	9.3	2.1	1.4	1.5	1.9	2.8	0.2	80.8
2016	50.0	88.2	8.8	2.4	1.5	1.2	1.9	2.5	0.2	81.6
2017	50.0	88.5	9.0	2.0	1.2	1.1	1.5	2.1	0.2	82.9

Source: Based on Eurostat database extracted on 25.09.18

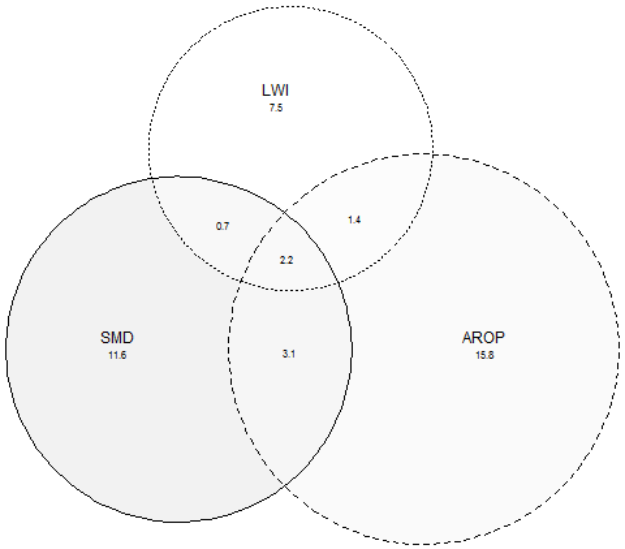
Sensitivity, specificity and intersections.



CYPRUS							GDP per capita in PPS = 84			
Year	Se	Sp	$AROP \wedge \neg SMD \wedge \neg LWI$	$AROP \wedge \neg SMD \wedge LWI$	$AROP \wedge SMD \wedge \neg LWI$	$AROP \wedge SMD \wedge LWI$	$\neg AROP \wedge \neg SMD \wedge LWI$	$\neg AROP \wedge SMD \wedge \neg LWI$	$\neg AROP \wedge SMD \wedge LWI$	$\neg AROP \wedge \neg SMD \wedge \neg LWI$
2008	35.2	86.0	11.5	1.2	2.5	0.7	1.4	5.4	0.5	76.7
2009	31.9	86.0	11.8	0.9	2.3	0.7	1.3	6.1	0.3	76.5
2010	34.5	86.7	10.3	1.5	3.4	0.5	1.7	7.0	0.4	75.4
2011	31.4	87.4	9.8	1.3	3.2	0.5	1.8	7.7	0.4	75.4
2012	31.1	88.2	8.5	1.5	3.8	0.9	2.1	9.6	0.8	72.9
2013	37.0	88.8	8.0	1.4	4.0	2.0	2.3	9.4	0.8	72.2
2014	32.7	88.9	7.2	2.2	3.1	1.9	2.7	9.1	1.2	72.6
2015	37.0	87.5	8.0	2.6	3.7	2.0	3.0	8.3	1.4	71.1
2016	38.2	87.5	8.4	2.4	3.3	1.9	3.2	7.5	0.9	72.3
2017	45.7	88.1	9.1	1.4	3.1	2.2	3.2	5.6	0.7	74.8

Source: Based on Eurostat database extracted on 25.09.18

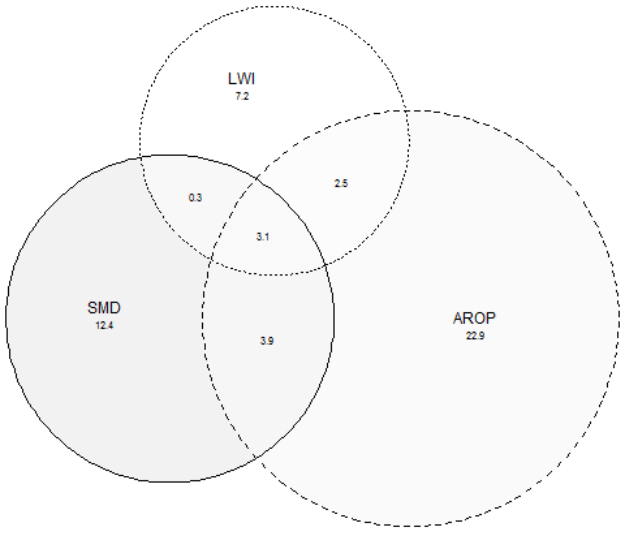
Sensitivity, specificity and intersections.



LITHUANIA							GDP per capita in PPS = 78			
Year	Se	Sp	$AROP \wedge \neg SMD \wedge \neg LWI$	$AROP \wedge \neg SMD \wedge LWI$	$AROP \wedge SMD \wedge \neg LWI$	$AROP \wedge SMD \wedge LWI$	$\neg AROP \wedge \neg SMD \wedge LWI$	$\neg AROP \wedge SMD \wedge \neg LWI$	$\neg AROP \wedge SMD \wedge LWI$	$\neg AROP \wedge \neg SMD \wedge \neg LWI$
2008	50.4	83.3	12.8	1.8	4.8	1.5	1.2	6.0	0.2	71.7
2009	46.2	84.5	11.3	1.8	5.0	2.2	0.9	7.7	0.7	70.4
2010	39.2	84.0	10.6	2.2	5.2	2.6	1.4	10.9	1.2	66.0
2011	37.9	85.2	8.9	3.1	4.0	3.2	2.1	10.5	1.3	66.9
2012	38.4	86.4	8.8	2.1	4.1	3.5	1.7	10.9	1.3	67.5
2013	48.4	84.8	10.0	2.8	5.2	2.6	2.0	7.5	0.8	69.2
2014	50.0	85.6	9.9	2.5	4.6	2.2	1.3	6.2	0.6	72.7
2015	57.6	83.5	11.4	2.8	5.6	2.4	1.2	5.4	0.5	70.7
2016	47.4	82.1	12.2	3.3	3.8	2.6	1.2	6.5	0.6	69.9
2017	56.5	81.8	13.4	2.5	3.9	3.1	1.3	5.1	0.3	70.4

Source: Based on Eurostat database extracted on 25.09.18

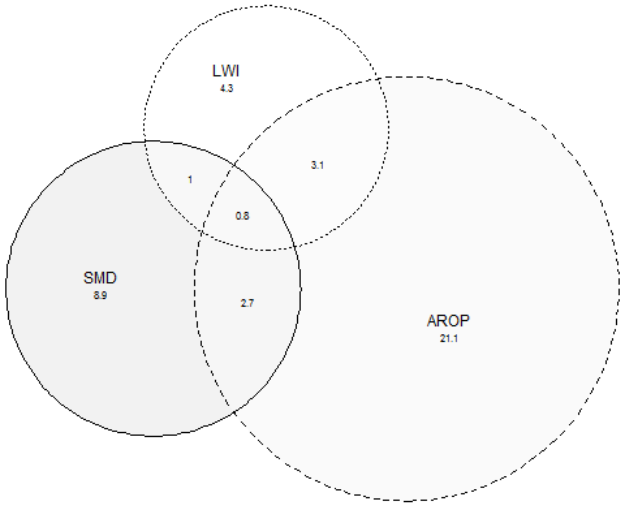
Sensitivity, specificity and intersections.



ESTONIA										GDP per capita in PPS = 77
Year	Se	Sp	$AROP \wedge \neg SMD \wedge \neg LWI$	$AROP \wedge \neg SMD \wedge LWI$	$AROP \wedge SMD \wedge \neg LWI$	$AROP \wedge SMD \wedge LWI$	$\neg AROP \wedge \neg SMD \wedge LWI$	$\neg AROP \wedge SMD \wedge \neg LWI$	$\neg AROP \wedge SMD \wedge LWI$	$\neg AROP \wedge \neg SMD \wedge \neg LWI$
2008	63.3	82.9	14.1	2.2	2.0	1.1	0.6	1.7	0.1	78.2
2009	56.5	82.7	14.4	1.8	2.1	1.4	1.1	2.6	0.1	76.6
2010	52.8	87.8	8.2	2.9	2.8	1.9	1.7	3.8	0.4	78.3
2011	54.7	86.1	9.5	3.2	2.3	2.4	1.7	3.5	0.4	76.9
2012	54.3	86.3	9.7	2.7	2.7	2.4	1.6	4.0	0.3	76.6
2013	50.0	83.9	11.6	3.3	2.0	1.8	1.1	3.5	0.3	76.5
2014	56.5	80.5	15.8	2.5	2.0	1.5	1.5	2.5	0.2	74.0
2015	63.6	80.3	16.0	2.8	1.8	1.0	1.0	1.5	0.1	75.8
2016	59.6	80.2	16.6	2.3	1.8	1.0	0.8	1.8	0.1	75.6
2017	65.9	80.8	16.1	2.3	1.9	0.8	1.0	1.2	0.2	76.6

Source: Based on Eurostat database extracted on 25.09.18

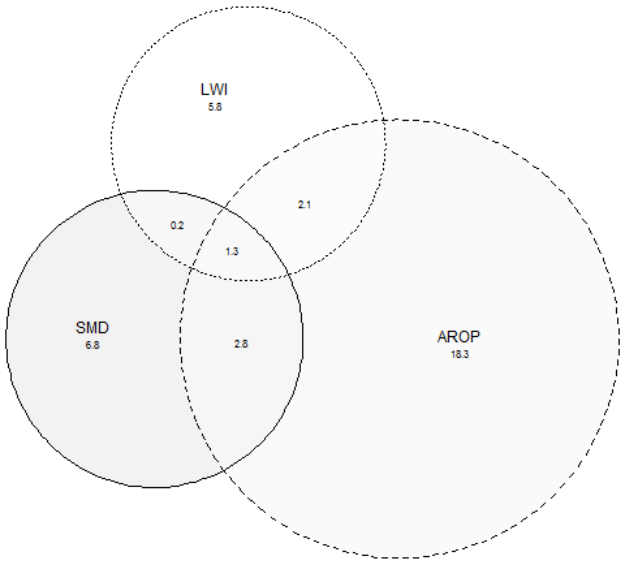
Sensitivity, specificity and intersections.



PORTUGAL							GDP per capita in PPS = 77			
Year	Se	Sp	$AROP \wedge \neg SMD \wedge \neg LWI$	$AROP \wedge \neg SMD \wedge LWI$	$AROP \wedge SMD \wedge \neg LWI$	$AROP \wedge SMD \wedge LWI$	$\neg AROP \wedge \neg SMD \wedge LWI$	$\neg AROP \wedge SMD \wedge \neg LWI$	$\neg AROP \wedge SMD \wedge LWI$	$\neg AROP \wedge \neg SMD \wedge \neg LWI$
2008	44.3	84.3	12.8	1.4	3.2	1.1	2.1	5.2	0.2	74.0
2009	45.1	84.8	11.9	1.9	2.9	1.2	2.1	4.7	0.3	75.1
2010	43.3	84.7	11.8	2.1	2.2	1.7	2.4	4.7	0.4	74.7
2011	50.6	85.0	11.7	2.1	2.9	1.3	2.4	3.6	0.5	75.6
2012	48.8	85.0	11.2	2.5	2.6	1.6	3.0	4.0	0.4	74.7
2013	50.0	85.2	10.1	3.1	3.2	2.2	3.2	5.0	0.4	72.5
2014	54.7	84.8	10.7	2.9	3.4	2.4	3.2	4.4	0.4	72.5
2015	56.3	84.4	11.4	2.7	3.5	1.9	3.0	3.8	0.4	73.4
2016	56.6	84.5	11.9	2.3	3.3	1.4	2.5	3.2	0.4	74.9
2017	60.3	84.7	12.1	2.1	2.8	1.3	2.2	2.5	0.2	76.7

Source: Based on Eurostat database extracted on 25.09.18

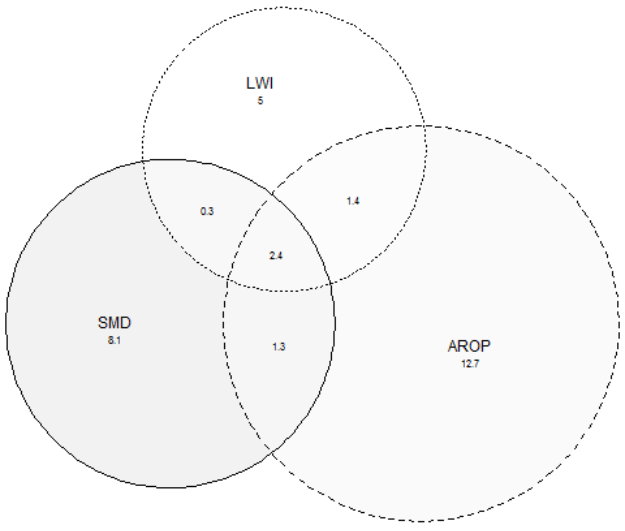
Sensitivity, specificity and intersections.



SLOVAKIA										GDP per capita in PPS = 77
Year	Se	Sp	$AROP \wedge \neg SMD \wedge \neg LWI$	$AROP \wedge \neg SMD \wedge LWI$	$AROP \wedge SMD \wedge \neg LWI$	$AROP \wedge SMD \wedge LWI$	$\neg AROP \wedge \neg SMD \wedge LWI$	$\neg AROP \wedge SMD \wedge \neg LWI$	$\neg AROP \wedge SMD \wedge LWI$	$\neg AROP \wedge \neg SMD \wedge \neg LWI$
2008	29.9	91.7	6.4	0.9	2.2	1.3	1.5	7.7	0.5	79.4
2009	36.0	92.2	6.2	0.7	2.1	1.9	1.5	6.8	0.3	80.4
2010	40.4	91.6	6.1	1.3	1.9	2.7	1.8	6.1	0.7	79.4
2011	41.9	90.5	7.0	1.5	1.9	2.5	1.5	5.5	0.6	79.4
2012	44.2	90.4	7.3	1.3	1.9	2.7	1.4	5.5	0.3	79.5
2013	45.6	90.9	6.6	1.6	1.8	2.9	1.3	5.3	0.3	80.2
2014	51.0	91.7	6.2	1.3	1.9	3.1	1.0	4.6	0.2	81.6
2015	46.2	91.1	6.6	1.5	1.6	2.6	1.3	4.6	0.3	81.6
2016	45.7	90.2	7.6	1.4	1.3	2.4	0.9	4.1	0.3	81.9
2017	:	:	:	:	:	:	:	:	:	:

Source: Based on Eurostat database extracted on 25.09.18

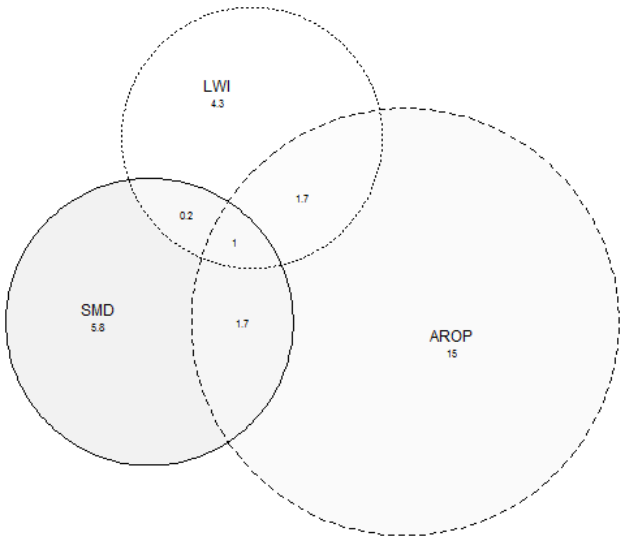
Sensitivity, specificity and intersections.



POLAND							GDP per capita in PPS = 70			
Year	Se	Sp	$AROP \wedge \neg SMD \wedge \neg LWI$	$AROP \wedge \neg SMD \wedge LWI$	$AROP \wedge SMD \wedge \neg LWI$	$AROP \wedge SMD \wedge LWI$	$\neg AROP \wedge \neg SMD \wedge LWI$	$\neg AROP \wedge SMD \wedge \neg LWI$	$\neg AROP \wedge SMD \wedge LWI$	$\neg AROP \wedge \neg SMD \wedge \neg LWI$
2008	36.9	87.5	9.0	1.3	4.7	1.9	2.4	10.4	0.9	69.5
2009	42.0	87.3	9.5	1.3	4.6	1.7	2.1	8.2	0.5	72.2
2010	43.3	86.7	9.9	1.5	4.3	1.8	2.1	7.6	0.4	72.2
2011	42.3	86.0	10.6	1.6	4.0	1.5	2.0	7.1	0.4	72.8
2012	41.8	86.8	9.9	1.5	3.9	1.7	1.8	7.4	0.4	73.3
2013	42.9	86.3	10.5	1.6	3.4	1.7	1.8	6.3	0.5	74.2
2014	45.2	86.3	10.6	1.7	3.2	1.5	2.1	5.2	0.5	75.3
2015	51.2	85.3	11.6	1.9	2.8	1.4	1.8	3.7	0.3	76.6
2016	53.7	85.3	11.7	2.0	2.4	1.2	1.5	2.9	0.2	78.1
2017	46.6	86.9	10.6	1.7	1.7	1.0	1.4	2.9	0.2	80.5

Source: Based on Eurostat database extracted on 25.09.18

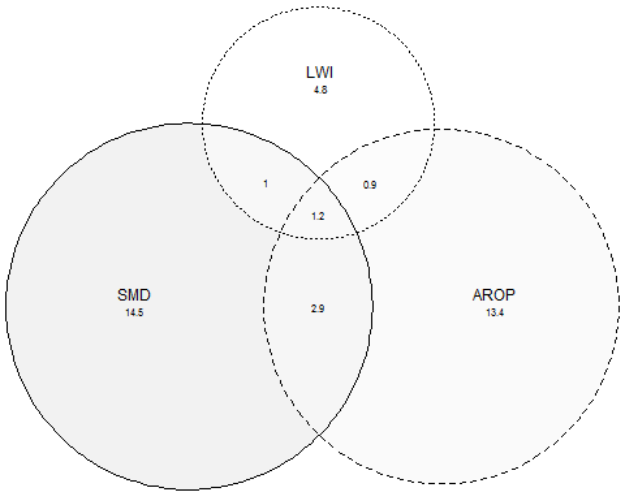
Sensitivity, specificity and intersections.



HUNGARY							GDP per capita in PPS = 68			
Year	Se	Sp	$AROP \wedge \neg SMD \wedge \neg LWI$	$AROP \wedge \neg SMD \wedge LWI$	$AROP \wedge SMD \wedge \neg LWI$	$AROP \wedge SMD \wedge LWI$	$\neg AROP \wedge \neg SMD \wedge LWI$	$\neg AROP \wedge SMD \wedge \neg LWI$	$\neg AROP \wedge SMD \wedge LWI$	$\neg AROP \wedge \neg SMD \wedge \neg LWI$
2008	30.7	91.5	5.0	2.0	2.8	2.7	3.4	10.9	1.5	71.8
2009	30.5	92.2	4.7	1.5	3.4	2.8	3.0	12.6	1.5	70.4
2010	30.9	92.7	4.2	1.5	3.3	3.4	2.6	13.3	1.7	70.1
2011	37.3	93.1	4.0	1.3	4.6	4.1	2.8	13.1	1.5	68.5
2012	36.6	93.6	3.4	1.3	4.8	4.8	2.5	14.9	1.7	66.5
2013	37.1	93.6	3.4	1.2	5.4	4.9	2.4	15.7	1.8	65.2
2014	39.2	92.8	4.1	1.4	4.7	4.7	2.3	13.3	1.3	68.2
2015	41.8	91.4	5.9	1.0	5.1	3.0	2.0	10.2	1.1	71.8
2016	38.9	90.3	6.8	1.3	4.4	1.9	1.9	9.0	0.9	73.7
2017	28.3	89.1	8.4	0.9	2.9	1.2	1.7	9.4	1.0	74.4

Source: Based on Eurostat database extracted on 25.09.18

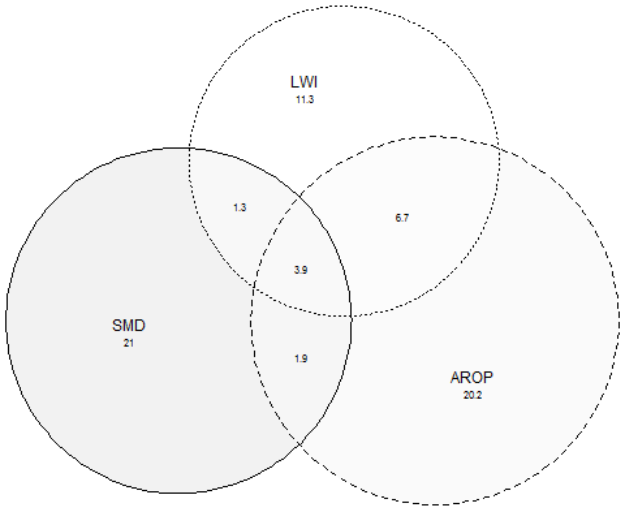
Sensitivity, specificity and intersections.



GREECE										GDP per capita in PPS = 67
Year	Se	Sp	$AROP \wedge \neg SMD \wedge \neg LWI$	$AROP \wedge \neg SMD \wedge LWI$	$AROP \wedge SMD \wedge \neg LWI$	$AROP \wedge SMD \wedge LWI$	$\neg AROP \wedge \neg SMD \wedge LWI$	$\neg AROP \wedge SMD \wedge \neg LWI$	$\neg AROP \wedge SMD \wedge LWI$	$\neg AROP \wedge \neg SMD \wedge \neg LWI$
2008	56.8	84.5	12.7	1.1	5.2	1.1	3.1	4.5	0.3	71.9
2009	55.5	84.6	12.7	1.0	5.3	0.8	2.9	4.7	0.2	72.4
2010	61.2	85.3	11.8	1.2	6.1	1.0	3.1	4.1	0.4	72.3
2011	59.9	85.4	10.3	2.1	6.3	2.8	3.5	5.5	0.6	69.0
2012	60.5	86.0	8.9	2.4	8.2	3.6	3.8	6.9	0.8	65.4
2013	60.8	86.4	8.1	2.7	7.2	5.2	4.6	7.0	1.0	64.3
2014	59.3	88.2	7.3	2.0	8.4	4.4	5.2	7.8	1.0	64.0
2015	57.0	88.8	7.0	1.7	8.0	4.6	4.8	8.4	1.1	64.3
2016	58.0	89.4	6.7	1.5	8.0	5.0	4.9	8.4	1.0	64.4
2017	50.5	87.8	7.7	1.9	6.7	3.9	4.2	9.1	1.3	65.2

Source: Based on Eurostat database extracted on 25.09.18

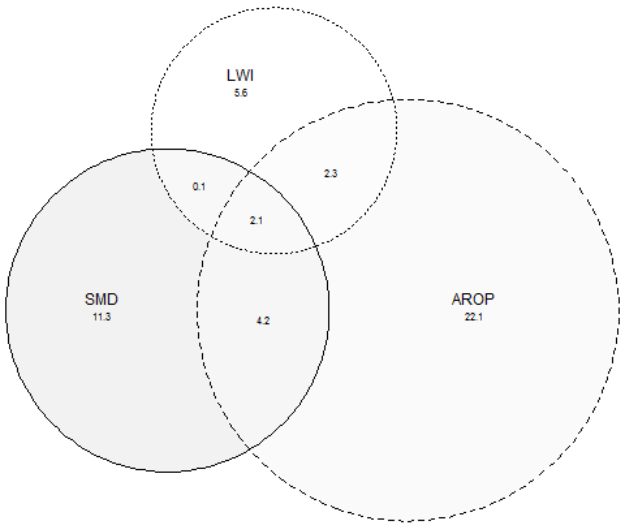
Sensitivity, specificity and intersections.



LATVIA							GDP per capita in PPS = 67			
Year	Se	Sp	$AROP \wedge \neg SMD \wedge \neg LWI$	$AROP \wedge \neg SMD \wedge LWI$	$AROP \wedge SMD \wedge \neg LWI$	$AROP \wedge SMD \wedge LWI$	$\neg AROP \wedge \neg SMD \wedge LWI$	$\neg AROP \wedge SMD \wedge \neg LWI$	$\neg AROP \wedge SMD \wedge LWI$	$\neg AROP \wedge \neg SMD \wedge \neg LWI$
2008	59.9	82.2	13.1	1.3	9.3	2.2	0.6	7.6	0.1	65.8
2009	50.9	80.6	13.3	1.8	8.3	2.9	0.7	10.5	0.3	62.1
2010	43.7	87.7	6.9	2.0	7.2	4.9	1.7	14.5	1.1	61.8
2011	37.6	89.3	5.6	1.8	7.1	4.6	1.8	18.0	1.4	59.9
2012	40.1	87.9	7.0	2.0	6.1	4.2	1.6	14.4	1.0	63.8
2013	40.0	87.1	8.0	1.8	6.3	3.3	1.4	13.4	1.0	64.9
2014	46.6	84.9	9.8	2.4	6.1	2.9	1.3	9.6	0.7	67.3
2015	54.9	83.8	11.8	1.7	6.2	2.8	0.9	7.0	0.4	69.1
2016	56.3	83.2	12.6	2.1	5.2	2.0	1.1	5.4	0.2	71.5
2017	55.8	82.2	13.5	2.3	4.2	2.1	1.1	4.9	0.1	71.8

Source: Based on Eurostat database extracted on 25.09.18

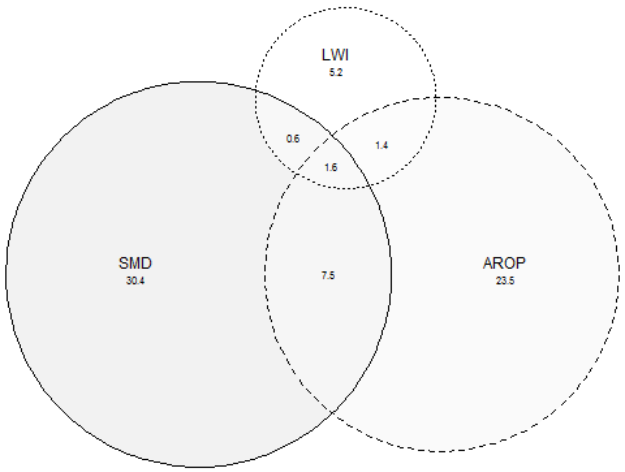
Sensitivity, specificity and intersections.



ROMANIA							GDP per capita in PPS = 63			
Year	Se	Sp	$AROP \wedge \neg SMD \wedge \neg LWI$	$AROP \wedge \neg SMD \wedge LWI$	$AROP \wedge SMD \wedge \neg LWI$	$AROP \wedge SMD \wedge LWI$	$\neg AROP \wedge \neg SMD \wedge LWI$	$\neg AROP \wedge SMD \wedge \neg LWI$	$\neg AROP \wedge SMD \wedge LWI$	$\neg AROP \wedge \neg SMD \wedge \neg LWI$
2008	44.0	86.3	8.3	0.9	12.0	2.4	2.3	17.2	1.1	55.8
2009	41.9	87.3	7.7	0.9	11.6	1.9	2.3	17.5	1.2	57.0
2010	42.1	87.4	7.9	0.9	11.3	1.5	2.3	16.3	1.3	58.5
2011	43.7	86.8	8.2	1.1	11.4	1.5	2.0	15.6	1.0	59.1
2012	41.6	85.5	9.1	0.9	10.9	2.0	2.1	17.1	1.0	56.8
2013	43.3	85.6	9.5	0.6	10.7	2.2	2.0	15.9	1.0	58.1
2014	47.5	82.7	12.0	0.8	10.1	2.2	1.6	12.8	0.8	59.7
2015	55.3	83.3	11.9	1.0	10.0	2.6	1.8	9.6	0.6	62.6
2016	49.8	82.3	12.2	1.3	9.0	2.8	1.6	11.4	0.5	61.2
2017	46.2	82.1	13.0	1.4	7.5	1.6	1.6	10.0	0.6	64.3

Source: Based on Eurostat database extracted on 25.09.18

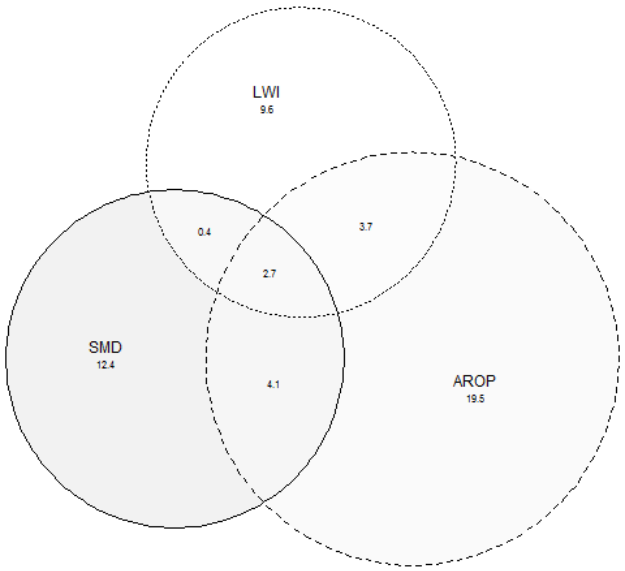
Sensitivity, specificity and intersections.



CROATIA							GDP per capita in PPS = 61			
Year	Se	Sp	$AROP \wedge \neg SMD \wedge \neg LWI$	$AROP \wedge \neg SMD \wedge LWI$	$AROP \wedge SMD \wedge \neg LWI$	$AROP \wedge SMD \wedge LWI$	$\neg AROP \wedge \neg SMD \wedge LWI$	$\neg AROP \wedge SMD \wedge \neg LWI$	$\neg AROP \wedge SMD \wedge LWI$	$\neg AROP \wedge \neg SMD \wedge \neg LWI$
2008	:	:	:	:	:	:	:	:	:	:
2009	:	:	:	:	:	:	:	:	:	:
2010	46.5	83.7	9.9	4.1	3.4	3.3	2.9	7.3	0.4	68.9
2011	47.4	83.8	9.6	4.1	3.7	3.5	3.7	7.1	0.9	67.4
2012	48.4	85.0	9.0	3.6	3.5	4.2	4.1	7.4	0.8	67.4
2013	52.7	86.2	8.1	3.7	4.2	3.6	3.4	6.4	0.6	70.1
2014	51.8	85.8	8.5	3.7	3.8	3.4	3.2	5.8	0.9	70.7
2015	54.7	85.6	8.6	3.8	4.1	3.4	2.9	5.6	0.6	70.9
2016	54.8	85.5	9.0	3.7	4.1	2.7	2.8	5.2	0.4	72.1
2017	:	:	:	:	:	:	:	:	:	:

Source: Based on Eurostat database extracted on 25.09.18

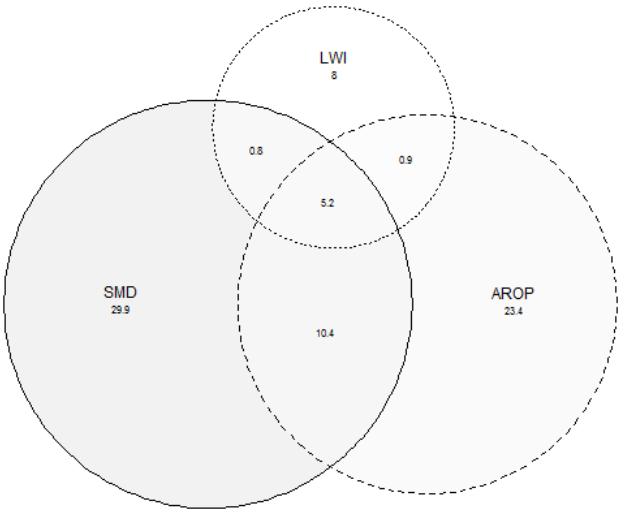
Sensitivity, specificity and intersections.



BULGARIA							GDP per capita in PPS = 49			
Year	Se	Sp	$AROP \wedge \neg SMD \wedge \neg LWI$	$AROP \wedge \neg SMD \wedge LWI$	$AROP \wedge SMD \wedge \neg LWI$	$AROP \wedge SMD \wedge LWI$	$\neg AROP \wedge \neg SMD \wedge LWI$	$\neg AROP \wedge SMD \wedge \neg LWI$	$\neg AROP \wedge SMD \wedge LWI$	$\neg AROP \wedge \neg SMD \wedge \neg LWI$
2008	44.4	94.7	2.8	0.3	13.8	4.5	0.5	22.0	0.9	55.2
2009	43.1	93.6	3.3	0.4	14.3	3.7	0.6	23.3	0.5	53.8
2010	38.9	94.7	2.6	0.3	13.5	4.3	0.6	27.1	0.8	50.8
2011	40.1	91.7	4.2	0.5	11.7	5.8	0.8	24.9	1.2	50.9
2012	38.3	92.1	3.6	0.8	11.1	5.8	0.8	25.4	1.8	50.7
2013	39.5	93.0	3.4	0.6	10.7	6.3	1.0	24.3	1.7	52.0
2014	48.2	91.3	5.0	0.8	10.7	5.2	1.3	15.5	1.6	59.9
2015	47.4	91.3	5.3	0.4	10.3	5.9	1.4	17.3	0.7	58.7
2016	49.5	89.6	6.4	0.7	10.3	5.5	1.4	15.0	1.1	59.6
2017	52.2	88.9	6.9	0.9	10.4	5.2	1.1	13.5	0.8	61.1

Source: Based on Eurostat database extracted on 25.09.18

Sensitivity, specificity and intersections.

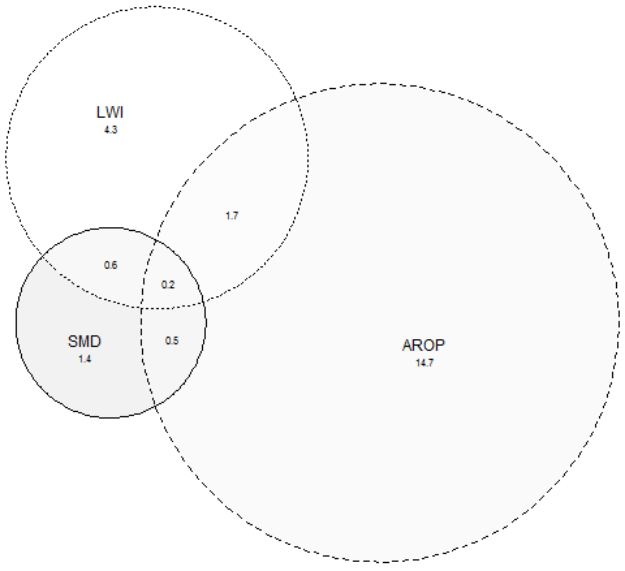


SWITZERLAND — GDP per capita in PPS = 158

Year	Se	Sp	$AROP \wedge \neg SMD \wedge \neg LWI$	$AROP \wedge \neg SMD \wedge LWI$	$AROP \wedge SMD \wedge \neg LWI$	$AROP \wedge SMD \wedge LWI$	$\neg AROP \wedge \neg SMD \wedge LWI$	$\neg AROP \wedge SMD \wedge \neg LWI$	$\neg AROP \wedge SMD \wedge LWI$	$\neg AROP \wedge \neg SMD \wedge \neg LWI$
2008	42.9	84.9	13.9	0.9	0.6	0.3	1.2	1.1	0.1	81.9
2009	50.0	85.2	13.2	1.3	0.7	0.3	1.3	0.8	0.2	82.1
2010	52.9	85.7	12.5	1.6	0.8	0.1	1.4	0.7	0.1	82.8
2011	45.5	85.3	12.9	1.7	0.2	0.3	1.6	0.5	0.1	82.8
2012	55.6	84.4	14.2	1.3	0.4	0.1	1.2	0.3	0.1	82.5
2013	42.9	85.8	12.6	1.5	0.2	0.1	1.5	0.3	0.1	83.7
2014	53.8	86.7	11.9	1.2	0.3	0.4	2.0	0.5	0.1	83.6
2015	57.1	85.0	13.3	1.5	0.5	0.3	2.0	0.3	0.3	81.8
2016	35.7	85.6	12.7	1.5	0.3	0.2	2.2	0.5	0.4	82.2
2017	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.

Source: Based on Eurostat database extracted on 25.09.18

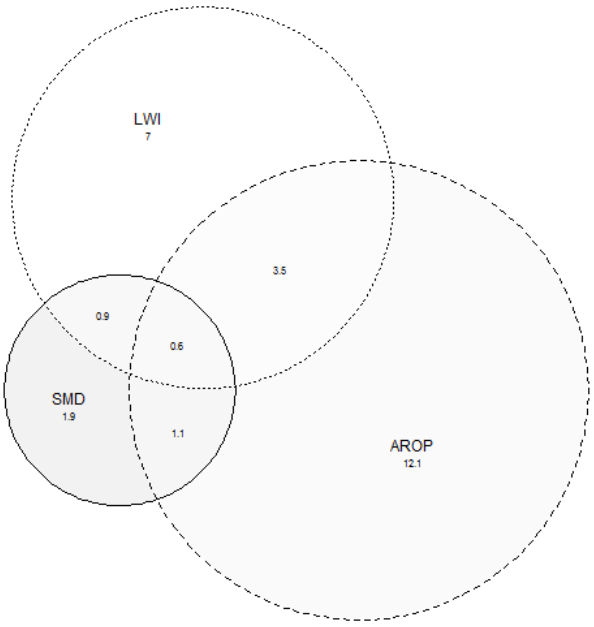
Sensitivity, specificity and intersections.



NORWAY							GDP per capita in PPS = 150			
Year	Se	Sp	$AROP \wedge \neg SMD \wedge \neg LWI$	$AROP \wedge \neg SMD \wedge LWI$	$AROP \wedge SMD \wedge \neg LWI$	$AROP \wedge SMD \wedge LWI$	$\neg AROP \wedge \neg SMD \wedge LWI$	$\neg AROP \wedge SMD \wedge \neg LWI$	$\neg AROP \wedge SMD \wedge LWI$	$\neg AROP \wedge \neg SMD \wedge \neg LWI$
2008	45.0	89.3	8.9	1.6	0.4	0.5	2.5	0.7	0.4	85.0
2009	50.0	89.2	8.7	1.9	0.4	0.7	2.4	0.7	0.4	84.8
2010	52.4	89.7	8.1	2.0	0.5	0.6	2.8	0.6	0.4	85.1
2011	47.8	90.5	7.9	1.4	0.3	0.8	2.8	0.8	0.4	85.5
2012	41.2	90.6	7.4	1.8	0.3	0.4	2.7	0.5	0.5	86.3
2013	55.0	89.9	8.1	1.8	0.6	0.5	2.3	0.6	0.3	85.9
2014	61.5	89.7	8.4	1.8	0.4	0.4	2.1	0.3	0.2	86.5
2015	47.1	88.7	8.3	2.8	0.3	0.5	2.2	0.6	0.3	85.0
2016	68.4	88.9	8.7	2.2	0.5	0.8	2.5	0.4	0.2	84.7
2017	57.9	88.8	8.1	2.9	0.5	0.6	3.2	0.5	0.3	83.9

Source: Based on Eurostat database extracted on 25.09.18

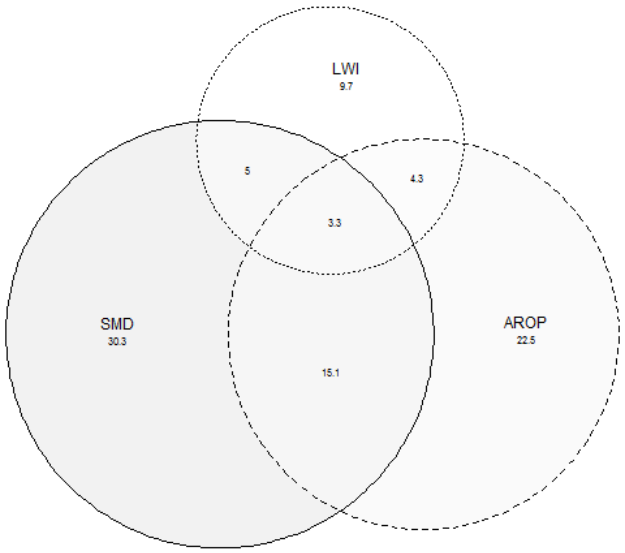
Sensitivity, specificity and intersections.



TURKEY							GDP per capita in PPS = 65			
Year	Se	Sp	$AROP \wedge \neg SMD \wedge \neg LWI$	$AROP \wedge \neg SMD \wedge LWI$	$AROP \wedge SMD \wedge \neg LWI$	$AROP \wedge SMD \wedge LWI$	$\neg AROP \wedge \neg SMD \wedge LWI$	$\neg AROP \wedge SMD \wedge \neg LWI$	$\neg AROP \wedge SMD \wedge LWI$	$\neg AROP \wedge \neg SMD \wedge \neg LWI$
2008	37.2	92.2	2.8	0.5	17.5	3.9	3.0	32.9	3.3	36.1
2009	37.6	91.0	3.3	0.6	17.6	3.7	3.2	32.3	3.1	36.1
2010	35.5	91.9	3.0	0.3	17.1	4.0	3.1	34.6	3.7	34.3
2011	35.6	93.3	2.4	0.4	16.5	4.1	3.0	34.2	3.1	36.2
2012	36.7	92.0	3.1	0.5	16.5	3.7	2.8	32.0	2.8	38.7
2013	42.9	92.3	3.9	0.4	15.4	3.4	3.1	22.7	2.3	48.8
2014	50.3	88.3	7.2	1.1	12.1	2.7	4.0	13.1	1.5	58.4
2015	49.8	89.4	6.4	1.0	11.8	3.3	3.7	13.5	1.7	58.7
2016	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
2017	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.

Source: Based on Eurostat database extracted on 25.09.18

Sensitivity, specificity and intersections.

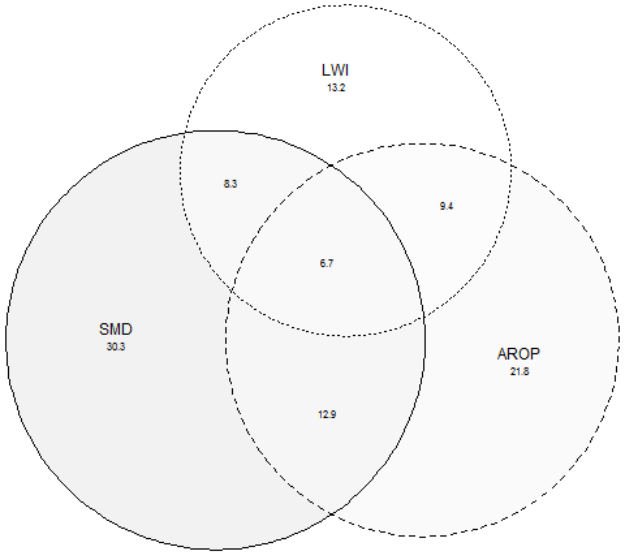


FORMER YUGOSLAV REP. OF MACEDONIA GDP per capita in PPS = 37

Year	Se	Sp	$AROP \wedge \neg SMD \wedge \neg LWI$	$AROP \wedge \neg SMD \wedge LWI$	$AROP \wedge SMD \wedge \neg LWI$	$AROP \wedge SMD \wedge LWI$	$\neg AROP \wedge \neg SMD \wedge LWI$	$\neg AROP \wedge SMD \wedge \neg LWI$	$\neg AROP \wedge SMD \wedge LWI$	$\neg AROP \wedge \neg SMD \wedge \neg LWI$
2008	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
2009	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
2010	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
2011	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
2012	45.3	87.1	5.1	2.5	8.5	10.0	1.8	20.3	2.0	49.7
2013	42.3	86.5	5.7	2.7	7.6	8.3	2.1	20.4	1.3	51.9
2014	47.1	91.8	3.6	1.7	8.6	8.2	2.3	17.0	1.9	56.7
2015	44.4	88.6	5.1	2.8	7.0	6.5	3.1	15.3	1.6	58.4
2016	43.0	87.3	6.2	2.7	6.2	6.7	2.2	15.5	1.6	58.9
2017	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.

Source: Based on Eurostat database extracted on 25.09.18

Sensitivity, specificity and intersections.

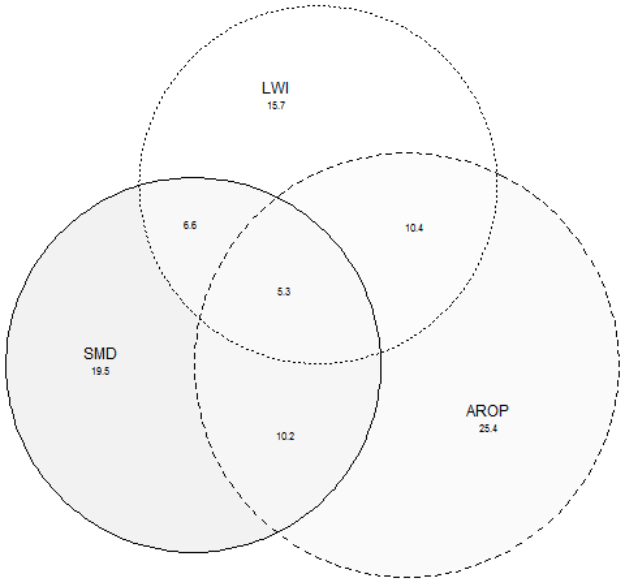


SERBIA GDP per capita in PPS = 37

Year	Se	Sp	$AROP \wedge \neg SMD \wedge \neg LWI$	$AROP \wedge \neg SMD \wedge LWI$	$AROP \wedge SMD \wedge \neg LWI$	$AROP \wedge SMD \wedge LWI$	$\neg AROP \wedge \neg SMD \wedge LWI$	$\neg AROP \wedge SMD \wedge \neg LWI$	$\neg AROP \wedge SMD \wedge LWI$	$\neg AROP \wedge \neg SMD \wedge \neg LWI$
2008	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
2009	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
2010	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
2011	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
2012	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
2013	46.8	83.7	9.1	2.8	6.7	5.9	3.3	12.6	1,7	58.0
2014	46.2	82.1	9.8	3.4	6.2	6.0	3.6	11.9	2,3	56.9
2015	49.4	82.1	9.3	4.3	6.1	5.8	3.8	10.4	1,8	58.7
2016	52.3	81.1	10.1	5.1	4.9	5.3	4.0	8.0	1,3	61.3
2017	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.

Source: Based on Eurostat database extracted on 25.09.18

Sensitivity, specificity and intersections.



Chapter 7

REFERENCES AND BIBLIOGRAPHY.

Agarwal, V., Taffler, R., 2008. Comparing the performance of market-based and accounting-based bankruptcy prediction models. J. Bank. Finance 32, 1541–1551. https://doi.org/10.1016/j.jbankfin.2007.07.014

Alkire, S., Santos, M.E., 2013. A Multidimensional Approach: Poverty Measurement & Beyond. Soc. Indic. Res. 112, 239–257. https://doi:10.1007/s11205-013-0257-3

Alkire, Sabina, Foster, James, Seth, Suman, Santos, Maria Emma, Roche, JosÃ© Manuel, Ballon Paola, 2015. Multidimensional Poverty Measurement and Analysis. Oxford.

Atkinson, A.B., Guio, Anne-Catherine, Marlière, Èric, 2017. Monitoring social inclusion in Europe. Eurostat. https://doi.org/10.2785/60152

Atkinson, A.B., Marlier, E., 2010. Income and living conditions in Europe. Publications Office of the European Union, Luxembourg.

Atkinson, A.B., Marlier, E., Europäische Kommission, Statistisches Amt, 2010. Income and living conditions in Europe. Publ. Office of the European Union, Luxembourg.

Battiston, D., Cruces, G., López-Calva, L.F.L., Ana Lugo, M.A., Santos, M.E., 2013. Income and beyond: multidimensional poverty in six Latin American countries. Soc. Indic. Res. 112. https://doi.org/10.1007/s11205-013-0249-3

Berardi, N., Sevestre, P., Thébault, J., 2017. The Determinants of Consumer Price Dispersion: Evidence from French Supermarkets. In: Matyas L. (eds) The Econometrics of Multi-dimensional Panels. Advanced Studies in Theoretical and Applied Econometrics, vol 50. Springer, Cham. https://doi.org/10.1007/978-3-319-60783-2_15

Betson, D.M., Muellbauer, J., 2004. Poverty equivalence scales: adjustment for demographic differences across families, in: National Research Council Workshop on Experimental Poverty Measures, June. Citeseer, pp. 15–16.

Bossert, W.a, Chakravarty, S.R., D'Ambrosio, C., 2012. Poverty and time. J. Econ.

Inequal. 10, 145–162.
https://doi:10.1007/s10888-011-9175-2

Buhmann, B., Rainwater, L., Schmaus, G., Smeeding, T., 1988. Equivalence Scales, Well-Being, Inequality, and Poverty: Sensitivity Estimates across Ten Countries Using the Luxembourg Income Study (LIS) Database. Rev. Income Wealth 34, 115–142.

Chzhen, Y., de Neubourg, C., Plavgo, I., de Milliano, M., 2016. Child Poverty in the European Union: the Multiple Overlapping Deprivation Analysis Approach (EU-MODA). Child Indic. Res. 9, 335–356.
https://doi.org/10.1007/s12187-015-9321-7

Costa, A., García, J, López, X., Raymond, J, 2015. Estimación de las paridades de poder adquisitivo para las comunidades autónomas españolas. Presented at the International Conference of Regional Science, Facultad de Economía y Empresa de la Universitar Rovira e Virgill, Reus (Spain).

Dür, A., Baccini, L., Elsig, M., 2014. The design of international trade agreements: Introducing a new dataset. Rev. Int. Organ. 9, 353–375.
https://doi.org/10.1007/s11558-013-9179-8

ECB, 2016. The Household Finance and Consumption Survey: results from the second wave.
https://doi.org/10.2866/177251

European Commission, 2010. EUROPE 2020: A strategy for smart, sustainable and inclusive growth.

European Commission, 2017. Regional indicators of socio-economic well-being.

Eurostat (Ed.), 2015. Regions in the European Union: nomenclature of territorial units for statistics; NUTS 2013/EU–28, Eurostat manuals and guidelines. Publ. of the Europ. Union, Luxembourg.

Eurostat, 2013. European system of accounts. ESA 2010.
https://doi.org/10.2785/16644

Eurostat, 2017a. Towards a harmonised methodology for statistical indicators –Part 2: Communicating through indicators.

https://doi.org/10.2785/799718

Eurostat, 2017b. Sustainable development in the European Union. Monitoring report on progress towards the SDGs in an EU context.
https://doi.org/10.2785/237722

Eurostat, 2018. Handbook on the compilation of statistics on illegal economic activities in national accounts and balance of payments.
https://doi.org/10.2785/0256

Eurostat, OECD, 2012. Eurostat-OECD methodological manual on purchasing power parities. Publications Office, Luxembourg.

Fawcett, T., 2006. An introduction to ROC analysis. Pattern Recognit. Lett. 27, 861–874.
https://doi.org/10.1016/j.patrec.2005.10.010

Fusco, A., Guio, A., Marlier, E., 2010. Income poverty and material deprivation in European countries. Eurostat.
https://doi.org/10.2785/1144

Gillie, A., 2008. Identifying the poor in the 1870s and 1880s1. Econ. Hist. Rev. 61, 302–325.

Guio, A., Gordon, D., Marlier, E., 2012. Measuring material deprivation in the EU: indicators for the whole population and child-specific indicators. Publications Office, Luxembourg.

Guio, A.C., Marlier, E., Gordon, D., Fahmy, E., Nandy, S., Pomati, M., 2016. Improving the measurement of material deprivation at the European Union level. J. Eur. Soc. Policy.
https://doi.org/10.1177/0958928716642947

Hand, D.J., Till, R.J., 2001. A simple generalisation of the area under the ROC curve for multiple class classification problems. Mach. Learn.
https://doi.org/10.1023/A:1010920819831

Harness, Nathaniel J., Chatterjee, Swarn, Finke, Michael, 2008. Household Financial Ratios: A Review of Literature. J. Pers. Finance 6, 77 –97.

Heikkilä, M., Ahlström, C.S., Gissler, M., Kiikkala, I., Kortteinen, M., Sintonen, H., Taskinen, S., Virtanen, M., 2004. Poverty dynamics according to direct, indirect and subjective measures.

INE, 1991. Censo de población de la Corona de Castilla "Marqués de la Ensenada". 1752.

Izquierdo Llanes, G., Blasco Torrejón, B., Recio Rapún, M.L., 2009. Economía de la vivienda en España. Instituto de Estudios Económicos.

Jolliffe, D., Prydz, E.B., 2015. Global poverty goals and prices: how purchasing power parity matters. World Bank Policy Res. Work. Pap.

Krumer-Nevo, M., Gorodzeisky, A., Saar-Heiman, Y., 2017. Debt, poverty, and financial exclusion. J. Soc. Work 17, 511 –530.
https://doi.org/10.1177/1468017316649330

López-Ratón, M., Rodríguez-Álvarez, M.X., Cadarso-Suárez, C., Gude-Sampedro, F., 2014. OptimalCutpoints: an R package for selecting optimal cutpoints in diagnostic tests. J. Stat. Softw. 61, 1–36.
https://doi.org/10.18637/jss.v061.i08

Lang, V.F., Lingnau, H., 2015. Defining and Measuring Poverty and Inequality Post-2015: Poverty and Inequality Post-2015. J. Int. Dev. 27, 399–414.
https://doi.org/10.1002/jid.3084

Lee, M.P., Sabri, M.F., 2017. Review of Financial Vulnerability Studies. Arch. Bus. Res. 5.
https://doi.org/10.14738/abr.52.2784

Lelkes, O., Gasior, K., 2018. Income Poverty in the EU: What Do We Actually Measure? Empirical Evidence on Choices, Underlying Assumptions and Implications (Based on EU-SILC 2005 –2014).
https://doi.org/10.1007/978-3-319-65006-7_6

López-Ratón, M., Rodríguez-Álvarez, M.X., Cadarso-Suárez, C., Gude-Sampedro, F., 2014. OptimalCutpoints: an R package for selecting optimal cutpoints in diagnostic tests. J. Stat. Softw. 61, 1 –36.
https://doi.org/10.18637/jss.v061.i08

Meyer, Bruce D., Sullivan, James X., 2017. Consumption and Income Inequality in the U.S. Since the 1960s. NBER.
https://doi.org/10.3386/w23655

Moatsos, M., 2016. Global Absolute Poverty: Behind the Veil of Dollars. J. Glob. Dev. 7.
https://doi.org/10.1515/jgd-2016-0033

Moisio, P., 2004. Poverty dynamics according to direct, indirect and subjective measures: Modelling Markovian processes in a discrete time and space with error. Stakes.

Mood, C., Jonsson, J.O., 2016. The Social Consequences of Poverty: An Empirical Test on Longitudinal Data. Soc. Indic. Res. 127, 633–652.
https://doi:10.1007/s11205-015-0983-9

Notten, G., 2016. How Poverty Indicators Confound Poverty Reduction Evaluations: The Targeting Performance of Income Transfers in Europe. Soc Indic Res.
https://doi.org/10.1007/s11205-015-0996-4

OECD, 2013. OECD Framework for Statistics on the Distribution of Household Income, Consumption and Wealth.
https://doi.org/10.1787/9789264194830-en

OECD, 2016. Society at a Glance 2016: OECD Social Indicators.
https://doi.org/10.1787/soc_glance-2016-17-en

Osier, G., Berger, Y., Goedemé, T., European Commission, Eurostat, 2013. Standard error estimation for the EU-SILC indicators of poverty and social exclusion: 2013 edition. Publications Office, Luxembourg.

Ravallion, M., 2011. Weakly relative poverty. Rev. Econ. Stat. 93, 1251–1261.
https://doi:10.1162/REST_a_00127

Ravallion, M., 2015. On testing the scale sensitivity of poverty measures. Econ. Lett. 137, 88–90.
https://doi:10.1016/j.econlet.2015.10.034

Ravallion, M., Himelein, K, Beegle, K, 2016. Can Subjective Questions on Economic Welfare Be Trusted? Econ. Dev. Cult. Change 64, 697–726.

Ravallion, M., Lokshin, M., 2006. Testing poverty lines. Rev. Income Wealth. https://doi.org/10.1111/j.1475-4991.2006.00196.x

Roberts, G.P., Barnes, H.A., Mackie, C., 2001. Using the microsoft excel "solveÂ´Â´ tool to perform non-linear curve fitting, using a range of non-Newtonian flow curves as examples. Appl. Rheol. 11, 271–276.

Salcedo, A.M., Izquierdo Llanes, G., 2017. An Empirical Approach to the Poverty Indicators Based on Revealed Parameters: The Case of Spain. Soc. Indic. Res. https://doi.org/10.1007/s11205-017-1675-4

Salcedo, A.M., Izquierdo Llanes, G., 2018. Drawing the optimal monetary poverty lines based on empirical data: an application to Spain. Qual Quant. https://doi.org/10.1007/s11135-018-0766-x

Salcedo, A.M., Gómez, Y., Izquierdo Llanes, G., 2017. National statistical system coordination: a tool or a fundamental principle?. Boletín de Estadística e Investigación Operativa. Vol. 33, No. 1, Marzo 2017, pp. 43-56.

Santos, M.E., Villatoro, P., 2018. A Multidimensional Poverty Index for Latin America: A Multidimensional Poverty Index for Latin America. Rev. Income Wealth 64, 52 –82. https://doi.org/10.1111/roiw.12275

Shiu, S.-Y., Gatsonis, C., 2008. The predictive receiver operating characteristic curve for the joint assessment of the positive and negative predictive values. Philos. Trans. R. Soc. Math. Phys. Eng. Sci. 366, 2313–2333. https://doi.org/10.1098/rsta.2008.0043

Šileika, A., Bekerytė, J., 2013. The theoretical issues of unemployment, poverty and crime coherence in the terms of sustainable development. J. Secur. Sustain. Issues 2, 59–70. https://doi.org/10.9770/jssi.2013.2.3(5)

Thorat, A., Vanneman, R., Desai, S., Dubey, A., 2017. Escaping and Falling into Poverty in India Today. World Dev. 93, 413–426. https://doi.org/10.1016/j.worlddev.2017.01.004

Törmälehto, V.-M., Sauli, H., 2013. The distributional impact of imputed rent in

EU-SILC 2007-2010. Eurostat.
https://doi.org/10.2785/21725

Ukpere, Wilfred I., Slabbert, Andre D., 2009. A relationship between current globalisation, unemployment, inequality and poverty. Int. J. Soc. Econ.
https://doi.org/10.1108/03068290910921172

UNECE, 2017. Guide on Poverty Measurement.
https://www.unece.org/index.php?id=47512&L=0

United Nations, 2015. Transforming our world: the 2030 Agenda for Sustainable Development.

United Nations, 2017. The Sustainable Development Goals Report 2017, UN, New York.
https://doi.org/10.18356/4d038e1e-en

Verma, V., Betti, G., Gagliardi, F., European Commission, Eurostat, 2010. Robustness of some EU-SILC based indicators at regional level. Publications Office, Luxembourg.

Walsh, S., Diamond, D., 1995. Non-linear curve fitting using Microsoft Excel Solver. Talanta 42, 561–572.
https://doi:10.1016/0039-9140(95)01446-I

Weziak-Bialowolska, D., Dijkstra, L., European Commission, Joint Research Centre, Institute for the Protection and the Security of the Citizen, 2014. Monitoring multidimensional poverty in the regions of the European Union. Publications Office, Luxembourg.

World Bank, 2015. Global Monitoring Report 2015/2016: Development Goals in an Era of Demographic Change. The World Bank.

Zimmer, J., 2015. The German Fiscal Equalisation System: Taking Account of Differences in Governmental Purchasing Powers. Available SSRN 2731109

Chapter 8

ABSTRACT / RESUMEN.

8.1 Abstract.

This doctoral research is titled "Data science applied to refining socio-economic indicators based on empirical data".

The construction of socio-economic well-being indicators to represent reality in the most accurate way constitutes an important challenge in the production of statistics, as recognized by prestigious international organizations. The production of different harmonized surveys in the last years allows us to have enough empirical data that offer an interesting opportunity to try to refine, if possible, the existing classic models of indicators compilation.

Putting the attention on one of the aspect with greater degree of complexity - its current paradigm of objective measurement of poverty - the main elements and approaches are analysed and the following three specific issues are addressed:

- Equivalization: This technique allows to transform an economic variable from households to individuals giving a different weighting to the members of a household. The choice by convention of different equivalence scales has direct effects on the results obtained.
- Thresholds: The at-risk-of monetary poverty rate can be considered as a socio-economic indicator based on a concept of economic distance, which is understood as a percentage p of the median equivalent income per adult. In the European Unions it's calculated by placing the percentage p at 60% in all EU countries, but the United Nations and the OECD recommend the use of a 50% of the median of equivalent income. The fact that such

prestigious institutions recommend by convention different percentages offers an opportunity to find out improvements to the calculation method based on the data science techniques.

- Multidimensional approach: In the European Union the use of an unidimensional indicator based on disposable income has certain limitations when representing such a complex phenomenon by means of a single indicator. It has been suggested that a multidimmensional framework based on Income-Consumption-Wealth might provide a possible solution since it would integrate the three variables clearly linked to poverty.

At this point it's worth mentioning that the results of this research should be considered from a perspective of complementarity rather than substitutability. The main results include an application of a model of non-linear parametric equivalence scale that represents a generalization of the existing models proposed by the Organization for Economic Cooperation and Development (OECD) allowing additionally the inclusion of purchasing power parities, which as a whole offers a better adjustment by autonomous communities. Additionally, the application of ROC (Receiver Operating Characteristic) curves to determine optimal cut-off points allows to have results that improve time consistency in relation to the indicators of severe material deprivation. Finally, the use of a multidimensional model based on the ICW paradigm implies an extension of the current model, that also offers an improvement in the results, particularly with regard to the sensitivity and specificity of the model.

The thesis is structured around eight chapters: the introduction is presented in chapter 1; chapter 2 analyzes the situation of the equivalence scales and their influence on the coherence of the model; chapter 3 studies the determination of thresholds based on different percentages of the median of equivalent income; chapter 4 generalizes the model applying a multidimensional approach; chapter 5 presents the conclusions; chapter 6 includes an annex with sensitivity and specificity data, ordered by GDP per capita and membership of the European Union; chapter 7 shows the bibliography and references used; to conclude, chapter 8 offers the summary in English and Spanish.

Finally, it should be noted that this thesis conclusions, particularly those related to the equivalence scales and the application of ROC curves, can be easily extrapolated to other statistical topics that present similar difficulties when constructing quantitative indicators based on empirical data.

8.2 Resumen.

Esta investigación doctoral se presenta bajo el título "Data science applied to refining socioeconomic indicators based on empirical data".

La construcción de indicadores socio-económicos de bienestar que representen la realidad de la forma más precisa posible constituye un importante reto en la producción de estadísticas, conforme se reconoce por prestigiosos organismos internacionales. La realización de distintas encuestas armonizadas a los largo de los últimos años nos permite disponer de suficientes datos empíricos que ofrecen la interesante oportunidad de tratar de refinar, en caso posible, los modelos clásicos existentes de producción de indicadores.

Poniendo la atención en uno de los aspectos con mayor grado de complejidad - el actual paradigma de medición objetiva de la pobreza - se analizan sus principales elementos y enfoques y se pretende buscar una respuesta a tres situaciones concretas:

- Equivalencia o "equivalización": Es una técnica utilizada para transformar una variable económica de hogares a individuos dando un peso diferente a los miembros del hogar. La elección por convención de distintas escalas de equivalencia tiene efectos directos en los resultados obtenidos.
- Umbrales: La tasa de riesgo de pobreza monetaria puede considerarse como un indicador socio-económico basado en un concepto de distancia económica, entendido como un porcentaje p de la mediana del ingreso equivalente por adulto. En la Unión Europea se calcula situando el porcentaje p en el 60% en todos los países, pero Naciones Unidas y la OCDE recomiendan utilizar un 50%

de la mediana. El hecho de que estas prestigiosas instituciones recomienden distintos porcentajes por convención supone una oportunidad para buscar mejoras a su cálculo por medio de la aplicación de la ciencia de datos.

- Enfoque multidimensional: En la Unión Europea el uso de un indicador unidimensional basado en la renta disponible ofrece limitaciones al representar un fenómeno muy complejo mediante un único dato. Se ha planteado si un enfoque multidimensional basado en la terna Ingreso - Consumo - Riqueza podría proporcionar una posible solución dado que integraría tres variables claramente relacionadas con este fenónemo.

En este punto conviene destacar que los resultados de esta investigación deben considerarse en general desde una óptica de complementariedad en lugar de sustituibilidad. Como principales resultados se consigue aplicar un modelo de escala de equivalencia paramétrico no lineal que supone una generalización de los actuales modelos existentes propuestos por la Organización para la Cooperación y el Desarrollo Económicos (OCDE), permitiendo adicionalmente la inclusión de paridades de poder de compra que en su conjunto ofrece un mejor ajuste por comunidades autónomas. Adicionalmente la aplicación de curvas ROC (Receiver Operating Characteristic) para determinar puntos de corte óptimo permite disponer de resultados que mejoran la coherencia temporal frente a los indicadores de carencia material severa. Finalmente el empleo de un modelo multidimensional basado en el paradigma ICW supone una ampliación del modelo clásico y también ofrece una mejora los resultados, particularmente en lo que se refiere a la sensibilidad y especificidad del modelo.

La tesis se estructura entorno a ocho capítulos: la introducción se presenta en el capítulo 1; el capítulo 2 analiza la situación de las escalas de equivalencia y su influencia en la coherencia del modelo; el capítulo 3 estudia la determinación de umbrales basados en diferentes porcentajes de la mediana de la renta equivalente; el capítulo 4 generaliza el modelo aplicando un enfoque multidimensional; el capítulo 5 presenta las conclusiones; el capítulo 6 incluye un anexo con los datos de sensibilidad y especificidad ordenados según PIB per cápita y pertenencia a la Unión Europea; el capítulo 7 muestra la bibliografía y referencias utilizadas; por último, el capítulo 8 ofrece el resumen en inglés y español.

Finalmente conviene señalar que las conclusiones de esta tesis, particularmente las relacionadas con la determinación de las escalas de equivalencia y la aplicación de curvas ROC, pueden ser fácilmente extrapolables a otros ámbitos que presenten dificultades similares a la hora de construir indicadores estadísticos cuantitativos basados en datos empíricos.