Springer Proceedings in Mathematics & Statistics

Nicola Salvati
Cira Perna
Stefano Marchetti
Raymond Chambers *Editors*

Studies in Theoretical and Applied Statistics

SIS 2021, Pisa, Italy, June 21–25

SIS Società Italiana di Statistica

Springer

Springer Proceedings in Mathematics & Statistics

Volume 406

This book series features volumes composed of selected contributions from workshops and conferences in all areas of current research in mathematics and statistics, including data science, operations research and optimization. In addition to an overall evaluation of the interest, scientific quality, and timeliness of each proposal at the hands of the publisher, individual contributions are all refereed to the high quality standards of leading journals in the field. Thus, this series provides the research community with well-edited, authoritative reports on developments in the most exciting areas of mathematical and statistical research today.

Nicola Salvati · Cira Perna · Stefano Marchetti ·
Raymond Chambers
Editors

Studies in Theoretical and Applied Statistics

SIS 2021, Pisa, Italy, June 21–25

Società Italiana di Statistica

Springer

Editors
Nicola Salvati
Department of Economics and Management
University of Pisa
Pisa, Italy

Cira Perna
Department of Economics and Statistics
University of Salerno
Fisciano, Salerno, Italy

Stefano Marchetti
Department of Economics and Management
University of Pisa
Pisa, Italy

Raymond Chambers
School of Mathematics and Applied Statistics
University of Wollongong
Wollongong, NSW, Australia

ISSN 2194-1009 ISSN 2194-1017 (electronic)
Springer Proceedings in Mathematics & Statistics
ISBN 978-3-031-16608-2 ISBN 978-3-031-16609-9 (eBook)
https://doi.org/10.1007/978-3-031-16609-9

Mathematics Subject Classification: 62-06

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

This book gathers selected peer-reviewed papers presented during the 50th Scientific Meeting of the Italian Statistical Society (SIS2021). Due to the Covid-19 pandemic, which limited the mobility of the staff of many universities and research centres, SIS2021 was conducted remotely from the 21st to the 25th of June 2021.

This biennial conference is a traditional meeting for promoting interactions among national and international researchers in statistics, demography, and applied statistics in Italy. The aim of the conference is to bring together national and foreign researchers and practitioners to discuss recent developments in theoretical and applied statistics as well as in demography and statistics for the social sciences.

The Scientific Program Committee and the Organizing Committee of SIS2021 put together a balanced and stimulating program which was of great interest to all participants.

The conference program included 4 plenary sessions, 15 specialized sessions, 20 solicited sessions, 37 contributed sessions, and the poster exhibition. The meeting also hosted three Satellite Events on 'Measuring uncertainty in key official economic statistics', 'Covid-19: the urgent call for a unified statistical and demographic challenge', and 'Evento SIS-PLS Statistica in classe: verso un insegnamento laboratoriale'. There were 323 submissions accepted by the Scientific Program Committee, including 128 that were presented at invited plenary, specialized and solicited sessions, and 195 that were submitted as contributed papers for oral presentation and for the poster sessions.

This book of selected papers from those presented at SIS2021 covers a wide variety of subjects and provides an overview of the current state of Italian scientific research in theoretical and applied statistics. The papers contained in this book cover areas that include Bayesian models, survey methods, time series models, spatial models, finance models, clustering methods, and new methods and applications to Covid-19.

The Scientific Program Committee, the Organizing Committee, and many volunteers contributed to the organization of SIS2021 and to the refereeing of the papers included in this book. Our heartfelt thanks go to all of them. A special thank you goes to Francesco Schirripa Spagnolo for his continuous assistance and support in the organization of the conference and in the editing of this book.

Wishing you a productive and stimulating reading experience.

Pisa, Italy	Nicola Salvati
Salerno, Italy	Cira Perna
Pisa, Italy	Stefano Marchetti
Wollongong, Australia	Raymond Chambers

Contents

A Composite Index of Economic Well-Being for the European Union Countries

Andrea Cutillo, Matteo Mazziotta, and Adriano Pareto

Abstract The measurement of Equitable and Sustainable Well-being (BES) in Italy is one of the most appreciated monitoring tools by the Scientific Community. The focus on the Economic Well-being domain seems essential around the last serious economic crisis. The use of an innovative composite index can help to measure the multidimensional phenomenon and monitor the situation at European level.

Keywords Composite index · Ranking · Economic well-being

1 Introduction

In this paper, the economic well-being in Europe is focused, taking as a reference point the economic domain of the project BES (Equitable and Sustainable Well-being in Italy) of the Italian National Institute of Statistics (Istat). The BES aims at evaluating the progress of societies by considering different perspectives through twelve relevant theoretical domains, each one measured through a different set of individual indicators. The BES project is inspired by the Global Project on Measuring the progress of Societies of the Oecd (2007), with the idea that the economic well-being is not enough for the developed Countries. However, since 2007, two huge economic crises have affected the households' economic well-being: the international economic crisis (about 2008–2009) derived from the Lehman Brothers failure; and

A. Cutillo · M. Mazziotta (✉) · A. Pareto
Italian National Institute of Statistics, Rome, Italy
e-mail: mazziott@istat.it

A. Cutillo
e-mail: cutillo@istat.it

A. Pareto
e-mail: pareto@istat.it

N. Salvati et al. (eds.), *Studies in Theoretical and Applied Statistics*, Springer Proceedings in Mathematics & Statistics 406, https://doi.org/10.1007/978-3-031-16609-9_1

the European crisis of the sovereign debts, whose effects were more intense in 2011–2012, and can be considered solved in 2014.[1] In the meantime, the EU fiscal and monetary policies have completely changed, going from very restrictive ones in the international economic crisis and in the first part of the sovereign debt crisis, to more expansive ones (especially the monetary policy) starting from the second part of the sovereign debt crisis till nowadays. This fact has reflected in a great improve of the European household' economic conditions, as can be seen in the next paragraphs. Then, the economic domain of the well-being still deserves particular relevance within the other dimensions. Following the timeliness described above, the longitudinal analysis is set at 4 relevant years: 2007, 2010, 2014 and 2019.

2 Theoretical Framework

The starting point of our framework is the Istat BES: it measures the economic domain through a set of ten indicators. However, we operate some changes due to operational issues (data availability and comparability with the other countries for the wealth indicators and the absolute poverty indicator) as well as theoretical issues (a couple of indicators can hardly be considered as economic well-being indicators and another one has been excluded in order to avoid a double counting of inequality). Changes and restriction are extensively described in the depiction of the adopted sub-domains and indicators. In this paper, we measure the economic well-being through four sub-domains (Purchasing power, Inequality, Poverty and Subjective evaluation), each one represented by a single indicator coming from the Eu-Silc (European Statistics on Income and Living Conditions) system. Purchasing power can give an evaluation of the average economic standard of a Country; Inequality is an important issue even in case of the rich Countries, since it measures the share of people who are relatively disadvantaged in respect of their social and economic context; poverty measures the share of people who can't reach a minimum standard of living; and the subjective evaluation is important in order to capture people who feel to have economic problems, even when they do not have difficulties under an objective point of view.

1. Sub-domain *Purchasing Power*; indicator: *Median equivalised income in purchasing power standards (Pps).* The Istat *average income per capita* is replaced for three reasons. First, the median is a better indicator of a monetary distribution, given its robustness to extreme values. Second, the equivalised form (through the modified Oecd scale) is better in order to consider the different sizes and needs of the households. In the opinion of the authors, also the Istat BES could benefit in changing accordingly. Finally, in the European context, it is essential to consider the different cost of life and purchasing powers in the

[1] Obviously, we cannot forget the current crisis deriving from the Covid19 pandemic situation. However, the adopted indicators are not still available for 2020 in all the Countries. Moreover, it should be evaluated when the pandemic situation will be officially declared as finished.

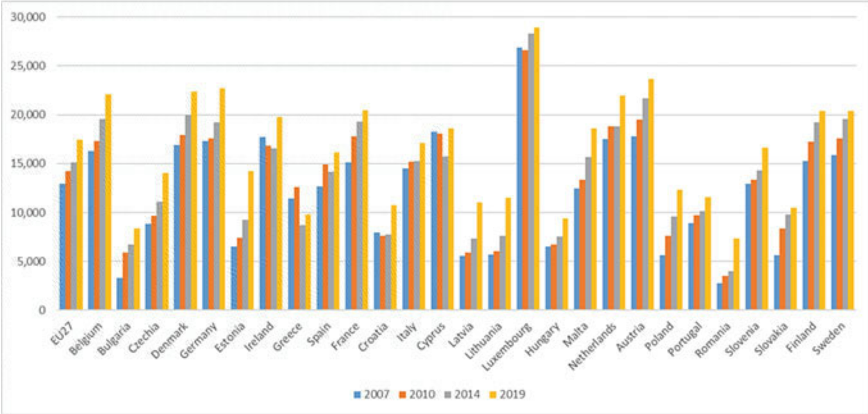

Fig. 1 Median equivalised income in Pps in the EU27 Countries. Years 2007, 2010, 2014 and 2019. Values in euros

Countries. The values of this indicator (Fig. 1) positively defined in respect of well-being, range between 2,783 euros (Romania in 2007) to 28,943 (Luxemburg in 2019). Romania presents the lowest values in all the years, even if this Country multiplies its value by 2.6 in the entire period (7,338 euros in 2019), while Luxemburg presents the highest values in all the years. The values in the entire EU27 are 12,927 euros in 2007, 14,235 in 2010, 15,137 in 2014 and 17,422 euros in 2019.

2. Sub-domain *Inequality*; indicator: *At risk of poverty rate (ARP)*. It is a relative measure of poverty: its threshold is set dependently on the income distribution and, therefore, it merely captures how many individuals are far from the others. That is, relative poverty is an inequality indicator rather than a poverty indicator [7]. Istat measures inequality also through the *Disposable income inequality* (S80_S20 index, which is the ratio of total equivalised income received by the 20% of the population with the highest income to that received by the 20% of the population with the lowest income). Since they are both representative of the same sub-domain and in order to avoid a double counting of the same domain, only the ARP is selected (as a matter of fact, the correlation coefficient between the two indicators is more than 0.90). As a general rule, it is a good practice to strictly select indicators in the construction of composite indicators. The ARP generally shows the lowest degree of variability across the Countries as well as across the years (Fig. 2). The values of ARP, negatively defined, range between 9.0% (Czechia in 2010) to 25.1% (Romania in 2014). Czechia presents the lowest values in all the years, while Romania presents the highest values in all the years. The values in the entire EU27 are 16.3% in 2007, 16.5% in 2010, 17.3% in 2014 and 16.5% euros in 2019.
3. Sub-domain *Poverty*; indicator *Severe material deprivation (SMD)*, that is the share of population living in households lacking at least 4 items out of 9 economic

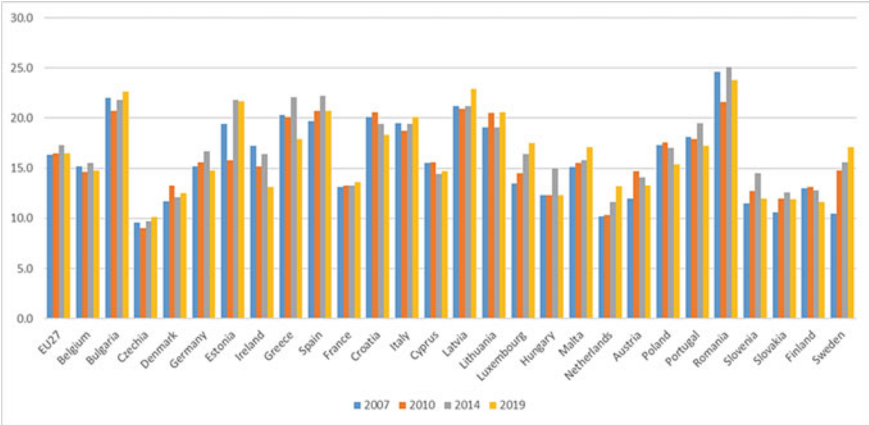

Fig. 2 At risk of poverty rate in the EU27 Countries. Years 2007, 2010, 2014 and 2019. Values in percentages

deprivations. Far from being a perfect indicator, it is the most similar indicator to the concept of absolute poverty in the EU. Unwillingly, Istat *Absolute poverty rate* cannot be used even if it is a better measure of poverty (the poverty lines are set independently of the monetary distribution, and also consider the different cost of life in different areas). However, the absolute poverty is officially measured only in Italy and USA, because of the difficulties in its definition, and the European Commission project "Measuring and monitoring absolute poverty—ABSPO" is still in the phase of study [4]. Leaving aside the World Bank measure, which does not fit for developed Countries, the SMD is the only indicator that permits European comparison in this domain, entailing data comparability. The values of this indicator (Fig. 3), negatively defined, range between 0.5% (Luxemburg in 2010) to 57.6% (Bulgaria in 2007). Bulgaria presents the highest values in all the years, but also shows a dramatic fall in the course of the years (19.9% in 2019), partly filling the gap with the other Countries. The values in the entire EU27 are 9.8% in 2007, 8.9% in 2010, 9.1% in 2014 and 5.4% in 2019.

4. Sub-domain *Subjective evaluation*; the indicator *Index of self-reported economic distress*, that is the share of individuals who declare to get to the end of the month with great difficulty. The subjective sub-dimension is considered an important one, since it can capture a worsening in well-being for people who feel to have economic problems, even when they have not difficulties with an objective point of view. This is particularly relevant especially in years in which the economic crises could have highly changed the perceptions of the households in a different way between Countries. The values of this indicator (Fig. 4), negatively defined, range between 1.4% (Germany in 2019) to 39.5% (Greece in 2014). Indeed, Germany is the Country that appeared as the leading one in EU in these years, and this fact reflects in the perceptions of the households. At the opposite, the dramatic jump in the 2014 Greek indicator indicates the uncertainty deriving

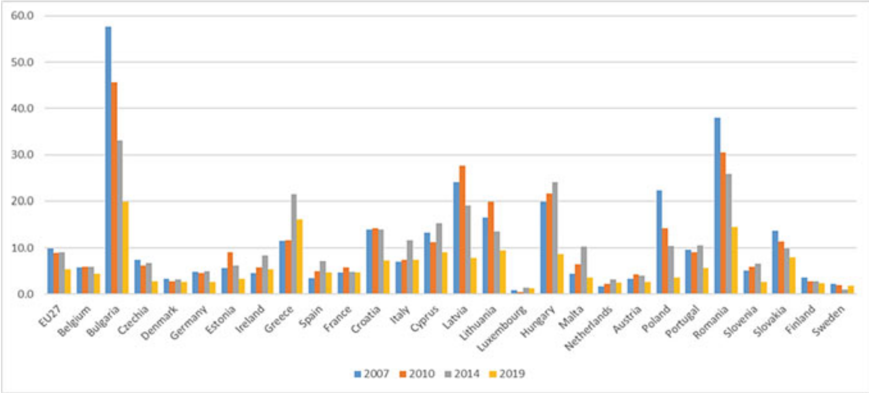

Fig. 3 Severe material deprivation in the EU27 Countries. Years 2007, 2010, 2014 and 2019. Values in percentage

from the consequences of the crisis for the Greek households. The values in the entire EU27 are 9.8% in 2007, 11.2% in 2010, 11.8% in 2014 and 6.5% in 2019.

The remaining four Istat indicators are removed for the following reasons: *Per capita net wealth*: the sub-domain wealth is certainly a pillar of the households' monetary well-being. However, correctly measuring the value of wealth is extremely complex [1], since some types of wealth are statistically hidden (e.g., paintings, jewellery etc.), and attributing a value to wealth is arbitrary when some types of wealth, e.g. houses, are not sold/bought. Unfortunately, this exclusion is a relevant issue in the European context, considering the different weight between financial

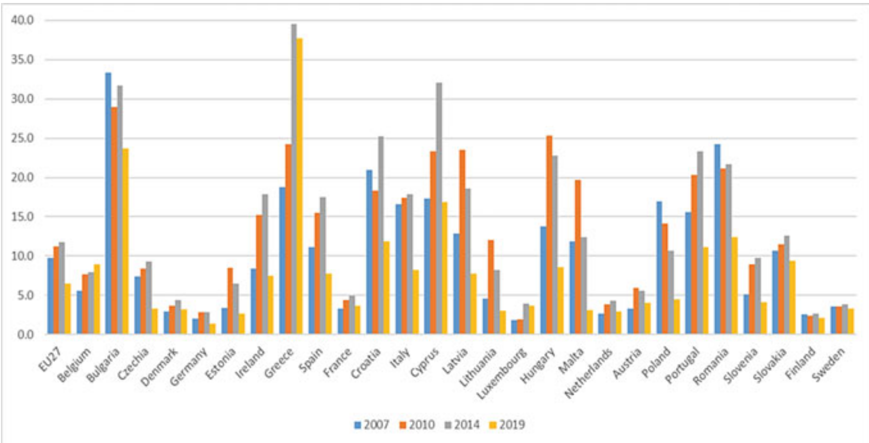

Fig. 4 Households declaring to get to the end of the month with great difficulty in the EU27 Countries. Years 2007, 2010, 2014 and 2019. Values in percentage

wealth and real estate wealth in the different countries. *People living in financially vulnerable households*, measured through the percentage of households with debt service greater than 30% of disposable income: to the best of our knowledge, there is not such indicator in the Eu-Silc database. *Severe housing deprivation* (Share of population living in overcrowded dwellings and also exhibits at least one of some structural problem) and *Low work intensity* (Proportion of people 0–59 living in households in which household members of working age worked less than 20% of the number of months that could theoretically have been worked) measure important topics, but, according to our views, they can't be considered as indicators of economic well-being from a theoretical point of view.

3 Methodological Aspects

The composite index was constructed using the Adjusted Mazziotta-Pareto Index—AMPI [5]. This aggregation function allows a partial compensability, so that an increase in the most deprived indicator will have a higher impact on the composite index (imperfect substitutability). Such a choice is advisable whenever a reasonable achievement in any of the individual indicators is considered to be crucial for overall performance [3]. The most original aspect of this index is the method of normalization, called "Constrained Min–Max Method" [6]. This method normalizes the range of individual indicators, similarly to the classic Min–Max method, but uses a common reference that allows to define a 'balancing model' (i.e., the set of values that are considered balanced). Thus, it is possible to compare the values of the units, both in space and time, with respect to a common reference that does not change over time.

Let us consider the matrix $\mathbf{X} = \{x_{ijt}\}$ with 27 rows (countries), 4 columns (individual indicators), and 4 layers (years) where x_{ijt} is the value of individual indicator j, for country i, at year t. A normalized matrix $\mathbf{R} = \{r_{ijt}\}$ is computed as follows:

$$r_{ijt} = 100 \pm \frac{x_{ijt} - x_{j0}}{\max\limits_{it}(x_{ijt}) - \min\limits_{it}(x_{ijt})} 60$$

where $\min\limits_{it}(x_{ijt})$ and $\max\limits_{it}(x_{ijt})$ are, respectively, the overall minimum and maximum of indicator j across all times (goalposts), x_{j0} is the EU average in 2007 (reference value) for indicator j, and the sign $\pm$ depends on the polarity of indicator j.

Denoting with $\mathrm{M}_{r_{it}}$, $\mathrm{S}_{r_{it}}$, $\mathrm{cv}_{r_{it}}$, respectively, the mean, standard deviation, and coefficient of variation of the normalized values for country i, at year t, the composite index is given by:

$$\mathrm{AMPI}_{it}^{-} = \mathrm{M}_{r_{it}} - \mathrm{S}_{r_{it}} \mathrm{cv}_{r_{it}}$$

where:

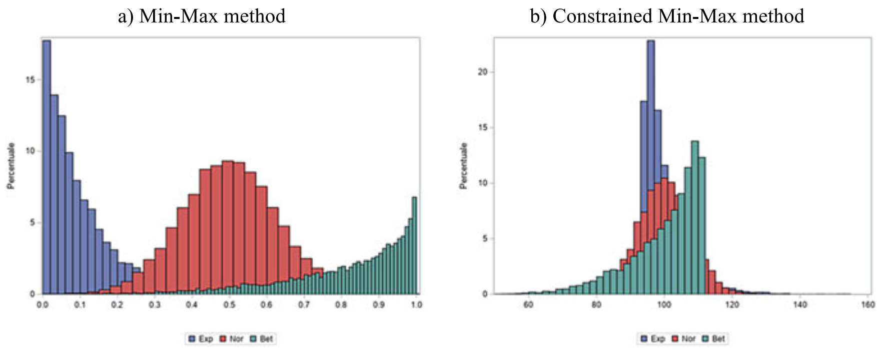

Fig. 5 Comparing the classic and the constrained Min–Max method

$$\mathrm{M}_{r_{it}} = \frac{\sum_{j=1}^{4} r_{ijt}}{4} \quad \mathrm{S}_{r_{it}} = \sqrt{\frac{\sum_{j=1}^{4} (r_{ijt} - \mathrm{M}_{r_{it}})^2}{4}} \quad \mathrm{cv}_{r_{it}} = \frac{\mathrm{S}_{r_{it}}}{\mathrm{M}_{r_{it}}}.$$

The version of AMPI with a negative penalty was used, as the composite index is 'increasing' or 'positive', i.e., increasing values of the index correspond to positive variations of the economic well-being. Therefore, an unbalance among indicators will have a negative effect on the value of the index [5].

Figure 5 shows the effect of normalization on three individual indicators with different shape generated in a simulation.[2] The first has an exponential distribution with $\lambda = 0.0125$ (Exp), the second has a normal distribution with $\mu = 150$ and $\sigma = 15$ (Nor) and the third has a Beta distribution with $\alpha = 4$ and $\beta = 0.8$ (Bet). The indicators have different parameters, as they represent the most disparate phenomena. In Fig. 5a, indicators are normalized by the classic Min–Max method in the range [0, 1], and in Fig. 5b, they are normalized by the constrained Min–max method with a reference (the mean) of 100 and a range of 60.

As we can see, the Min–Max method bring all values into a closed interval, but the distributions of indicators are not 'centred' and this leads to the loss of a common reference value, such as the mean. It follows that equal normalized values (i.e., balanced normalized values) can correspond to very unbalanced original values. For example, the normalized value 0.2 for the Exp indicator corresponds to a high original value; whereas for the Nor and Bet indicators it corresponds to a very low original value. Moreover, the normalized value 0.5 is the mean of the range, but not of distributions, and then it cannot be used as a reference for reading results (e.g., if the normalized value of a country is 0.3., we cannot know if its original value is

[2] Note that socio-economic indicators are basically of two types: per capita indicators and percentage indicators. Per capita indicators tend to be open-ended, in particular at the upper end of the range (e.g., *GDP per capita*); percentage type indicators tend to have severe constraints operating at the upper end of the range, with consequent piling up of observation there (e.g., *Adult literacy*). Therefore, most of individual indicators have positively or negatively skewed distributions [2].

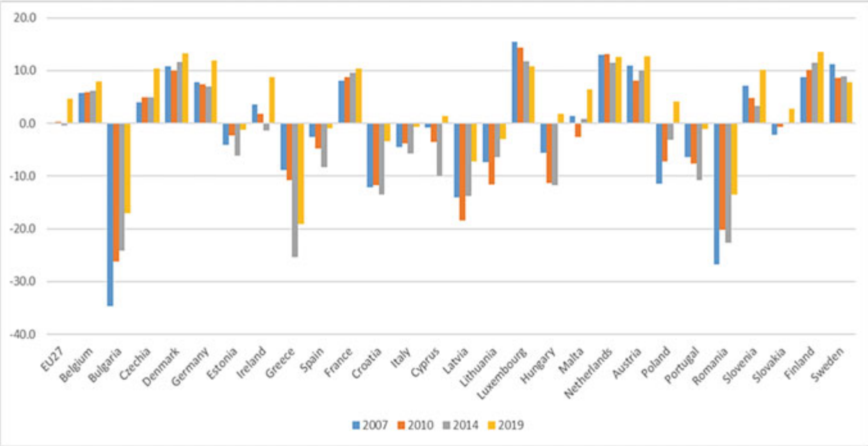

Fig. 6 AMPI indicator. Distance from the reference value (EU27 in 2007 = 100) in the EU27 countries. Years 2007, 2010, 2014 and 2019

above or below the mean). On the other hand, normalized values by the constrained Min–Max method are not forced into a closed interval, they are 'centred' with respect to a common reference, and they are easier to interpret: if the normalized value of a country is greater than 100, then it is above the reference value, else it is below the reference value (Fig. 6). Finally, the comparability across time is maintained when new data become available (the goalposts do not need to be updated).

4 A Longitudinal Analysis

In the analysis, the reference value is the Eu27 in 2007 (=100), and each value can be evaluated as the relative distance to 100 (Table 1). The Eu27 indicator is not far from 100 neither in 2010 (100.2) nor in 2014 (99.6). The last year shows instead an increase of the overall index of about 5 point (104.7 in 2019).

Before commenting the different phases, it can be of interest to observe which indicators have the greatest impact in the AMPI. All the four primary indicators (one positively defined and three negatively defined in respect of economic wellbeing) are obviously highly correlated with the AMPI. However, the one that shows the highest correlation is the poverty indicator (severe material deprivation), about −0.90 in the four years, while the one with the lowest correlation is the inequality index (at risk of poverty rate), that decreases from −0.80 in 2007 to −0.75 in 2019 (Table 2).

The first phase, corresponding to the international economic crisis, is the most stable. Indeed, the ranking, based on the AMPI, shows a low level of variability between 2007 and 2010, as well as the values of the AMPI. The highest jump in the AMPI absolute value is observed for Poland, which also passes from the 23rd

Table 1 AMPI value and ranking in the EU Countries, years 2007, 2010, 2014 and 2017

Country	Year							
	2007		2010		2014		2019	
	AMPI	Rank	AMPI	Rank	AMPI	Rank	AMPI	Rank
Belgium	105.7	10	105.9	9	106.1	9	107.9	11
Bulgaria	65.3	27	73.7	27	75.9	26	83.0	26
Czechia	104.0	11	104.9	10	105.0	10	110.3	8
Denmark	110.9	5	109.9	4	111.7	2	113.2	2
Germany	107.8	8	107.4	8	106.9	8	111.8	5
Estonia	95.9	17	97.7	14	93.9	17	98.8	21
Ireland	103.6	12	101.8	12	98.6	14	108.8	10
Greece	91.1	22	89.2	21	74.6	27	80.9	27
Spain	97.4	16	95.3	18	91.7	19	99.0	19
France	108.1	7	108.7	5	109.6	6	110.4	7
Croatia	87.9	24	88.2	24	86.4	23	96.6	23
Italy	95.4	18	96.2	17	94.2	16	99.4	18
Cyprus	99.2	14	96.4	16	90.0	20	101.4	17
Latvia	86.0	25	81.6	25	86.2	24	92.8	24
Lithuania	92.7	21	88.3	23	93.6	18	97.1	22
Luxembourg	115.4	1	114.3	1	111.7	1	110.7	6
Hungary	94.5	19	88.6	22	88.3	22	101.8	16
Malta	101.3	13	97.4	15	100.8	12	106.4	13
Netherlands	113.0	2	113.1	2	111.5	3	112.5	4
Austria	111.0	4	108.1	7	110.0	5	112.8	3
Poland	88.5	23	92.8	19	96.8	15	104.1	14
Portugal	93.7	20	92.4	20	89.3	21	98.9	20
Romania	73.2	26	79.8	26	77.3	25	86.5	25
Slovenia	107.2	9	104.8	11	103.3	11	110.2	9
Slovakia	97.9	15	99.4	13	99.9	13	102.8	15
Finland	108.8	6	110.2	3	111.5	4	113.5	1
Sweden	111.3	3	108.6	6	109.0	7	107.7	12
EU27	**100.0**		**100.2**		**99.6**		**104.7**	

position to the 19th in 2010. In the second phase, corresponding to the crisis of the sovereign debt, there is a greater mobility in the ranking. Greece shows the highest jump, from 21st to 27th and last position. The Greek AMPI decreased dramatically from 89.2 to 74.6. This fall was mainly due to a dramatic fall in the purchasing power of the households (the median equivalised income in Pps decreased from 12,598 to 8,673 euros). Also, the SMD and the subjective economic distress greatly worsened, respectively from 11.6% to 21.5% and from 24.2% to 39.5%). Indeed Greece was

Table 2 Correlation coefficient between the primary indicators and the AMPI in the different years

Indicator	Year			
	2007	2010	2014	2019
Median equivalised income in pps	0.82	0.83	0.83	0.80
At risk of poverty rate	−0.80	−0.76	−0.79	−0.75
Severe material deprivation	−0.91	−0.90	−0.90	−0.90
Index of self-reported economic distress	−0.90	−0.89	−0.89	−0.79

the first country to be hit by the equity markets distrust on the debt sustainability, later followed by Portugal and Ireland and successively by Italy and Spain. In the 2010–2014 phase, Ireland loses two positions (from 12 to 14th), Spain and Portugal one position (respectively, from 18 to 19th and from 21st to 22nd), while Italy gained one position (from 17 to 16th). However, also Italy showed a decrease in the synthetic index, from 96.2 to 94.2 and the overall Italian situation was somewhat preserved by the fact that only the SMD indicator worsened (from 7.4 to 11.6%), while the other three were substantially unchanged. In this time, we can observe a new great advance of Poland (+4 in the ranking, from 19 to 15th), which shows an increase of the MPI from 92.8 to 96.8.

In the opinion of the authors, these data clearly show that the European response to the sovereign debt crisis has done more harm than good. The vexatious conditions imposed to Greece by the European Commission, European Central Bank and International Monetary Fund highly worsened the household economic conditions of the Country and were badly used as a warning for the other indebted Countries. Unsurprisingly, they were instead used by the stock markets' operators as a sign of permit towards speculation, which quickly enlarged against the other Countries. Luckily, when the entire Eurozone was in doubt, the European institutions changed their policies. IMF was involved less intensely; the ECB completely changed its monetary policy, which originally just looked at an about non-existent inflation and did not foresee an intervention on the stock markets (the Quantitative Easing started in 2012 in order to support the financial system and to save the Euro area; somewhat enlarged its effects on the productivity system in 2014; and started its second and stronger phase in 2015, with an always greater intervention); and the Eurozone, even in a context of a formally stricter balance observation through the *fiscal compact*, contemplated a series of adjustments which allowed to keep in account a number of factors (e.g., the years of general economic crisis as well as the notion of "potential GDP") rather than applying in aseptic way the treaties. The new policies facilitated the growth of the GDP as well as an improvement in the households' economic conditions in Europe, as observed in the data. Indeed, in the last phase, till 2019, the overall Eu27 index passed from 99.6 to 104.7, showing a general increase on the economic well-being of the households, and all the Countries, but Sweden and Luxemburg, increased the value of the index. Some Countries had a particularly great increase (Hungary, Cyprus, Croatia and Ireland, more than +10 points). As concerning the ranking, Hungary showed the greatest increase, + 6 positions, especially due to an improvement in median purchasing power, SMD and subjective economic distress; Luxemburg and Sweden showed the greatest decrease, −5 positions, especially due to an increase of inequality as measured by the ARP rate in a context of general decrease of inequality in the European zone.

Considering the entire time frame, 2007–2019, some Countries greatly increased their economic well-being, in particular, and somewhat obviously given that the economic convergence is one of the targets of the EU, the Countries that started from a disadvantaged situation: Bulgaria (+17.7), Poland (+15.6) and Romania (+13.3). In the case of Poland, this also pushed the ranking, from the 23rd to the 14th position; Bulgaria and Romania still remain at the bottom tail of the ranking, respectively 26th and 25th in 2019, +1 position for both the Countries, but strongly filled the gap in respect of the overall EU27. At the opposite, the Greek indicator has fallen down by 10.1 point (even if it is growing in the last sub-time), completely due to the 2011–2014 time frame. At present, Greece is in the last position of the ranking, 27th (vs 22nd in 2007), while the first position is occupied by Finland. The other two Countries with an important decrease in the MPI indicator are Luxemburg, −4.8 points, and Sweden, −3.5 points, which shifted, respectively, from 1st to 6th position and from 3rd to 12th position.

Summarizing, the overall time frame is divided in the following phases of the international economic crisis: 2007–2010; the phase of the Eurozone crisis, 2011–2014; and the last phase of economic stability, 2015–2019. The first two appear to be a long period of unique crisis, slightly softer and more diffuse in the first part; more intense and localized in a fewer number of Countries in the second part. The particularity of this second phase is that even the Countries which didn't face the crisis of the sovereign debts didn't improve their economic well-being, showing that the entire Europe has faced it as well, and the only way for advancing is solidarity and reasonability. The last phase was indeed characterized by a general increase of the European households' economic well-being, mostly due to a more reasonable and rational use of the fiscal policies and, especially, of the monetary policies. Such measures have permitted to relax the economic distress on the European households.

As concerning Italy, it started 18th in 2007 and is 18th in 2019, with a negative gap in respect of the EU27 which ranged from −4 to −5.5 points in the phase. Looking at the sub-domains, Italy has improved its purchasing power, even though with a grow rate lower than the Eu27; the ARP is very stable along the phase; the SMD indicator shows a high value only in 2014 (11.6% vs a little more than 7% in the other years); the subjective economic distress about halves in the last phase (from 17.9% to 8.2%), even due to a more stable economic situation which reflects on the opinion of the households.

5 Conclusions

In this paper we analysed the economic well-being in Europe in the 2007–2019 time frame at 4 relevant years: 2007, 2010, 2014 and 2019. In order to do so, we have considered the economic domain of the project BES (Equitable and Sustainable Well-being) of the Italian National Institute of Statistics (Istat), changing its definition in accordance with data comparability and theoretical issues. Such domain was measured through four sub-domains: *Purchasing Power is measured by* the *Median equivalised income in purchasing power standards (Pps). Inequality is measured by* the *At risk of poverty rate (ARP). Poverty* is measured by the *Severe material deprivation (SMD);* and *Subjective evaluation* is measured by the *Index of economic distress.*

Following the aforementioned years, the analyses considers 3 relevant phases: 2007–2010, which comprises the international economic crisis; 2011–2014, which comprises the crisis of the European sovereign debts; and 2015–2019, that is characterized by relative stability and recovery. The first two phases appear to be a unique long time frame of crisis, slightly softer and more diffuse in the first part; more intense and localized in a fewer number of Countries in the second part, particularly heavy for Greece. The peculiarity of this second phase is that even the Countries which didn't face the crisis of the sovereign debts didn't improve their economic well-being. This fact clearly show that the European response was far from being satisfactory, and the vexatious conditions imposed to Greece by EC, ECB and IMF highly worsened the Greek household economic conditions and were badly used as a warning for other indebted Countries. On the other hand, such measures stimulated the stock markets' speculation, which quickly enlarged against the other Countries, and the entire Eurozone was put in doubt. Luckily, fiscal and monetary policies have completely changed since then. The IMF was involved less intensely; the Eurozone, even in a context of a formally stricter balance observation through the fiscal compact, contemplated several adjustments which allowed to keep in account different factors; and, mainly, the ECB completely changed its monetary policy (through the quantitative easing that started in 2012 and enlarged its dimension starting from 2015). Such measures have permitted to relax the economic distress on the European households and all European countries have resumed the normal path towards the higher and generalized economic well-being that characterized the whole post-war period. In this regard, it has to be noted that Germany, the Country that undoubtedly has leaded the EU in the considered time frame, does not improve its position neither in the value of the indicator nor in the ranking (8th) till 2014, while it increased the value of the indicator (+4.9 points) and the ranking (+3 positions) in the last phase, when "less German" fiscal and monetary policies were applied (certainly with the agreement of Germany itself), as a further confirmation of the fact that the only way for economic advancing in EU is solidarity and reasonability.

References

1. Canberra Group: Handbook on Household Income Statistics. United Nations, Geneva (2011)
2. Casacci, S., Pareto, A.: A nonlinear multivariate method for constructing profiles of country performance. Int. J. Multicriteria Decis. Mak. **8**, 295–312 (2021)
3. Chiappero-Martinetti, E., von Jacobi, N.: Light and shade of multidimensional indexes. How methodological choices impact on empirical results. In: F. Maggino, G. Nuvolati (eds), Quality of life in Italy, Research and Reflections, pp. 69–103, Springer, Cham (2012)
4. Cutillo, A., Raitano, M., Siciliani, I.: Income-based and consumption-based measurement of absolute poverty: insights from Italy. Soc. Indic. Res. (2020). https://doi.org/10.1007/s11205-020-02386-9
5. Mazziotta, M., Pareto, A.: On a generalized non-compensatory composite index for measuring socio-economic phenomena. Soc. Indic. Res. **127**, 983–1003 (2016)
6. Mazziotta, M., Pareto, A.: Everything you always wanted to know about normalization (but were afraid to ask). Ital. Rev. Econ., Demogr. Stat., LXXV **1**, 41–52 (2021)
7. Sen, A.: Poor, relatively speaking. Oxford Econ. Ser. **35**(2), 153–169 (1983)

A Dynamic Power Prior for Bayesian Non-inferiority Trials

Fulvio De Santis and Stefania Gubbiotti

Abstract Non-inferiority trials compare new experimental treatments to active controls. Previous information on the control treatments is often available and, as long as the past and the current experiments are sufficiently homogeneous, historical data may be useful to reserve resources to the new therapy's arm and to improve accuracy of inference. In this article we propose a Bayesian method for exploiting historical information based on a dynamic power prior for the parameter of the control arm. The degree of information-borrowing is tuned by a quantity based on the Hellinger distance between the two posterior distributions of the control arm's parameter, obtained respectively from the current and the historical experiments. Pre-posterior analysis for type-I error/power assessment and for sample size determination is also discussed.

Keywords Clinical trials · Hellinger distance · Historical data · Power prior · Sample size determination

1 Introduction

A Non-inferiority (NI) trial is an experiment where a new treatment is compared to an existing active therapy (control). Unlike trials in which a new effective treatment must be shown to be superior to the placebo, the objective of a NI trial is to establish that the difference between the effects of the new and the control treatments is small enough to conclude that the new drug is also effective. NI trials are therefore typically used to draw inference on the unknown parameter: if the hypothesis that the unknown parameter is below a given *non-inferiority margin* is rejected, one can conclude for NI. The importance of "borrowing" information from previous studies is that "with

F. De Santis · S. Gubbiotti (✉)
Dipartimento di Scienze Statistiche, Sapienza Universitá di Roma, Piazzale Aldo Moro N. 5, 00185 Roma, Italy
e-mail: stefania.gubbiotti@uniroma1.it

F. De Santis
e-mail: fulvio.desantis@uniroma1.it

N. Salvati et al. (eds.), *Studies in Theoretical and Applied Statistics*, Springer Proceedings in Mathematics & Statistics 406, https://doi.org/10.1007/978-3-031-16609-9_2

historical data providing information on the control arm, more trial resources can be devoted to the novel treatment while retaining accurate estimates of the current control arm parameters" [28]. This may result in more accurate inference, as long as historical information is sufficiently similar to the current control data. Techniques to borrow information from historical data have been developed both from the frequentist and the Bayesian perspectives. One advantage of the Bayesian approach is that it allows a very natural way to exploit this historical information on the control parameter: data from previous experiments can in fact be used to define a distribution for the effect of the control treatment to be employed as prior in the current experiment. In this regard, see, among others [4, 8–11, 18, 19, 22]. Bayesian methodology also provides several approaches to discount the level of borrowing from historical data. These methods take into account the degree of compatibility between data from current and past experiments. Such a problem has been addressed, for instance, by [14–16, 20, 28]. Among the available alternative borrowing approaches, in this article we focus on the power priors methodology. The idea, originally proposed in the seminal paper [17], prescribes to define the prior for the control parameter to be proportional to a starting density (typically a non-informative prior) times the likelihood associated to historical data raised to a parameter that ranges in [0, 1] and weights the historical data relative to the likelihood of the current study. The power prior parameter tunes the influence of the past data on the distribution of the control parameter: a value equal to 0 is equivalent to no incorporation of historical data in the prior; a value equal to 1 corresponds to full borrowing; intermediate values imply partial borrowing. The choice is then crucial especially when there is heterogeneity between the previous and the current trial or when the sample sizes of the two studies are significantly different. Several methods have been proposed to deal with this choice. In addition to the two main strategies, that consider either a fixed value or a random variable with a given density on the unit interval, some authors have recently proposed to consider a function of a measure of congruence between historical and current data. This approach yields the so-called dynamic power prior: see, for instance, [12, 13, 19, 21, 23, 24].

The present article is a wholly Bayesian conversion of the hybrid frequentist-Bayes method proposed by [19] borrowing ideas from [23]. The main features of Liu's approach in [19] are: (i) implementation of a frequentist test for NI; (ii) instrumental use of a dynamic power prior *only* for the selection of the amount of borrowing from historical data (no posterior analysis is considered); (iii) definition of the power prior parameter as an arbitrary function of the p-value for testing the hypothesis of equivalence between the current and historical control true response rates. Features (ii) and (iii) present some controversial aspects. Specifically, for (ii) one can object that an instrumental use of the power prior does not have a clear justification outside a Bayesian context; with respect to (iii), one can call into question the arbitrary choice of the p-value function that may yield any value of the power prior parameter in [0, 1]. For these reasons, in this paper we propose: (i) to make use of a Bayesian test of NI, based on a credible interval for the unknown effects difference; (ii) to consider a power prior to build the posterior distributions of the parameter necessary for feature (i); (iii) to define a new dynamic fraction based on a sensible measure of compatibility between historical and current data using the Hellinger distance (see [23]).

The outline of the paper is as follows. In Sect. 2 we describe the Bayesian methodology that we propose for the NI test. Section 2.1 provides details on the construction of the priors and on the derivation of the corresponding posteriors. Specifically, we assume the dynamic fraction to be a function of the Hellinger distance between the posterior densities of the control parameter given the current and historical data of the control arms, respectively. We also consider the possibility of setting an upper bound to the amount of information borrowed from previous studies in order to avoid that current data is overwhelmed by pre-experimental information. We explore in Sect. 3 the main (posterior) features of the proposed approach in a real NI study on vaccine considered in [19]. In order to address the requirements of regulatory agencies [2, 7], in Sect. 4 we also investigate frequentist properties of our proposal in terms of type-I error and power. In Sect. 5 we introduce a sample size determination criterion. Discussions on Bayesian experimental design and sample size determination can be found, among others, in [1, 4–6, 10, 11, 18, 25, 27, 29]. Finally Sect. 6 contains a discussion.

2 Methodology

Let us consider a two-arms trial where an experimental drug (e) is compared to a standard therapy, here used as control (c). Let θ_e and θ_c denote the corresponding unknown probabilities of success and let X_e and X_c denote the random number of positive responses out of n_e and n_c observations in the two arms. We assume that $X_j|\theta_j \sim Bin(n_j, \theta_j),\ j = e, c$ and that $X_e \perp X_c \,|\, \theta_e, \theta_c$. Non-inferiority of drug e with respect to drug c is assessed if the null hypothesis of the test

$$H_0 : \ \theta_e - \theta_c \leq -\delta \quad vs. \quad H_1 : \ \theta_e - \theta_c > -\delta \tag{1}$$

is rejected, where $\delta > 0$ is a selected NI margin. Adopting the Bayesian paradigm, we proceed as follows. We determine a credible interval $C = [L, U]$ for $\theta = \theta_e - \theta_c$ of level $1 - \gamma$ and we reject H_0 if $L > -\delta$. Determination of C requires the posterior distributions of θ_e and θ_c. We assume that no information on θ_e is available, whereas historical data regarding θ_c can be retrieved. Under these assumptions we construct the prior distributions for θ_e and θ_c and derive the corresponding posteriors, as detailed in the following subsection. Based on these posterior distributions, the lower limit L of the equal tails interval for $\theta = \theta_e - \theta_c$ is simply computed via Monte Carlo. Then, if $L > -\delta$ the null hypotesis of the NI test is rejected.

2.1 Priors Construction

Let $\pi_e(\cdot)$ be the non-informative $Beta(1, 1)$ density prior for θ_e. Given the experimental data x_e and using $\pi_e(\cdot)$ we obtain the posterior $\pi_e(\cdot|x_e)$ that is a $Beta(1 + x_e, 1 + n_e - x_e)$ density. Furthermore, let us assume that a previous study provides historical data (n_h, x_h) yielding information on the control parameter θ_c, where n_h and x_h are the size and the number of successes. As prior for θ_c in the current experiment we then consider its posterior density given x_h. In order to take into account potential heterogeneity between current and historical information on θ_c, we consider the power prior originally defined by [17] as

$$\pi_c^P(\theta_c|x_h) \propto \pi_c^o(\theta_c) \times [f(x_h|\theta_c)]^a, \quad a \in [0, 1] \tag{2}$$

where $\pi_c^o(\theta_c)$ is a starting prior (typically a non-informative prior), $f(x_h|\theta_c)$ the likelihood function of θ_c given the historical data x_h and $a \in [0, 1]$ a discount parameter. The smaller a, the lighter the degree of incorporation of historical information: $a = 0$ corresponds to no borrowing, whereas $a = 1$ implies full borrowing. Noting that $[f(x_h|\theta_c)]^a \propto \theta_c^{ax_h}(1 - \theta_c)^{a(n_h - x_h)}$ and assuming $\pi_c^o(\cdot)$ to be the $Beta(1, 1)$ density, we have that $\pi_c^P(\theta_c|x_h, x_c)$ is the $Beta(1 + ax_h + x_c, 1 + a(n_h - x_h) + n_c - x_c)$ density.

The choice of a is crucial in determining the impact of historical data on the analysis. As an extreme case, if x_h and x_c can be considered fully exchangeable, then we set $a = 1$. The opposite extreme case is obtained by setting $a = 0$, that corresponds to total discard of historical information on θ_c. In the basic definition of power priors, the tuning parameter a is either fixed or random, but it does not depend on the available data. In the *dynamic* power prior, on the contrary, a measures the homogeneity between historical and current control data. A natural choice is to consider a measure of agreement between $\pi_c(\cdot|x_c)$ and $\pi_h(\cdot|x_h)$, where $\pi_j(\cdot|x_j)$ are the posterior densities for the control parameter obtained by updating $\pi_j(\cdot)$ with x_j, $j = h, c$. We here consider $\pi_j(\cdot)$ to be $Beta(1, 1)$ densities. Therefore $\theta_c|x_j \sim Beta(\bar{\alpha}_j, \bar{\beta}_j)$, where $\bar{\alpha}_j = 1 + x_j$ and $\bar{\beta}_j = 1 + n_j - x_j$, $j = c, h$. With this purpose, following [23], we first consider a measure based on the Hellinger distance between the two posterior densities, i.e.

$$d[\pi_c(\cdot|x_c), \pi_h(\cdot|x_h)] = \left(1 - \int_{\mathbb{R}} \sqrt{\pi_c(\theta|x_c) \cdot \pi_h(\theta|x_h)} d\theta\right)^{\frac{1}{2}}. \tag{3}$$

Then, we define the power prior parameter as the product of two factors: the first is a *static* coefficient $\kappa \in [0, 1]$ that provides an upper limit to the *quantity of information* that we are willing to borrow; the second is a *dynamic* fraction that depends on the *commensurability* between current and historical data, i.e.

$$a(x_c, x_h) = \kappa \cdot (1 - d[\pi_c(\cdot|x_c), \pi_h(\cdot|x_h)]). \tag{4}$$

Table 1 Current data (Rotavirus vaccine example)

Arm	j	n_j	x_j	$\hat{\theta}_j$
Experimental	e	558	415	0.74
Control	c	592	426	0.72

Since $d[\cdot,\cdot]$ is a relative distance and $\kappa \in [0, 1]$, then $a(x_c, x_h) \in [0, 1]$: for a given value of κ, the more compatible information provided by $\pi_c(\cdot|x_c)$ and $\pi_h(\cdot|x_h)$, the larger $a(x_c, x_h)$. Note that, for instance, if we set $\kappa = 1$ the amount of borrowing is fully determined by $(1 - d[\pi_c(\cdot|x_c), \pi_h(\cdot|x_h)])$; conversely, if we set $\kappa < 1$ we impose an upper limit to the fraction to be borrowed. This choice makes sense for instance when $n_h \gg n_c$ and we want to downweight historical prior information so that current data are not overwhelmed. It can be easily checked that under our assumptions (3) becomes

$$d[\pi_c(\cdot|x_c), \pi_h(\cdot|x_h)] = \left(1 - \frac{B\left(\frac{\bar{\alpha}_c+\bar{\alpha}_h}{2}, \frac{\bar{\beta}_c+\bar{\beta}_h}{2}\right)}{\sqrt{B(\bar{\alpha}_c, \bar{\beta}_c) \cdot B(\bar{\alpha}_h, \bar{\beta}_h)}}\right)^{\frac{1}{2}},$$

where $B(\cdot,\cdot)$ is the Beta function.

3 Application

In this section we consider an example described in [19], where a NI study is conducted to compare a pentavalent vaccine (RotaTeq) with a placebo against Rotavirus, both administered together with routine pediatric vaccines. The data are the number of subjects in the two groups who give a positive response to vaccination. Let $\hat{\theta}_j = x_j/n_j$, $j = e, c, h$ denote the response rates.

Table 1 reports current data for both experimental and control arms, whereas Table 2 summarizes data on the control related to four different historical studies that are also combined using a meta-analytic model (*pooled*) as in [19]. In addition, for the sake of the following illustration, we consider the *cumulative* data that are obtained by crude aggregation of the four historical datasets. In the original example, Liu sets $\delta = 0.10$. Here with no loss of generality we consider a stricter NI margin by setting $\delta = 0.03$.

Table 3 reports the values of a computed with Eq. (4) for each single historical study and for cumulative and pooled data. Correspondingly the table shows the bounds L and U of the 0.95-credible intervals and $P(H_1|x_c, x_e)$, i.e. the posterior probability of H_1 computed with respect to (2). Cases $a = 0$ (no borrowing) and $a = 1$ (full borrowing) are also considered for comparison. In Study 1, the degree of borrowing is close to 1 due to the high compatibility between current and histor-

Table 2 Historical data (Rotavirus vaccine example)

Study	n_h	x_h	$\hat{\theta}_h$
1	576	417	0.724
2	111	90	0.811
3	62	49	0.790
4	487	376	0.772
Pooled	483	367	0.759
Cumulative	1236	932	0.754

Table 3 Values of a (with $\kappa = 1$ and $\kappa = 0.8$), bounds of C (with $1 - \gamma = 0.95$) and $P(H_1|x_c, x_e)$ for different historical studies

Study	a	L	U	$U - L$	$P(H_1|x_c, x_e)$
1	1	−0.023	0.065	0.088	0.989
2	1	**−0.041**	0.058	0.099	0.940
3	1	**−0.033**	0.067	0.100	0.968
4	1	**−0.045**	0.044	0.089	0.901
Pooled	1	**−0.039**	0.050	0.089	0.938
Cumulative	1	**−0.042**	0.041		0.921
	($\kappa = 1$)				
1	0.917	−0.024	0.066	0.090	0.987
2	0.171	−0.029	0.072	0.101	0.978
3	0.331	**−0.030**	0.071	0.101	0.976
4	0.213	**−0.033**	0.064	0.097	0.969
Pooled	0.346	**−0.034**	0.062	0.096	0.964
Cumulative	0.307	**−0.036**	0.056	0.092	0.958
	($\kappa = 0.8$)				
1	0.734	−0.025	0.067	0.092	0.985
2	0.137	−0.029	0.073	0.102	0.977
3	0.265	−0.028	0.074	0.102	0.979
4	0.170	**−0.031**	0.067	0.098	0.973
Pooled	0.277	**−0.031**	0.066	0.097	0.972
Cumulative	0.245	**−0.035**	0.058	0.093	0.962
—	0	−0.027	0.075	0.102	0.981

ical data, both in terms of response rate ($\hat{\theta}_h \approx \hat{\theta}_c$) and study dimension ($n_h \approx n_c$). Conversely, in Study 2 the degree of borrowing is the lowest due to heterogeneity in both sample size and response rate. Study 3 represents an intermediate situation with respect to the previous cases. It is interesting to note that even though the data from Study 4 and the pooled case are very similar in sample size, the values of a are substantially different as a consequence of the difference between the two response rates. Finally, when $n_h > n_c$, as in the cumulative historical data case, the value of a is smaller than in the pooled case as a combined effect of two conflicting determinants: (i) a response rate (slightly) closer to $\hat{\theta}_c$, which is supposed to yield a (slightly) larger a; and (ii) a sample size n_h much larger than n_c, that reduces the commensurability between the two posterior distributions thus yielding a smaller a. This dynamic results in the desired downweighting of historical prior information. As commented in Sect. 2.1 one may be willing to fix an upper limit to the degree of borrowing by setting $\kappa < 1$. For instance, if $\kappa = 0.8$, in Study 1 a reduces to 0.734 with respect to the previous case, which is equivalent to a 26% reduction in the prior sample size (from 576 to $a \cdot n_h = 0.734 \cdot 576 = 423$). See Table 3 for other numerical examples. For a deeper insight in Fig. 1 (top panel) we consider prefixed values of $\hat{\theta}_h$, we compute x_h for each value of n_h from 10 to 1300, we determine the corresponding value of $a(x_c, x_h)$ and plot $a(x_c, x_h)$ as a function of n_h. Circles denote the values of a obtained in Table 3 for the different historical data sets. Note that this plot shows how the concurrent effect of n_h and x_h determines very different steepness of the curves that describe $a(x_c, x_h)$ as a function of n_h and illustrates how in certain cases even very small changes in n_h and x_h produce significant variations in a (as an example compare the values of a corresponding to Studies 2 and 3). As expected the maximum level of compatibility is achieved for $\hat{\theta}_h = \hat{\theta}_c$ and $n_h = n_c$ (solid line) i.e. when $\pi_c(\cdot|x_c) = \pi_h(\cdot|x_h)$. The same conclusion can be drawn by looking at the dotted curve corresponding to $n_h = 592$ in Fig. 1 (bottom panel), where $a(x_c, x_h)$ is now plotted against $\hat{\theta}_h$ for several fixed value of n_h. First of all, note that the maximum value of a is not monotone with n_h. The level of borrowing is due to the combined effect of $\hat{\theta}_h$ and n_h for given values of $\hat{\theta}_c$ and n_c. For fixed values of n_h, the plots of a are symmetric with respect to the values of $\hat{\theta}_h$, i.e. the level of borrowing only depends on the absolute value of the difference $\hat{\theta}_h - \hat{\theta}_c$. Finally, note that in all the empirical studies considered in this example $\hat{\theta}_h > \hat{\theta}_c$.

Let us now comment on the conclusions of the NI test recalling that values of $L < -\delta$ (bold character in Table 3) do not allow to reject the null hypothesis. First of all, if we consider the full borrowing case ($a = 1$) all studies but the first prevent one from rejecting H_0, whereas ignoring historical information ($a = 0$) implies rejection ($L = -0.027 > -0.03 = -\delta$). Secondly, consider the dynamic borrowing case. In Study 1 the high compatibility with the current control data, both in terms of sample sizes and response rates, implies conclusions consistent with the full and the no borrowing cases (regardless of κ). Conversely, results from Study 2 are very different from current control data ($\hat{\theta}_h \gg \hat{\theta}_c$): full borrowing yields a value of L much lower than $-\delta$, whereas, thanks to the small degree of borrowing ($a = 0.171$), one is able to reject the null hypothesis consistently with the total discount case. Similar considerations apply to Study 3, in which $\hat{\theta}_h$ is smaller than in Study 2 and closer to $\hat{\theta}_c$. Due to the